CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

The Draft Online Safety Bill: Carnegie UK Trust initial analysis

The UK government published its draft Online Safety Bill on 12th May 2021.[1] It will now undergo a three-month period of pre-legislative scrutiny, undertaken by a joint Parliamentary committee, before the revised, final Bill will be introduced later in the year. Pre-legislative scrutiny is intended to examine "big picture" design and structure issues with the government then deciding whether to amend or reject the recommendations of the Committee's report when it revises the Bill for introduction. This analysis sets out some of the issues which we think should be considered.

Leading issues

- The draft Bill has the potential to develop into an effective, evidence-based framework for the regulation of social media companies and search engines to prevent harm arising to people in the UK. This is an achievement, given that the Bill was drafted during a national crisis. On a spectrum of regulation, the regime would sit appropriately between regulation of broadcasting and self-regulation of the press.

- The draft Bill is hard for a lay reader to understand. This will hinder scrutiny and increase regulatory burden. The government should structurally simplify the Bill's three safety duties, three counterbalancing factors and its byzantine commencement process. Specifically, there should be a general safety duty to orientate and give coherence to the regime.

- To meet the UK's international commitments on free speech, there should be a separation of powers between the Executive and a communications regulator. The draft Bill takes too many powers for the Secretary of State. These should be reduced, removing in particular the Secretary of State's power to direct OFCOM to modify its codes of practice to bring them in line with government policy.

- The thresholds for harm underpin the entire regime for children and adults but no process is described for defining "significant harm". The government's intention is that the threshold is low enough to be effective, but this needs more than a side comment in the explanatory notes.

- Harms to adults on the largest platforms are not well covered (Clause 11). The government needs to spell out how huge volumes of racism, misogyny, antisemitism etc – that are not criminal but are oppressive and harmful, particularly to prominent figures – will be addressed. No special treatment is given to protect politicians, candidates and journalists involved in the democratic process.

- The regulator does not have enough powers to address threats to public safety, public security and national security. The Prime Minister and President Biden recently signed a new Atlantic

[1] https://www.gov.uk/government/publications/draft-online-safety-bill

Charter[2] to "oppose interference through disinformation or other malign influences, including in elections". Building such capability into the risk assessments in this Bill would be an easy way of meeting that obligation, demonstrating to the USA how this can be done while respecting free speech. The UK has an opportunity to be a world-leader on this agenda.

- The regime described in the draft Bill could be employed by other regulators to make markets work better and protect the public. The December policy document[3] said that the power to 'co-designate' a regulator to act under the OSB regime would be available, but we cannot see it in the draft Bill. This is strategically important and the government should confirm that this ability is inherited from the Communications Act.

About Carnegie UK Trust

1. Over the past three years, Carnegie UK Trust has shaped the debate in the UK on reduction of online harm through the development of, and advocacy for, a proposal to introduce a statutory duty of care. We believe this is of critical importance to our mission to improve wellbeing. Our proposal is for social media companies to design and run safer systems – not for government to regulate individual pieces of content. Companies should take reasonable steps to prevent reasonably foreseeable harms that occur in the operation of their services, enforced by a regulator.[4] The proposal has been developed by Professor Lorna Woods (Professor of Internet Law, University of Essex), William Perrin (Carnegie UK Trustee) and Maeve Walsh (Carnegie UK Trust Associate), working with the Carnegie UK Trust team. It draws on well-established legal concepts to set out a statutory duty of care backed by an independent regulator, with measuring, reporting and transparency obligations on the companies. A focus on the outcome (harm) makes this approach futureproof and necessarily systemic. We propose that, as in health and safety regulation companies should run their systems in a proportionate, risk-based manner to reduce reasonably foreseeable harm. Broadcast regulation demonstrates that a skilled regulator can work to assess harm in context, regulate it and balance this with maintaining free speech. Proportionality in regulation allows for innovation and market entry by SMEs.

2. We are pleased that the UK government has adopted – in part – our approach. But, as we set out below, it either excludes or omits several components that we have proposed or advocated for. In particular, it does not address: harms to democracy or the electoral process[5]; financial fraud and scams, except where these are via user-generated content[6]; mis- or disinformation

[2] Clause 3 '*Third, we remain united behind the principles of sovereignty, territorial integrity, and the peaceful resolution of disputes. We oppose interference through disinformation or other malign influences, including in elections, and reaffirm our commitment to debt transparency, sustainability and sound governance of debt relief.*' https://www.gov.uk/government/publications/new-atlantic-charter-and-joint-statement-agreed-by-the-pm-and-president-biden/the-new-atlantic-charter-2021

[3] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response

[4] All our work is available here: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/.

[5] See our March 2021 blog post on protecting those involved in the democratic process: https://www.carnegieuktrust.org.uk/blog/increased-online-safety-for-people-involved-in-the-democratic-process-in-the-uk/; and our January 2021 blog post on freedom of speech and political mis/-disinformation: https://www.carnegieuktrust.org.uk/blog/freedom-of-expression-speech-rights-modern-regulation/

[6] Carnegie UK Trust co-signed a letter to the Home Secretary and DCMS Secretary of State on this issue in April 2021: https://www.carnegieuktrust.org.uk/news/cukt-joins-online-scams-coalition/

that has a societal impact[7]. Hate crime is covered, but it is unclear how well hatred short of the criminal threshold will be covered.[8] There is no obvious way for regulators to work together[9] and a reference to "co-designation" powers for OFCOM that appeared in the December policy document has not been transposed to the draft Bill. This paper sets out, in narrative form, our analysis of the structure and intent of the Bill and then goes through a series of specific questions which – we believe – are integral to the effective functioning of the regulations when they come into force and should be addressed during the process of pre-legislative scrutiny. It does not aim to provide a comprehensive analysis of all aspects of the Bill (e.g., we do not address reporting and transparency or what the freedom of expression duties might mean) or drafting issues. As with all our work, we offer this analysis to policymakers, legislators and civil society colleagues in the spirit of our long-term commitment to ensure the introduction of proportionate, systemic regulation that protects the widest number of users from online harms. There will be areas of harm and/or technical or operational issues on which others will have more expertise and we look forward to assessing how their analyses "dock" with our overarching view. As always, we welcome feedback on our work and would invite those with an interest in this area to contact us to discuss further.

Overview of the draft Bill

3. We welcome the fact that the draft Bill describes a regime taking a systemic approach to the regulation of online harms. This is a significant starting point and one for which Ministers and DCMS officials should be commended. For such an approach to achieve its maximum effect it should be linked into an international multilateral framework to build upon the recent G7 Technology Ministers' Statement.[10] The draft Bill is limited, however, by the choices the government has made on where it wants that regulation to bite and – notably – by the areas that it has excluded and those that are ill-defined. These are policy and political choices that will come under intense scrutiny in the weeks and months ahead. We set out some of those we feel are most important in the second section of this analysis but start here by looking at the design of the regulatory framework.

Structural complexity

4. The draft Bill is a framework Bill, providing a structure in primary legislation under which supporting secondary legislation and a plethora of codes and guidance from OFCOM will sit. The draft Bill's complexity flows from the choice the government has made to break the obligations on relevant operators down into multiple duties, some of which apply only to a limited group of operators. The Bill then sets out, for each of these duties and groups, a description of the process OFCOM and the Secretary of State has to follow to fill-in the framework.

---

[7] See our thoughts on how a systemic duty of care would tackle the Covid "infodemic": https://www.carnegieuktrust.org.uk/blog/addressing-the-infodemic-through-a-focus-on-online-system-design/

[8] We have published a draft Code of Practice for Hate Crime and wider legal harms alongside this full response: https://www.carnegieuktrust.org.uk/blog/draft-code-of-practice-in-respect-of-hate-crime-and-wider-legal-harms/

[9] Our proposal for "regulatory interlock" is outlined in this blog post from September 2020: https://www.carnegieuktrust.org.uk/blog/online-harms-interlocking-regulation/

[10] https://www.gov.uk/government/publications/g7-digital-and-technology-ministerial-declaration

5. The intricate nature of the Bill and the difficulty of reading it from end to end as a narrative will make scrutiny and deliberation more difficult. It makes it hard to ascertain what is meant to happen and allows room for people to assert meaning that may not be there. Equally, some of the complexity may lead to unintended outcomes. The regulatory burden is in any event increased. The government could make less complex design choices and provide better explanatory tools during the pre-legislative process – the occasional diagram and timeline would help.

6. There are three separate thematic duties of care, each with an underpinning risk assessment duty. Further, there are a number of procedural duties. Three counterbalancing considerations apply unevenly across categories of operator. The duties for user-to-user services differ depending on whether the service is "Category 1" (the largest/riskiest) or not. The basic range of duties are repeated in a different form for search engines, leading to much repetition and cross-referencing.

7. Unlike the deliberately all-encompassing statutory duties of the Health and Safety at Work or Occupiers Liability (Land) Acts, the government is at pains, first, not to be all-encompassing and, secondly, to make some duties (much) stronger than others. This is also different from the proposal in the Online Harms White Paper[11] and from the Carnegie proposal[12] both of which envisaged an overarching duty of care for all services, but a duty which might apply differently depending on the service.

8. The risk assessment and safety duties for user-to-user services target three areas of concern:
    - Child sexual abuse and terrorism offences and crimes that impact individuals (clauses 5(2), 7(1), and 9);
    - Harm to individual children (cls 5(4)), 7(3)-(4), and 10); and
    - Harm to individual adults (on Category 1 only) (cls 5(5), 7(6)-(7), and 11).

    While OFCOM has a role in determining which user-to-user services are Category 1, Schedule 4 provides constraints on the criteria that will be used in such determination. The risk assessment and safety duties for search engines exclude adult harms altogether (but contain parallel illegal content (cl 17(2), 19(1), and 21) and children's duties (cl 17(3), 19(2) and 22)).

Scope and Thresholds

9. To be of relevance to the duties, the content must meet certain thresholds. For criminal law, in addition to terrorism and child sexual exploitation and abuse (CSEA) offences, relevant offences are when the intended victim is an individual (cl 41(4)(d)) or the Secretary of State specifies the offence falls within scope (cl 41(4)(c) and cl 44) and is not an offence under cl 41(6). Content falls within relevant offences when the service provider has "reasonable grounds" for believing that the content constitutes a relevant offence. Clause 41(5) also envisages that the Secretary of State may use the Cl 44 process to specify content to be "priority illegal content" – we assume this is for the Secretary of State and Parliament to set priorities for OFCOM's regulatory work. CSEA content and terrorism content are not automatically priority illegal content. Special

[11] https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper
[12] Our full reference paper of April 2019 sets this out in detail: https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

obligations arise in relation to priority illegal content, particularly for user-to-user platforms (cl 9 (3)(a)-(c), 21(3)(a)). It appears that the hate crime offences would be covered by clause 41(4)(d), as well as a much broader range of offences, e.g. coercive or controlling behaviour, threats to kill and harassment.

10. For other content, the threshold is that the content gives rise to psychological or physical harm (cl 45(3) in relation to children and cl 46(3) in relation to adults). Content that is harmful to children is content in respect of which there are reasonable grounds for believing there is a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on a child of ordinary sensibilities (cl 45(3)). Similar wording is used in relation to the definition of content harmful to adults, though the reference point is an adult of ordinary sensibilities (cl 46(3)). In both cases, the Secretary of State may designate further content; there is also a priority content Category. For these two categories, the Secretary of State must consult OFCOM before making the regulations and OFCOM is under an obligation to review the incidence of types of designated harmful content. The focus on individual harms may result in boundaries being drawn through certain categories of content – notably misinformation and hate speech.

Risk assessment

11. OFCOM will do a broadly-based market risk assessment (cl 61) that underpins much of the regime and then produce guidance for service providers to assist them in carrying out (cl 62) their own risk assessment for each safety duty. This will take a year or more after Royal Assent. (See annex A for our best guess on the timeline for commencement.)

12. Armed with this guidance from OFCOM, services then have to carry out their own risk assessments. Risk assessments should cover the 'characteristics' of the service, including the impact of the characteristics on the risk of harm from certain defined types of content. Clause 61(6) lists user base, business model, governance and other systems and processes, as well as its functionalities – i.e. the system design, operation of the algorithm etc. It seems, however, that there is little ability for the regulator to critique the providers' own risk assessments.

13. There needs to be a clear connection between the risk assessments, the Online Safety Objectives (cl 30), the carrying out of the duties and enforcement of those by OFCOM to ensure that a 'backstop' exists and applies to systems, processes and content. One way of doing this might be an overarching duty of care.

Safety and other duties

14. The illegal harm duty (cl 9) and the harm to children duty (cl 10) are strong, albeit complex. The (quite different) harms to adults duty (cl 11) is ill-defined and weak and is imposed only on Category 1 providers. This means that the criteria on which this categorisation will rest has great significance for the impact of the regime – many of the high-profile concerns may fall in the "harms to adults" duty.

15. There are three counterbalancing 'protective' duties intended to meet concerns from some about the impact of safety regulation. The duty about rights such as free speech and privacy (cl 12) applies to all user-to-user services (but imposes specific obligations on Category 1 providers). A second group of rights apply to Category 1 providers only: duties in relation to
    - content of democratic importance (cl 13); and
    - protecting content from traditional media that is regulated or self-regulated elsewhere (cl 14).

    Search engines are treated differently – with only a duty about freedom of expression and privacy (cl 23); there is no equivalent of clauses 13 and 14 (as there is no equivalent of a Category 1 provider).

16. OFCOM will also produce Codes of Practice relating to recommended steps for compliance with the safety duties; the Codes must be compatible with the pursuit of the Online Safety Objectives detailed in clause 30. These objectives underpin the codes and apply to any service claiming that it has met the codes of practice by other means. The codes in relation to the safety duties and certain other duties seem to be on a "comply or explain" basis (cl 36) and OFCOM is to take Codes of Practice into account (cl 37) as well as being admissible in court; these provisions do not apply to the guidance as regards risk assessments (cl 62). These Online Safety Objectives are the key to a "safety by design" approach and need to be considered closely. We note for example that there is no explicit obligation to consider the severity of harms posed in cl 30(2)(a). The Online Safety Objectives should be explicit about systems and processes being proportionate to the size of the threat.

17. Part 3 of the draft Bill contains other duties: the obligation to provide annual transparency reports, to notify OFCOM and to pay fees. OFCOM may specify what is to be included within the reports within the categories listed in clause 49(4). While the notification duty applies to all regulated services, the transparency report obligation applies only to specified services; this may limit to oversight of some small but problematic services.

18. In the Bill, the duties are relatively empty frameworks; most will be populated in part by OFCOM research, risk assessments and OFCOM guidance, Codes etc (the latter presented to the Secretary of State for them to lay as an SI). This means that the powers and resources made available to OFCOM will be a significant factor in the regime's success or not. It also means that OFCOM's independence must be safeguarded.

Powers of the Secretary of State

19. It is apparent from the description above that the role of the Secretary of State is significant. Although most Acts envisage some filling in of detail and a role for relevant Secretaries of State, the powers of the Secretary of State go much further than expected (see the table set out in Annex B). Part of this is over-reach (for instance, ensuring OFCOM's guidance is in line with "government policy"), part is a function of using secondary legislation to give Parliament a say in development of the framework.

Enforcement

20. OFCOM's enforcement powers at first glance seem logical, with good information gathering powers etc (Pt 4, Chapter 5). OFCOM, as well as being able to impose fines, is given broad

'business disruption' powers which seem like an effective enforcement suite. Powers are set out in the draft Bill for criminal sanctions against company executives, but the Secretary of State will only commence that part of the Bill in the future after a review by OFCOM and only if certain conditions are met.

21. The regime is focussed on systems and processes and OFCOM's enforcement route is to challenge whether the duties are being broken in general. OFCOM can only get specific items of content taken down in an emergency. The regime does not replace existing causes of action that individual users could take against other users nor the conditional immunity from liability that platforms have against those actions. Moreover, the draft Bill does not specifically envisage class action and/or representative actions. Instead, the draft Bill contains provisions to enable a super-complaint mechanism (cl 106); this allows designated bodies to bring to OFCOM's attention systemic problems – though these will be limited by the thresholds in the draft Bill too.

22. OFCOM's decisions with regard to inclusion on the register of services or to issue a use of technology notice may be appealed (to the Upper Tribunal – cl 104-5).

23. In the next section, we set out a series of detailed questions which need further thought or clarification during the pre-legislative scrutiny period.

Initial Analysis: Questions and issues

Does OFCOM remain independent?

24. The UK should ensure the same institutional standards to defend free speech as it does in other areas of media regulation: OFCOM's independence is central to that. The United Kingdom is a member of the Council of Europe. The Council of Europe recognises the importance of regulatory authorities as contributing to the fostering of an environment favourable to freedom of expression and supporting diversity and pluralism. Independence of the authority is key to it being able to carry out its responsibilities effectively. 'Recommendation Rec (2000) 23 to member states on the independence and functions of regulatory authorities for the broadcasting sector'[13] sets out the conditions for such independence (principles re-affirmed in a 2008 Declaration). Respecting these principles is an important part of the UK's international commitment to free speech, limiting the ability of the executive to interfere with broadcasting regulation. It is a form of self-restraint by the party in power from which they will benefit when in opposition. It allows the regulatory authority the space to take decisions based on the available evidence.

25. We suggest that a similar standard should apply in online safety regulation as it does in other regulatory contexts. Yet, in addition to the approval process for Codes, the draft Bill contains a number of provisions that allow the Secretary of State to issue directions to OFCOM as to how to fulfil its duties. Notably, clause 33(1)(a) allows the Secretary of State to modify a code of practice to "reflect government policy". While clause 33(3) contains some limitations on the exercise of this power, there is a concern that it might undermine OFCOM's independence. Removal of this provision is, in our view, desirable and would reaffirm of regulatory independence.

Is the regime systemic?

26. We welcome the systemic approach that the draft adopts, underpinned by risk assessments. We see the regulatory framework as "systemic" in two aspects. In its risk assessment obligations (cl 7(8)-(10), cl 16 (1)(6) and cl 61), as well as through the Online Safety Objectives, it focuses on the design and business model of the platforms, covering the processes and functioning of the platform as well as its 'characteristics'. The use of a risk assessment-based model emphasises the roles of systems and processes in governance.

27. Although this is positive, we do have questions as to whether the drafting consistently reflects this approach. The fact that the threshold for action is defined by specified types of content may make it difficult to assess the contribution of the systems design on content creation and dissemination. For example, an individual item of self-harm content, which is not illegal, may not trigger the threshold for adult or children's safety duties as, on its own, it does not lead to the required level of harm. However, repeatedly sending this content to an individual by a service provider's systems may however have such an impact. Another example can be seen in the response on Twitter to Yorkshire Tea's riposte to a user's criticism directed to Yorkshire Tea of their teabags appearing in a photograph with Rishi Sunak. Yorkshire Tea's tweet was popular, and was much re-tweeted, with many people tagging the original user leading to that user's account being temporarily inundated. In TattleLife, for example, the platform is structured so as

[13] https://www.coe.int/en/web/freedom-expression/committee-of-ministers-adopted-texts/-/asset_publisher/aDXmrol0vvsU/content/recommendation-rec-2000-23-of-the-committee-of-ministers-to-member-states-on-the-independence-and-functions-of-regulatory-authorities-for-the-broadcas

to direct criticism and gossip at named individuals; while the comments individually are unlikely to be harmful at the threshold specified in the definition of "content harmful to adults", the targeting might mean that cumulatively there is a problem – effectively, the platform facilitates, through its design, pile-ons. These are all consequences of the system design, not the individual items of content.

28. While OFCOM's Codes and guidance may recognise this issue, arguably they apply once the harm threshold has been triggered. So, while the operation of the system with regard to items of content that are in themselves sufficiently harmful is caught, it is unclear whether the situation where the system plays a part in getting the content to the severity threshold is. Given the sorts of considerations listed in clause 61 (as well as the Online Safety Objectives), it may be that this broader interplay between content and system is intended to be included. Yet, the obligation to take types of content into account refers to the defined terms – "content that is harmful to children" and "content that is harmful to adults" – both of which have a prior severity threshold built in.

29. Similar points may be made with regard to cross-platform harms; for example, funnelling of users on to other platforms (where more extreme or illegal content may be found) – and note even the illegal content risk assessment only refers to content encountered "by means" of the platform. Does that mean just illegal content encountered on the platform, or does it extent to illegal content on other platforms where the first platform has nonetheless played a role in the illegal content being encountered?

30. Moreover, there is no overarching duty of care. This choice may have implications beyond complexity. An over-arching duty would have the advantage of providing a clear indication of the orientation of the duty; that is, the operator has an obligation towards user safety. While it could be argued that the 'online safety objectives' (Cl 30) and the characteristics in cl 61 identify some design features, we have concerns about how they feed into the risk assessment and safety duties, especially as there are no quality requirements surrounding the risk assessment.

31. The Carnegie proposal and the White Paper (as well as the Health and Safety at Work Act), took as their starting point the obligation to take reasonable steps, implicitly referring back to principles found in tort law – that reasonable steps should be taken with respect to foreseeable harm. In the draft Bill, the obligations with regard to the safety duties is to take "proportionate" steps, rather than reasonable ones – but is unclear what proportionate refers to. Current legislation identifies conditions for proportionality[14] (e.g. the size of the platform, the significance of the threat, types of user), but those rules separately require service providers to take "appropriate measures". It is unclear whether the draft Bill deals with whether the measures are appropriate, though the objective of 'effective management' of risks may provide some comfort, as does the fact that the Codes are backed up by a "comply or explain" approach (see cl 36) (see also legal effect of codes in cl 37). The scope of obligation does depend on what are perceived to be risks in the first place (cl 9(2), 21(2)) and there are no qualitative requirements around the risk assessment; in particular, there is no requirement that the operator act reasonably in assessing whether a risk exists or not. OFCOM's guidance (based on its risk assessment under clause 61) would not be binding (cl 62). Would this lead to operators

[14] S 368Z1(4) Communications Act 2003

avoiding looking at problematic aspects of their services? Given the fundamental role of the risk assessment, a weakness here could be felt through the entire system.

32. A general duty avoids the risk that concerns – specifically those that are not well caught by a regime in part focussed on content and content classification - fall through the gaps, whether now or as technology or the way the services are used changes. As we argued in our 2019 report[15], adopting a general duty contains an element of future-proofing and we suggest that such a duty be re-instated, with the current duties providing the differential obligations that arise from the different types of risk and threat (for which types of content are essentially proxies).

What is the scope of harms included?

33. While some categories of criminal offence are identified in the draft Bill as is the threshold for harm to children and adults, the draft allows for the addition to and subsequent amendment of these categories. We note the evidence given to the Lords Digital and Communications Committee by the Digital Minister Caroline Dinenage that a deliberate decision was made not to add harms to the face of the Bill due to the speed at which the landscape evolves and to leave flexibility for new (content that is deemed to cause) harms to be added[16]. However understandable this is, it gives rise to a number of questions around the scope of these provisions and the process.

34. One issue as to scope concerns whether the Secretary of State can choose to include priority content which does not satisfy the test of the victim being an individual in cl 41(4)(d). While clause 44(1) specifies that the Secretary of State is to take into account of level of risk of harm being caused to individuals, the Secretary of State is not bound by this. Clause 44(5) does preclude the Secretary of State from trying to introduce offences excluded from the regime by clause 41(6). Possibly, the Secretary of State could also specify harms that did not otherwise meet the harm threshold. These possibilities raise the question of what sorts of things are envisaged in that, if the Secretary of State chooses to go down that route. Is this, for instance, where the Government intends to include racism and misogyny of the sort directed at footballers and other prominent figures? Or other forms of distressing abuse that fall short of a criminal offence?

35. The regulation-making process under clauses 44-46 gives Parliament some involvement in the harms to be emphasised in the regime. Nonetheless, this intervention is limited. Parliament needs to consider how it could exercise its views in an SI-making process where the Secretary of State has the initiative. One option is for the "Grimstone Rule" apply here, as described in relation to Trade Bill negotiations:

> "This Parliament will now need, in the way that it works, to address four major points in any future statutory system, although they will be covered by the Grimstone rule: approval of the initial objectives, review of the progress of negotiations, considerations of the final

[15] https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

[16] Caroline Dinenage evidence to Lords Digital and Communications Committee 11th May 2021 https://committees.parliament.uk/oralevidence/2187/pdf/

proposed agreement including changes to existing statutory provisions, and parliamentary approval of the deal and any subsequent changes to legislation that may be required."[17]

Another option might be a joint committee to examine proposed SIs.

36. We would welcome a process whereby the Secretary of State's intention with regard to priority harms is published during the pre-legislative scrutiny. It may indeed be possible to move to a process where at least some of the priority harms are included in the Bill. Where OFCOM's research is required, the government should consider whether some 'foundation' research could be done in the six months or more before the Bill is introduced so as to allow for more to be put on the face of the Bill. We note that OFCOM is taking small steps in this direction with research on disinformation published recently and explicitly flagged as preparation[18]. But a lot more will be required and OFCOM should have the confidence of the government's support.

37. We envisage that hate crime will come under the category of "priority illegal content" but, while draft codes of practice on CSEA and terrorism have already been published alongside the full government response, there is no such code in development on hate crime. This could potentially lead to delays in enforceable action by the companies in this area. Hate speech which falls short of the criminal threshold could be designated as priority harmful content (both in respect of adults and children). We have been working with civil organisations – prior to the publication of the draft Bill – to look at what a model draft Code of Practice for Hate Crime might include and we have published this alongside our response for feedback and refinement[19]. We do not expect that it will be picked up "as is" by OFCOM, not least as it does not align perfectly with the draft Bill, but we hope that it will provide additional substance to work with and expedite the development of the thinking in this area as the draft Bill goes through PLS.

<u>How is legal but harmful defined and addressed?</u>

38. Legal but harmful is dealt with in relation to children and adults separately, but both sets of provisions have the same structure and harm threshold. It seems the intention is that OFCOM's overarching risk profile (clause 61), on which they will be required to consult widely, which then leads to the cascade of risk assessment and risk mitigation duties that fall to companies. The continuous nature of the risk-assessment process will, to an extent, also provide an element of future proofing, whereby new risks can be identified and addressed. The mitigation duties differ, however, between the adults' safety duty and the children's safety duty.

39. The threshold of psychological or physical harm is significant – if this is too high then this part of the regime will be greatly limited in its effect. Note the requirement is that the adverse impact must be 'significant'. The threshold is not elaborated on the face of the Bill and only described

---

[17] Lord Stevenson of Balmacara, Debate on the Trade Bill, 23 February 2021, Hansard (HL) vol 810, col 728, available: https://hansard.parliament.uk/lords/2021-02-23/debates/8F92FB28-D51A-4996-8471-7CCAB472F224/TradeBill and Lord Grimstone https://hansard.parliament.uk/Lords/2021-02-23/debates/8F92FB28-D51A-4996-8471-7CCAB472F224/TradeBill Trade Bill Volume 810: debated on Tuesday 23 February 2021 Col 723

[18] https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/online-nation

[19] https://www.carnegieuktrust.org.uk/blog/draft-code-of-practice-in-respect-of-hate-crime-and-wider-legal-harms/

vaguely in the Explanatory Notes with regard to children (EN para 273[20]). The meaning of "psychological harm" is potentially problematic in this regard. Given the regime is based on the duty of care, existing meanings from tort law may affect the threshold. In tort law, similar sounding thresholds for psychological harm have been set so high as to be of little use: they tend to revert to something like 'a recognised psychiatric condition/injury' i.e. a medical definition. Similar concerns arise in the criminal law context. – the Law Commission has criticised both[21]. We understand that the government's intention is for the threshold to be below the much-criticised high threshold of the criminal law on emotional harm and that in tort. This is desirable. But such a vital threshold needs to be set out either as a task for OFCOM to define or, preferably, on the face of the Bill.

40. As noted earlier, the Bill is not clear as to whether an assessment of harm is to be done by considering the impact of an individual item of content, or the cumulative impact of such content taken together (note the word content is the same whether referring to either a single item or to multiple items). How OFCOM interprets these in a regulatory regime needs to be explained. These thresholds are central to the regime.[22] The Bill should be explicit that the relevant threshold of harm can be reached by the operation of the platforms' systems and not just be reference to content alone. The government or OFCOM should expand upon how systems and processes can cause harm.

41. The thresholds for children should be at the least the same as the existing rules for video-sharing platforms in the Communications Act.[23] Under-18s are to be protected from restricted material which includes material that might "impair the physical, mental or moral development of persons under the age of 18", following the principle that material that has the most potential to so harm those under the age of 18 must be subject to the strictest access control measures.[24] The wording of the draft Bill seems to set a higher threshold for intervention, lowering protection.

---

[20] NB explanatory notes para 273 and 275 provide more detail which suggest threshold will be lower, but this is not on the face of Bill*: e.g., "content risks directly or indirectly having a significant adverse physical or psychological impact on a child of ordinary sensibilities. This could be by indirectly resulting in physical injuries or by directly or indirectly resulting in a significant negative effect on the mental state of an individual. This could include causing feelings such as serious anxiety and fear, longer-term conditions such as depression and stress and medically recognised mental illnesses, both short term and permanent"; and "content may be harmful to children in the way in which it is disseminated, even if the nature of the content is not harmful, for example repeatedly sending apparently innocuous content to a user could be bullying and intimidating. in determining whether content is harmful, provider should take into account how many users could be encountering the service and how easily, quickly and widely the content can be disseminated on the service"*.)

[21] Law Commission Liability for Psychiatric Illness, 10 March 1998 (LC249); Law Commission, Harmful Online Communications: The Criminal Offences, 11 September 2020 (Consultation Paper 248).

[22] We note that the Digital Minister, in her appearance before the Lords Digital and Communications Committee on 11th May, responded to a question from Viscount Colville on this definition and confirmed that the harms covered by it would be subject to secondary legislation, with a list "compiled by Ofcom, working with expert advice, subject to democratic oversight and parliamentary debate". Companies would then be "free to decide" how to address the risk and set it out in their terms and conditions. Ms Dinenage also elaborated on the difference between preventing adults "being offended" (which was not the aim of the Bill) and the impact of extremely harmful or extremely emotive content that is often spread by algorithms or where pile-ons target an individual, concluding that "there is a clear distinction between finding something offensive and the potential to cause harm". Secondary legislation informed by Ofcom once they have taken expert insight. (https://committees.parliament.uk/oralevidence/2187/pdf/)

[23] https://www.legislation.gov.uk/ukpga/2003/21/contents

[24] S 368Z1 Communications Act 2003: https://www.legislation.gov.uk/ukpga/2003/21/section/368Z1

42. As regards harm to adults, there are two specific areas where we feel the Bill is (either deliberately or otherwise) unclear and user protections are weakened as a result. Firstly, Section 11 ("Safety duties protecting adults: Category 1 services") states that services have a "duty to specify in the terms of service" how "priority content" and "other content that is harmful to adults" should be "dealt with by the services."[25] We understand that the policy intention here is to ensure that the responsibility for risk assessment, and then setting the "tolerance threshold" for legal but harmful content, sits with companies rather than with the government. However, "dealt with" is a phrase that has no qualitative meaning: it does not state whether it has to be done positively, negatively or by deciding not to do anything about the problem. (There is precedent for the challenge of this type of language e.g. the current case with the Irish Information Commissioner of use of the term "handling" - many cases were deemed to be "handled" by not taking a decision).[26] Contrast the position for the children's safety duty where the obligation is to "mitigate and effectively manage" risks (cl 10(2)).

43. We have no desire to see all platforms having to set exactly the same threshold for speech and behaviour, but it is important to remember that safety duties are not just about moderation and take down. For example, a platform that wanted to adopt a more 'anything goes' approach, might want to ensure *effective* warnings at point of entry or provide their users with tools to self-curate as they adjust to risks within that online environment. It is unclear the extent to which the provisions outlining the effect of the codes (cl 37) (which should reflect the online safety objectives in clause 30) cut down platforms' choice in this context, especially taken against the context of a deficient or wilfully blind risk assessment. This part of the Bill relies upon platforms' enforcement of their own terms of service (as against users). In so doing, it loses close connection with the characteristics of the platform design and operation and their impact on content creation (e.g. through financial or other incentives), information flows and user empowerment (e.g. through usable curation tools) that flows from a systemic approach. By contrast, the illegal content and child safety duties emphasise the importance of these "characteristics".

44. Note also that the designation of types of content as priority content seems to have no impact on the platform's response to that problem – it just means that the platform must "deal with" the topic in its Terms of Service, irrespective of whether that issue had come up in its risk assessment.

45. Secondly, in clause 46, the meaning of "content that is harmful to adults" is set out, using the term that the provider of the service has "*reasonable grounds to believe that the nature of the content is such that there is material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities*". This is a different phrasing to that set out in the Full Government Response ("the legislation will set out that online content and activity should be considered harmful, and therefore in scope of the regime, where it gives rise to *a reasonably foreseeable risk of a significant adverse physical or psychological impact on individuals*"; para 2.2[27]). Does the shift indicate a different threshold is in play? Note

---

[25] We note a change in wording re the enforcement of companies' Terms and Conditions between the draft Bill and the Government response the Inquiry into Covid and Misinformation (https://publications.parliament.uk/pa/cm5801/cmselect/cmcumeds/894/89402.htm). The latter said platforms "will need to enforce these terms effectively, consistently and transparently" (p 2), while in the draft Bill it is just "consistently" (for example, S9, 5 (b)), which could well mean badly or not at all.

[26] https://noyb.eu/en/irish-dpc-handles-9993-gdpr-complaints-without-decision

[27] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response

that the Communications Act already imposes on some user-to-user platforms the obligation to "protecting the general public from videos and audio-visual commercial communications containing relevant harmful material". "Relevant harmful material" includes material containing violence or hatred against a group of persons or a member of a group of persons based on any of the grounds referred to in Article 21 of the Charter of Fundamental Rights of the European Union of 7 December 2000. While some of this material would fall under illegal content, not all the categories are protected by the criminal law. This means that any such types of content would fall to be assessed under 46(3), which might lead to difficulties in the context of lots of low-grade abuse. Arguably, this then constitutes a reduction in the level of protection. The commitments made in the G7 communique about tackling forms of online gendered abuse will be in part delivered by this clause and to set a strong international lead, the clause needs to be made to work.[28]

46. We note that the obligations with regard to harmful but legal, which are weak as regards user-to-user services, do not apply to search at all. This suggests that there is no obligation to do anything about antisemitic auto-completes, for example, or any safeguards around the return of results on suicide and self-harm.

How is the boundary for Category 1 determined?

47. We have noted the distinction between different categories of service and the fact that the criteria for the categories will be set down in regulations. Schedule 4 requires the regulations to contain rules relating to number of users and about functionality[29] (but not business ethos or character of the service). This may mean that is would not be possible to designate a service purely on the basis of its size, or on the basis of its riskiness. Does it mean that a small but very, very risky platform (e.g. chat roulette with video capability) could never be Category 1, or that a super-dominant platform could not be Category 1 by virtue of its size alone?

48. Note that similar types of criteria are used to determine the boundaries of Category 2A and 2B services, though this boundary has significance only for the transparency reporting duty. It is unclear how this system will cope with services that scale rapidly.

What measures are there to protect children?

49. Companies will be obliged to assess whether it is possible that children can access their service and then whether it is likely and that they will only be able to conclude "no" if there are systems and processes in place to prevent children's access. OFCOM will provide guidance to assist platform operators with this (cl 28), but the obligation on the platforms to make this assessment will not kick in until OFCOM has published the guidance. As guidance, this is not enforceable. The requirement to make the assessment and to keep it up to date is enforceable. While there is little prescription on the face of the draft Bill regarding the precise measures required to protect children, we are content that overall approach is systemic. The risk assessment (cl 61(2)(ii)) and Online Safety Objectives (cl 30(2)(a)(vi)) note the different developmental stage of children. We welcome the indication that some measure of age verification or age assurance will

[28] https://www.g7uk.org/wp-content/uploads/2021/06/Carbis-Bay-G7-Summit-Communique-PDF-430KB-25-pages-5.pdf

[29] Functionality is defined at cl 135(1)

be required for platforms to "be entitled to conclude" that it is not possible for children to access a service (cl 26(3)), but further clarification is required here, in particular whether standards for age verification will be set – it is not clear whether the clause 28 guidance would go this far.

50. While the risk assessment specifically flags the risk of adults contacting children, it assumes that contact will happen within the platform and fails to take into account the interaction between platforms: for example, where people intending to abuse or groom children use one platform to flag that activity and another where the activity is carried out. There is no cross-platform duty to collaborate, even on illegal risks such as CSEA and to risk assess on that basis. How will issues around cross-platform harm and cross-platform co-operation be addressed?

How does the draft Bill make up for the repeal of Part 3 of the Digital Economy Act?

51. Children's campaigners are concerned that sites which publish hardcore porn are not caught by the regulations and the requirement for age-verification unless they include user-generated content or other forms of user interactivity. It is highly likely that any porn sites which have either such functionality will remove it to dodge regulation, with no impact on their revenues or business model. This eventuality would negate a core part of Part 3, DEA 2017: addressing the prevalence of porn on social media is necessary, especially as it is where younger children first stumble across it, but the consumption of it occurs on sites like Pornhub. We note that Oliver Dowden told the DCMS Select Committee recently that he had heard these concerns regarding commercial pornography sites and he had an open mind re further measures being considered during PLS.[30]

52. One solution might be to extend the territorial scope of the regulation of On-Demand Programme Services[31]. Currently those provisions do not catch those services which have a significant UK user base, or which target UK users, but which are not established in the UK.

What are the exclusions?

53. We note that a few specific exclusions are listed at Section 46(8) where physical or psychological harm comes from "content's potential financial impact"; "safety and quality of goods featured in the content"; or "the way in which a service featured in the content may be performed". We presume that the concession signalled by the government in relation to online scams (where facilitated by user-generated content)[32] will lead to a revision of the first exemption, which we

[30] Oliver Dowden evidence to DCMS Select Committee hearing 13th May 2021 https://committees.parliament.uk/oralevidence/2185/pdf/

[31] See https://www.ofcom.org.uk/tv-radio-and-on-demand/information-for-industry/on-demand

[32] The Government's press release (11/05/2021) set out that: "Measures to tackle user-generated fraud will be included in the Bill. It will mean online companies will, for the first time, have to take responsibility for tackling fraudulent user-generated content, such as posts on social media, on their platforms. This includes romance scams and fake investment opportunities posted by users on Facebook groups or sent via Snapchat. … **Fraud via advertising, emails or cloned websites will not be in scope because the Bill focuses on harm committed through user-generated content**. The Government is working closely with industry, regulators and consumer groups to consider additional legislative and non-legislative solutions. The Home Office will publish a Fraud Action Plan after the 2021 spending review and the Department for Digital, Culture, Media and Sport will consult on online advertising, including the role it can play in enabling online fraud, later this year." (https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published)

welcome, subject to understanding the extent of that concession and how it will be reflected in the face of the Bill. The government should consider whether the experience of the pandemic suggests that ed-tech should in fact be in the scope of the regime to support teachers and parents in using innovative tools.

54. However, mis/disinformation, wider consumer harms and all other definitions of harmful content and illegal content, unless the Secretary of State designates otherwise, have to result in harm to individual: this leads to the exemption of broad swathes of societal harm that are either the result of co-ordinated campaigns or where the aggregation of individual harms is such that a distinct societal harm occurs. What, for example, would be the position for environmental or climate change related misinformation (bearing in mind the WHO identifies the consequences of climate change as including significant health threats)? The draft Bill envisages that harms need not be the direct consequence of the speech but may be indirect. While it is important that this possibility remains in the regime, some further guidance on how indirect such harms may be would be desirable. In his evidence to the DCMS Select Committee, the Secretary of State Oliver Dowden specifically said, in relation to misinformation where there is an "aggregation of harm", it would have to fit into the Category of "physical or psychological harm" to be in scope[33]. While disinformation on specific issues which lead to physical harm to an individual (e.g. Coronavirus misinformation) might be covered, there is a significant gap in relation to national security and electoral interference.

55. We have set out in detail risk management measures for the high-risk category of people involved in democratic processes – candidates, office-holders and journalists[34]. While the draft Bill describes how protections that already exist for such people as individuals would apply (in relation to criminal and psychologically harmful harassment and abuse) we judge that more work is required given the central importance to the functioning of democracy as set out in the Committee on Standards in Public Life 2017 report on Intimidation in Public Life.[35]

56. Concerns have also been raised about the ring-fencing of content of democratic importance and, relatedly, what the definition is of news content and whether it will include citizen journalism. If one cannot "de-amplify" disinformation, especially that which challenges democratic integrity (e.g. from foreign actors), that is an issue. We do not come to a firm conclusion on all the issues arising from these provisions, but note that it will be important to ensure that they are not open to abuse and that it will be challenging to reflect fully the public interest in journalism and in democratic debate. These provisions will require close scrutiny.

How is national security addressed?

57. The draft Bill does not offer a systemic duty to protect national security but gives the Secretary of State and the regulator powers in the curious form of a 'public statement notice' (clause 112). The public statement notice appears to allow the Secretary of State to ask OFCOM to ask

---

[33] https://committees.parliament.uk/oralevidence/2185/pdf/

[34] Online safety for people involved in the democratic process, 24th March 2021 https://www.carnegieuktrust.org.uk/blog/increased-online-safety-for-people-involved-in-the-democratic-process-in-the-uk/

[35] https://www.gov.uk/government/publications/intimidation-in-public-life-a-review-by-the-committee-on-standards-in-public-life

regulated services to make a statement on how they are complying with a threat set out in the notice to public safety, public health and national security. This feels like an emergency procedure, rather a rolling process of risk assessment based on the UK government's extensive risk assessment processes for these harms. It also does not seem to expect a minimum standard in the operators' respective response. The public statement notice addresses collective harms, and perhaps for this reason is connected to OFCOM's media literacy powers in the Communications Act rather than online safety which is focussed on harm to the individual. There is no indication of how a response to a public safety notice that is inadequate might be enforced.

58. The government said in the December policy document that:

> *Where disinformation and misinformation presents a significant threat to public safety, public health or national security, the regulator will have the power to act.* (para 2.84) [36]

59. While, as we have noted, some aspects of misinformation might be caught with regards to health and public safety, the draft Bill does not deliver a systemic response to that policy intention nor does it fulfil the Prime Minister's subsequent commitment in Parliament to the Chair of the APPG on Technology and National Security.

> *Darren Jones MP: …..can the Prime Minister confirm that the online safety Bill that will be presented to the House this year will contain sufficient powers to tackle collective online harms, including threats to our democracy?*
>
> *Prime Minister: Yes, I can.*[37]

60. The Government's consultation on Legislation to Counter State Threats[38] (contemporaneous with the draft Bill) says that:

> *Disinformation and information operations – increasingly, these have become core tools for state and non-state actors alike to sow discord, attempt to interfere in UK democracy, and disrupt the fabric of UK society through division and polarisation*

61. The USA intelligence reports[39] on Russian interference in the 2020 election demonstrated how social media services are used as attack vectors by the UK's adversaries. Media and analytical reports[40] suggest there may have been a disinformation attack on UK elections.

62. The new Atlantic Charter[41] commits the UK and the USA to:

> *oppose interference through disinformation or other malign influences, including in elections.*

63. The accompanying UK/USA joint statement says that:

---

[36] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response
[37] HC Deb, 16 March 2021, c175 https://www.theyworkforyou.com/debates/?id=2021-03-16a.161.0#g175.0
[38] See https://www.gov.uk/government/consultations/legislation-to-counter-state-threats
[39] Ref 'Foreign Threats to the 2020 U.S. Federal Elections' https://www.dni.gov/index.php/newsroom/reports-publications/reports-publications-2021/item/2192-intelligence-community-assessment-on-foreign-threats-to-the-2020-u-s-federal-elections
[40] FT coverage of report by Graphika https://www.ft.com/content/e190ac7e-154d-11ea-9ee4-11f260415385
[41] Atlantic Charter https://www.gov.uk/government/publications/new-atlantic-charter-and-joint-statement-agreed-by-the-pm-and-president-biden/the-new-atlantic-charter-2021

Building on the U.K. G7 Presidency's focus on open societies, and looking ahead to the U.S.-hosted Summit for Democracy, the U.S. and U.K. will continue to make practical efforts to support open societies and democracy across the globe. We will do this by defending media freedom, advancing a free and open internet, combatting corruption, tackling disinformation, protecting civic space, advancing women's political empowerment, protecting freedom of religion or belief, and promoting human rights of all people.

64. Civil regulation has an important role to play in national security by requiring proper security of infrastructure – for instance physical security of power stations, airports etc. There is a gap between traditional cyber security which defends services and networks themselves (such as the NIS regulations[42]) and defending against a disinformation attack where an adversary exploits the way a service works without having to attack its underlying software. A draft Bill on online safety would be an ideal place to address this vulnerability.

65. The Government suggests that attacks on elections are not within scope, pointing to the Defending Democracy programme[43] and the work of the counter-disinformation unit as their means of addressing this risk, a unit which works without formal oversight. We do not feel that either of these are sufficient for the scale of the threat. Nor do we understand why the draft Online Safety Bill cannot be used to address this clear harm. Instead, a perhaps unintended side effect of clause 13 on democratic speech is that platforms in fact might have to protect political disinformation, even if it has been catalysed by a foreign adversary. We suggest that it should be possible to separate disinformation attacks by state actors or their proxies from lower-level disinformation. The UK has a high-powered process for determining such threats and in the draft Bill has a potentially strong regime for addressing online harm it seems wasteful to separate these out. National security is also a matter that the state has unique competence to assess – the current informal mechanisms for sharing national security information with major social networks are overdue for formalisation.

<u>How are threats to public health and public safety assessed and mitigated?</u>

66. In the UK, the State is responsible for assessing whether a significant threat to public health or public safety exists. Where these pose relevant threats of harm to individuals we think that OFCOM's clause 61 risk analysis should include these within scope. The Bill or the process around the Bill should provide for better transmission from the public sector to companies about such risks, in particular when they arise from disinformation. Assessments of threats to public health are undertaken by the 122 Directors of public health and the country-level public health regimes. Assessments of threats to public safety are undertaken by the 45 territorial police forces and aspects of the security services. The current mechanism for transmitting information about such threats is for the government to ask the social media companies if they could be so kind as to consider the issues through bilateral contacts with the individual threat assessors or sometimes the Counter Disinformation Unit in DCMS. We understand that other channels might also exist. The creation of a regulatory regime allows for this system of co-ordination and

[42] The Security of Network and Information Systems Regulations (SI 2018/506) (NIS Regulations)
[43] See e.g. the Government response to the ISC Russia Report, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/902342/HMG_Russia_Response_web_accessible.pdf

transmission of the outcome of threat assessments to be put on a formal footing and made properly accountable to parliament. This could also reduce a regulatory burden and improve security by having an aggregated single point of contact for regulated services and the regulator. The government has not addressed this in the draft Bill.

67. It is not clear if using social media to incite damage to property or, say, a statue (such as during public disturbance) is covered by the regime. Social media was claimed to have a role in several large outbreaks of public disorder over the last ten years[44]; incitement to more general disturbance which could lead to harm to individuals might be caught. The government should clarify its intentions.

What powers does OFCOM have?

68. We welcome the fact that the draft Bill follows the standard practice of handing over decisions on implementation of legislation to the appointed regulators, though we reiterate our concerns about the powers of the Secretary of State. Although there are significant requirements that OFCOM will need to meet with regard to drafting, consulting on and implementing codes of practice and guidelines, these can be relatively high level, giving companies flexibility to take the steps necessary to meet the regulatory requirements – provided that there is some obligation on them to act reasonably or to take appropriate measures. Much of the groundwork done to date by the ICO "safety by design" guidelines can be built on here, as well as the work OFCOM itself has carried out under its video-sharing platform powers.

69. Given the upfront obligations on OFCOM with regard to bringing forward evidence to support secondary legislation to include specific harms within the regime, there are legitimate concerns re the capability and capacity it has to do this *at pace*, even within the major recruitment drive that it now has underway.

Is there a duty to cooperate?

70. There is a golden opportunity not taken in the draft to set a duty on regulated services to cooperate with other regulators – the police, the ASA, FCA etc. This would be simple to do. We note that there was no substantial answer given on this by Minister Dinenage when asked a similar question by the Lord Stevenson in the Lords Digital and Communications Committee recent evidence section.[45]

Can other regulators be co-designated to work under the regime?

71. The government's final response in December mentioned co-designation – OFCOM being able to give other specialist regulators the ability to work within the regime. This important power allows OFCOM to bring other regulators specialisms into the regime. This doesn't seem to be in the draft Bill and it isn't clear if the power to co-designate in the Communications Act also applies to Online Safety. We understand that the latter is the intention but the government should confirm this.

What is the redress system?

[44] 'The rules of engagement A review of the August 2011 disorders' HMIC 2011
[45] https://committees.parliament.uk/oralevidence/2187/pdf/

72. We note that services providers will be under a duty to provide redress mechanisms, though it seems that these can be limited to inhouse review. We note that video-sharing platforms are currently under an obligation to provide an impartial out-of-court procedure for the resolution of any dispute between a person using the service and the provider and also safeguards the right of the user to bring civil proceedings if that user so choses.[46] The provisions in the draft Bill do not provide this level of detail. We suggest that the requirements introduced into the Bill should not be lower than those that already apply at least to a sub-section of social media platforms.

How is online advertising addressed?

73. The current video sharing platform provisions impose on platforms two sets of obligations – one in relation to content that it controls and one in relation to adverts under the control of others. This implies that some providers should already be putting in place controls around how they control advertising across their platform (without imposing direct liability for content). OFCOM currently has responsibility for this and will continue to do so until the Online Safety Act comes into force, after which it will stop. The ASA self-regulatory regime for the content of ads will then apply. While compliance of the content of the ad with the relevant advertising standards may fall within the ASA's remit, OFCOM should remain backstop regulator and should clearly have responsibility as part of the draft Online Safety Bill with its focus on systems and business models for the ad placement and targeting aspect of adverts. This leads to three points:
    - the regime currently found in the Communications Act imposes stringent rules for ads targeted at children and includes requirements on harmful ads. Clause 39(2) in conjunction with clause 39(7) excludes paid for advertisements from the scope of "regulated content". Will the current standards be maintained following the entry into force of the Online Safety Act? This is important as some forms of advertising could trigger relevant (individual) harm – e.g. sale of skin lightening products. Furthermore, the boundary between ads and non-commercial content in some contexts (e.g. influencers) is fine.
    - It is unclear what will happen to the control element of the regime. Having systems in place as to, e.g. targeting of advertising at children, is part of the systems approach and should be part of this regime.
    - The ASA's ability to impose penalties is weak, especially by comparison with the regime proposed by the draft Bill (e.g. business disruption measures, as well as fines).

What might enable a quick start?

74. Much of OFCOM's research required in the first year of operation is generic and a first run could be started now in mid-2021 rather than waiting until mid-2022. We would urge the Secretary of State to write to OFCOM now to ask it to prepare to receive powers under future Online Safety legislation. This would be similar to the duty on proto-OFCOM established under the original Office of Communications[47] paving act, which had a duty to prepare to receive functions under the Communications Act a few years later. This would then allow OFCOM to show some of its working as parliamentary scrutiny progresses. There is a convention that government can only begin preparatory work for legislation after Second Reading of the relevant Bill. However, OFCOM has broad powers to undertake research, has already undertaken much research on matters relevant to the draft Bill and should not be so constrained.

75. It is possible that by performing some research now OFCOM could remove the need for some delegated powers which complicate the Bill. We are undertaking an analysis of where that might

[46] S 368Z1(7) Communications Act 2003.
[47] https://www.legislation.gov.uk/ukpga/2002/11/contents/enacted

work. Carnegie UK Trust supported Lord McNally's Online Harms Reduction Regulator (Report) Bill, which suggested a similar approach[48].

How does commencement work?

76. We feel that the sequencing of commencement is curious and could benefit from clarification, or indeed a flowchart. We set out in annex A the sequence of events that we think need to take place before meaningful action can be taken under this regime; in summary, once the Bill receives Royal Assent (likely mid-2022, or later) OFCOM has 6-18 months to makes its broadly based survey of the risks of platforms (clause 61), then a further six months to make guidance (clause 62) for regulated services to meet their risk assessments the approval of which is then the appointed day on which it seems the duties in the act come into force. The completion of these two steps then starts the clock for companies to do their risk assessments: they have three months from the day of publication of the risk profile or guidance, whichever is the later. It is not clear when the obligation to take action to respond to the risk assessment comes in either, or whether any duty (for example, taking into account human rights) is independent of the risk assessment.

77. The sequencing of these steps with additional steps required to ensure safeguards for children or to address content harmful to adults is also unclear. Regarding children, companies need to know whether or not they have children using the service: clause 27 says a company has three months from OFCOM's publication of guidance under clause 28. It is not clear whether this is supposed to happen before the OFCOM guidance on the risk assessments or whether the work on risk assessment would be split to get the illegal content guidance, which applies across the board, published first. Should companies who know they've got children using their services want to start thinking about it before they have to?

78. For content harmful to adults, a platform would presumably need to know that it was Category 1 before the obligations to do a risk assessment in relation to this would bite (though the text of the timing for the risk assessment doesn't say this; the risk assessment is specified to apply to Category 1 only). Schedule 4 gives a timeline for this: OFCOM has to do research and report within 6 months (though an extension of up to 18 months can be granted, though there is no oversight of that process in the Schedule) After receiving advice, the Secretary of State is to bring an SI forward "as soon as reasonably practicable" and then OFCOM establishes a register as soon as reasonably practicable thereafter. (There is, of course, the period for the laying of the SI before Parliament but OFCOM could presumably work on its analysis during that period and only publish its conclusions once the SI enters into force, so minimising any delay from that process.)

79. We presume these could run concurrently but this gives OFCOM a lot to do simultaneously. On top of that, there is the process for priority content: the system can work without it but it would be politically astute for it to be addressed quite fast. This again requires the Secretary of State to consult OFCOM – at least as regards priority content for the adults' and children's safety duties (cl 47(2)). None of this is clear in the draft Bill or its supporting documents and the sequencing of several of these events will have implications for the effective functioning of the regime.

80. There are a number of things that could expedite this process, including an instruction to OFCOM from the Secretary of State to start its generic research now and not wait until the second reading of the Bill. Analysis too of which bits of the process are conditional on prior steps

[48] https://www.carnegieuktrust.org.uk/news/online-harms-paving-Bill-introduced/

and which pieces of advice from OFCOM are helpful, but not essential, would also help in defining the critical path. We think that a timeline from DCMS would assist deliberations in the PLS stage, along with confirmation that Parliament can set the Secretary of State a deadline by when to do things.

ANNEX A: an indicative timeline

- Pre-legislative scrutiny (12 weeks, likely to start June 2021 with a gap for recess)
- Bill introduced in the autumn 2021 (passage may take up to 9 months)
- Commencement and Royal Assent outlining definitions and basic powers of OFCOM (earliest summer 2022)
- First big showpiece will be the Secretary of State setting out their priorities (some will come from OFCOM).
- OFCOM set out Section 61 risk assessment which will encompass the big systemic issues, design of platform and content. This may take 6 months but can take up to 18 months. OFCOM will then need to produce Section 62 guidance to providers on how they should carry out risk assessment (a further 6 months).
- Only when guidance on risk assessment is published are platforms obliged to undertake the risk assessments. Providers have 3 months to do the risk assessment, but extra time can be agreed by OFCOM.
- OFCOM also has to produce guidance on assessment of access by children, which will be interlocking but developed as a standalone process.
- Sometime around this time the Secretary of State can bring in regulations via secondary legislation not covered by the rest of the regime.
- Prior to doing that (though exact timings are unclear), OFCOM has to deliver Section 5 – advice on the tiers (Cat 1, 2a, 2b etc).
- OFCOM is very good at delivering robust processes, but may take 6 months to a year to get up and running. The shortest possible path is 9 months (though this feels unfeasible in reality).
- Overall looking at mid-2023, recognising that a number of issues will be contested.

ANNEX B: the role of the Secretary of State

This table sets out the various powers, roles and responsibilities conferred by the draft Online Safety Bill onto the Secretary of State. **Bold text identifies where OFCOM is informing the Secretary of State (SoS)**; the rest are when the Secretary of State (SoS) acts.

Provision	Purpose	Oversight	Impact
3(8)-(12)	Amendment/repeal of provisions providing exemption (Schedule 1); potential to affect cl 39	Condition; risk of harm is low cl 3(8), (10) Cl 132(4): **positive** resolution procedure	Scope
29(5)	**OFCOM to consult SoS (among others) on CoP, but note role of online safety objectives and approval by SoS (cl 32 & 32) (see cl 35 in re amendment)**	Cl 31 specifies things for OFCOM to take into account	Operation
30 (5)	Amendment of "online safety objectives" (leading to OFCOM to review Code of Practice (CoP))	Cl 132(4): **positive** resolution procedure	Operation
32(1)	Approval of CoP – to SoS (subject to cl 33) and then to Parliament	NB no time limits for laying CoP before Parliament, but then 40 days **negative** resolution process (cl 32(3)); CoP to be published within 3 days of issue (cl 34(1))	Operation
33	SoS power of (repeat) direction in re CoP, so as to 'reflect government policy' or 'for reasons of national security or public safety' – OFCOM explain its changes until SoS satisfied, then CoP goes to Parliament	Limitations in 33(2): can't direct particular step, reasons must be given CoP (plus OFCOM's explanations) to be laid 'as soon as reasonably practicable' (cl 33(5))	Operation
35(2)-(4)	**OFCOM notifies SoS of minor amendments to CoP,** SoS to agree	Entry into force 21 days after issue	Operation
Cl 39(12)-(13)	SoS may repeal 39(2)(d)/(5) [comments and reviews]; 39(2)(e)/(6) [oral communications]	Conditions around risk of harm Cl 132(4): **positive** resolution procedure	Scope
Cl 41(4)	Including offence (and constituting "priority illegal content")	See cl 44(1) for things SoS to take into account, cl 44(5) -excluded category cl 44(6). **Approval** of both houses required for first set of regs (Cl 132(6)); thereafter regulation made under ***negative*** resolution procedure	Scope
Cl 42(2)	Amendment of schedule 2 (list of terrorism offences)	Cl 132(4): **positive** resolution procedure	Scope
Cl 43(2)	Amendment of Part 1 or 3 of schedule 3 (list of CSAE offences) [NB Scottish Ministers may amend Part 2 Schedule 3]	Cl 132(4): **positive** resolution procedure Scottish ministers (cl 43(3) – **affirmative resolution** procedure	Scope

		under s 29 Interpretation and Legislative Reform (Scotland)Act	
Cl 45(2)(i) and (ii)	"Priority content harmful to children" & "Primary priority content that is harmful to children"	Cl 47(2): consult OFCOM; OFCOM must review regulations and publish report **Approval** of both houses required for first set of regs (Cl 132(6)); thereafter, regulation made under *negative* resolution procedure	Scope
Cl 46(2)	Specification of content harmful to adults as "priority content that is harmful to adults" (cl 46(9))	Cl 47(2): consult OFCOM; OFCOM must review regulations and publish report **Approval** of both houses required for first set of regs (Cl 132(6)); thereafter, regulation made under *negative* resolution procedure	Scope
Cl 49(6)-(7)	Amendment to transparency report requirements (frequency and/or content)	CL 49(8) consult OFCOM	Operation
Cl 51(5)-(6)	OFCOM's powers of exemption in re notification and fees requires SoS approval		Procedure
Cl 51(8)	Change to revenue thresholds	Consult OFCOM Cl 132(4): **positive** resolution procedure	Procedure
Cl 53	**Setting of threshold figure: OFCOM to notify SoS**, who approves (or not) (see also 85(14))	Consult OFCOM – cl 85(15)- Cl 132(4): **positive** resolution procedure	Procedure
Cl 54	SoS guidance on principles in re fees		Procedure
Cl 57(3)	Extension of period within which OFCOM must respond to SoS statement of Strategic Priorities (see cl 109)		Operation
Cl 59(7)/ Sch 4	Specification of criteria for classification of services as category 1, Category 2A, Category 2B	SoS must make regulations and must cover certain issues; OFCOM to carry out research and provide advice to SoS; SoS to provide statement explaining any difference from OFCOM advice. SI enacted as **negative** resolution	Impact on scope of obligations
Cl 66(5)	Statement of minimum standards of accuracy in re technology notices	Advice from OFCOM	Operation

Cl 69	**OFCOM's report about use of technology – to SoS and then to Parliament**	See cl 102	Procedure
Cl 78	Amendment of s24B Comms Act (formulation of policy) – provision of information to the SoS		
Cl 91(10)	**OFCOM to notify SoS after service restriction order made**		Enforcement/procedure
Cl 93(8)	**OFCOM to notify SoS after access restriction order made**		Enforcement/procedure
CL 97(4)	**OFCOM to consult SoS about enforcement guidance**		Enforcement/procedure
Cl 101(6)	**OFCOM to send report on researchers' access to SoS, then on to Parliament**		Review
Cl 106(3)	Designation of criteria for entity to make super-complaints	Criteria in cl 106(4) – public interest; Consultation of OFCOM and other relevant persons (cl 106(5)) Cl 132(4): **positive** resolution procedure	Enforcement
Cl 107	Procedures for super-complaints	Consultation (107(3)) Regulation made under **negative** resolution procedure	Enforcement
Cl 109	Statement of Strategic Priorities – published in a manner determined by SoS (109(5))	Frequency limitations on amendments; consultation and draft before Parliament (cl 110)	Operation
Cl 111	Directions about establishing advisory Committees	Consult OFCOM	Procedure
Cl 112	Directions where threat to health or safety of the public or to national security in re media literacy powers (Cl 112(2) or to give notice to providers (Cl 112(3)-(5))	SoS to publish reasons	Operation
Cl 113	Guidance to OFCOM about exercise of functions under OSB, under s 1(3) & 11 CA03 both in re online safety	Consult OFCOM; minimum frequency; lay before Parliament (no timings) – no limitations on considerations/nature of guidance – NB reporting on SoS functions (Cl 114)	Operation
Cl 115	Review of Online Safety Bill; media literacy powers in s. 11 CA03	Timings set down and considerations ((3)-(4)); consult OFCOM – report to be laid before Parliament (no time limits)	Review
Cl 140	Commencement		