CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

Internet Harm Reduction

January 2019

An updated proposal by Professor Lorna Woods and William Perrin

1. In February last year, we began publishing a series of blog posts[1] setting out our proposal that a statutory duty of care to take reasonable steps to prevent foreseeable harm from arising to users should be imposed on social media platforms. Drawing on the well-established approach of health and safety legislation, as well as the concept of the "precautionary principle" that has informed UK policy frameworks for a couple of decades in other novel areas, the underlying approach was risk-based: that is, rather than specifying particular regulatory mechanisms, the obligation on relevant companies was to strive towards a specified outcome. The duty of care should be backed up by a regulator, with measuring, reporting and transparency obligations on the company concerned (the 'harm reduction cycle').

2. This was a preliminary view, published at a time when there were almost no worked-up proposals for tackling harms and political, parliamentary and public concerns about the activities of tech companies in causing harms were beginning to coalesce[2]. In May 2018, the UK government announced[3] that it would bring forward proposals for new laws to tackle a wide range of internet harms. The government's announcement triggered a wide range of policy work by other stakeholders[4]. We refined our blog posts in evidence to the House of Lords Communications Committee in April 2018[5]. Now, following a number of discussions with a range of stakeholders in the subsequent months, including a meeting on 14th January 2019 to test our new thinking, we develop some issues that were not addressed in our first proposals and revisit some of our earlier thinking.

1 https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

2 See the Committee on Standards in Public Life report on intimidation in public life (December 2017) https://www.gov.uk/government/publications/intimidation-in-public-life-a-review-by-the-committee-on-standards-in-public-life. Also the current Commons DCMS Committee Inquiry into Fake News had begun to take evidence in Q4 2017.

3 Response to Internet Harms Green Paper consultation (20 May 2018) https://www.gov.uk/government/news/new-laws-to-make-social-media-safer

4 For example: https://www.nspcc.org.uk/what-we-do/campaigns/wild-west-web/; https://doteveryone.org.uk/project/regulating-for-responsible-technology/; https://institute.global/insight/renewing-centre/tony-blairs-foreword-new-deal-big-tech; https://webrootsdemocracy.org/kinder-gentler-politics/

5 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/communications-committee/the-internet-to-regulate-or-not-to-regulate/written/82684.pdf

3. This paper sets out new or revised thinking on:

   The case for regulation
   Scope of services subject to regulation
   Protecting a broader range of users
   Nature of harms
   Codes of practice
   Rights of action
   Interaction with criminal law
   Penalties and sanctions
   Implications for the regulator

The case for regulation

4. The case for regulation strengthened during 2018. There have been many instances of social media practice – whether in relation to content available, platform design or other behaviours – giving rise to concern.

5. In our original proposals, we drew from a wide range of regulatory practice to construct a risk-managed regime for social media. Listening in particular to the inquiries of the DCMS and Lords Communications Select Committees, and talking with 5Rights, NSPCC, Communications Chambers, Doteveryone, OFCOM and DCMS, we have been reminded of the centrality of software to the issues at stake.

6. We have revisited Lawrence Lessig's work from 1999[6]. 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. It is becoming increasingly apparent that it also nudges us towards certain behaviour. 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. Corporate decisions drive what content is displayed to a user. Service providers could choose not to deploy risky services without safeguards[7] or they could develop effective tools to influence risk of harm if they choose to deploy them.

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

6 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" (1999) and "Code: version 2.0" (2006)

7 '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?' (Sean Parker, a Facebook Founder, 2007) 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

8. At a structural level, OFCOM noted there is a problem of regulatory asymmetry when services are competing for the same eyeballs. On one device, users will consume multiple services subject to different regulatory frameworks and users are consequently protected to different degrees within each window. The flip side to this is a concern about level playing fields for competition between different content providers.[8]

Scope of services

9. In our initial outline, we proposed a statutory duty of care be imposed on only the largest social media services. We attempted a definition of these qualifying services; broadly, that a service should:

   - Have a strong two-way or multiway communications component;
   - Display user-generated content publicly or to a large member/user audience or group; and
   - A significant number of users or audience – more than, say, 1,000,000.

   We excluded from scope those services already subject to a regulatory regime – notably the press and broadcast media, insofar as the content produced by these actors is concerned. Our proposals do not displace existing laws – for example, the regulatory regime in relation to advertising overseen by the Advertising Standards Authority (ASA) whose rules are pertinent here (e.g. content and placement of ads, especially as regards children[9]). Applying this test meant that services such as Facebook and Twitter would be included, but also gaming platforms such as Twitch. On reflection some large games that contain messaging services might also fall within the definition.

10. The definition sought to exclude messaging services and search engines. During 2018, we met with a wide range of stakeholders who improved our understanding of the scope of risk and proportionality of the burden on business.

11. The UK government in May 2018 proposed addressing harms across a far broader canvas with:

    'legislation that will cover the full range of online harms, including both harmful and illegal content.'[10]

12. In the light of this, 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 Information Commissioner's Office (ICO) consultation on the Age Appropriate Design Code[11] also took this comprehensive approach to harm reduction for children. Throughout 2018, we took note of the DCMS Select Committee's inquiry into disinformation and

8 https://www.ofcom.org.uk/__data/assets/pdf_file/0022/120991/Addressing-harmful-online-content.pdf

9 See Advertising Standard Authority (ASA) rules on children and age-restricted ads online: https://www.asa.org.uk/resource/children-age-restricted-ads-online.html

10 https://www.gov.uk/government/news/new-laws-to-make-social-media-safer

11 Organisation response to ICO consultation: https://ico.org.uk/about-the-ico/responses-to-the-call-for-evidence-on-the-age-appropriate-design-code/

fake news, the multiplicity of work on ethics in AI and the early stages of implementing the GDPR through the Data Protection Act 2018. We continued throughout 2018 to examine regulation in other sectors.

13. Industry representatives pointed out to us that regulation might weigh heaviest on SMEs and new entrants, possibly inhibiting competition.

14. We now propose some changes to the scope of the duty of care.

1: Removing the de minimis user/customer threshold for duty of care and safety by design for all relevant service providers.

15. 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, we are struck that basic design and resourcing errors in a growth stage have caused substantial problems for larger services[12]. Much of the debate on AI ethics attempts to bake in ethical behaviour at the outset. 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[13].

We propose: broadening the scope of our original proposals to apply to all relevant service providers irrespective of size.

2: Proportionality

16. Some commentators have suggested that applying a duty of care to all providers might discourage innovation and reinforce the dominance of existing market players. While there is some justification in this view, 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.

17. Good regulators do take account of company size and regulation is applied proportionate to business size or capability[14]. 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.' [15]

12 This has been a consistent theme of the DCMS Committee fake news inquiry.

13 See for instance the Food Hygiene regulations 2006: http://www.legislation.gov.uk/uksi/2006/14/contents/made

14 HSE - '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.' http://www.hse.gov.uk/simple-health-safety/

15 The Code does not apply to OFCOM but sets out Government views on good regulation - https://www.gov.uk/government/publications/regulators-code

18. The European Commission also acknowledges this in its proposed directive on online terrorist content which requires 'economic capacity' to be taken into account[16] in deciding the adequacy of a company's response.

19. 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. 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.

20. 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.

21. We note that, in other sectors, 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[17]. Many commentators call for more media literacy training for children. 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 bring the fact that guidance and risk-tested code libraries would be available.

22. 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.

23. 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.

We propose: that in assessing compliance with the statutory duty of care, a regulator should adopt a proportionate approach which takes into account, inter alia, the severity of the harm and the size of risk as well as the size of, and resources available to, a service operator alongside the perceived ability to reasonably foresee the risk/harm suffered.

16 From an EU perspective, the recital 18 to the proposed terrorist content online regulation says "in assessing the effectiveness and proportionality of the measures, competent authorities should take into account relevant parameters including the number of removal orders and referrals issued to the provider, their economic capacity and the impact of its service in disseminating terrorist content (for example, taking into account the number of users in the Union)"

17 For instance Google, Microsoft, Facebook have long worked in hashing of child abuse images, now brokered by IWF (https://www.iwf.org.uk/news/tech-breakthrough-announced-on-20th-anniversary-of-iwfs-first-child-sexual-abuse-imagery). Similar action occurs on terrorism: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

We propose: The regulator should work with industry, civil society, the ICO and the Regulatory Policy Committee[18] to produce a statutory safety by design code and should share best practice. We further suggest that funding should be made available to researchers to understand what sorts of software create which sorts of on-line environments.

We propose: that the regulator should engage with the training and professional development industry to steer them to develop products for SMEs that cover risk management and ethical issues and introduce legal and ethical issues to design considerations.

3: Expanding the definition of services

Messaging

24. We excluded most messaging services from our proposals as we regarded them as private communication. We now consider that some of these services are not necessarily private and also give rise to risks to individuals. We encountered disturbing reports of harms arising in messaging services[19] including in large groups. Although some messaging service providers do carry out pro-active moderation,[20] 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.

25. In the last year, it has become clearer that messaging services have gone beyond small groups supporting existing relationships – familial, friendship or work[21]. 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. 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.

26. 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

18 Regulatory Policy Committee: "provides the government with external, independent scrutiny of new regulatory and deregulatory proposals". https://www.gov.uk/government/organisations/regulatory-policy-committee

19 Kik chat app 'involved in 1,100 child abuse cases' (Angus Crawford BBC News) https://www.bbc.co.uk/news/uk-45568276

20 'A WhatsApp spokesperson tells me that it scans all unencrypted information on its network — basically anything outside of chat threads themselves — including user profile photos, group profile photos and group information.' In: 'WhatsApp has an encrypted child porn problem Facebook fails to provide enough moderators' (Josh Constein, Techcrunch, 20 December 2018) https://techcrunch.com/2018/12/20/whatsapp-pornography/

21 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.

should be a far more responsive and effective notice and remedy process for people in a group who have experienced harm. A risk-managed harm reduction process would lead to different measures to those for traditional social media.

We propose: 'messaging' services that enable large groups or those that enable public groups are qualifying services and fall under our proposed regime. The regulator should work with industry, users and civil society on a specific harm reduction cycle for such messaging services.

Questions: should all multiway communications fall within the regime, with proportionality and a risk-based assessment ensuring that the regime is not too onerous? Or should some small multiway services lie outside the regime? If the latter, how do we define the boundary?

Search engines

27. The government's broad definition of online harms has led to us being asked whether a duty of care regime could apply to general search engines, of the likes of Google. 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[22]) is of particular concern. 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.[23] We also note the disturbing research by Anti-Toxin for Tech Crunch[24] into child abuse imagery on Bing, which Google had prevented returning in 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.

28. 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[25] and impartiality that may not be best considered through the lens of a statutory duty of care. Finally, could a regulator manage search as well as social media?

Question: Having recognised that there are questions surrounding search engines (and perhaps other vehicles for discovery), we have however not had the resources to consider fully whether search can or should come into this regime and do not have a proposal at this time. We would however welcome views on whether the regime should be expanded to cover search engines, and how the statutory duty of care might apply in that context.

---

22 'YouTube, the Great Radicalizer' (Zeynep Tufecki, New York Times March 2018) https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

23 'Google removed "Are Jews evil?" from its auto-complete function in December 2016 (following a series of articles on the Guardian/Observer website)' Community Security Trust https://cst.org.uk/news/blog/2019/01/11/hidden-hate-what-google-searches-tell-us-about-antisemitism-today

24 Microsoft statement: "Clearly these results were unacceptable under our standards and policies and we appreciate TechCrunch making us aware. We acted immediately to remove them, but we also want to prevent any other similar violations in the future. We're focused on learning from this so we can make any other improvements needed." https://techcrunch.com/2019/01/10/unsafe-search/

25 Such as due prominence https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/epg-prominence

Protecting a broader range of users

29. 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 services when they themselves are not customers of those services.

30. 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[26] 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[27].

31. 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.

32. One approach might be that already used by the Health and Safety at Work Act 1974 (HSAW74) to tackle harm to people outside the immediate duty of care. Section 2 of the Act 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.*

33. The connecting factor in the HSAW74 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. We note that the HSAW74 provides a lesser protection to third parties than it does to employees.

26 Twitch corporate blog announcing changes (8 February 2018) https://blog.twitch.tv/twitch-community-guidelines-updates-f2e82d87ae58; '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/)

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

34. We think this serves as a broad model and will give further thought to refining it in this context. The service provider could well not be aware of harm to someone with whom it has no relationship. But equally the service provider might have no route for someone who is not a user to complain. A simple improvement in a service offer to open up complaints to people who are affected might create routes for people who are being harmed to seek a solution.

We propose: extending the scope of the regime to cover harm occurring to people who are not users and will consider whether the HSAW74 approach works in this respect.

Harm

35. Our 2018 proposals set out an approach for companies to determine what was harmful and how to mitigate it in a risk-managed way, working with the regulator and civil society within bounds set by parliament. This remains the best approach – for the regulator to oversee what companies judge to be harm within parameters set by Parliament. However, we have been asked several times what harm is, perhaps unsurprisingly in a lightly-regulated sector.

36. Some discussions have focussed on removal of content that is contrary to the criminal law. While the criminal law may identify types of content that cause significant harm, we re-iterate that 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.

37. During 2018, we have seen thinking on harm take shape through other processes. We list these below. However, we share Baroness Grender's view (in Lord Stevenson's House of Lords short debate[28] on a social media duty of care) that competent regulators have had little difficulty in working out what harm means:

> *'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?'*

38. OFCOM's task[29] 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'.*

28 Debate: https://hansard.parliament.uk/Lords/2018-11-12/debates/DF630121-FFEF-49D5-B812-3ABBE43371FA/SocialMediaServices

29 Indeed, Baroness Grender refers to only one of OFCOM's duties set out in S319 Communications Act 2003 https://www.legislation.gov.uk/ukpga/2003/21/section/319

39. The amendment to the Audio-Visual Media Services Directive[30] was published[31] In November 2018. The Directive 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 identifies harms such as content which may impair the physical, mental or moral development of minors; content inciting violence or containing hate speech; and illegal content e.g. provocation to commit a terrorist offence.

40. In September 2018, OFCOM published[32] with the ICO a joint survey of online harms. 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.

41. We note that in the Irish Republic, Donnchadh Ó Laoghaire TD[33] published a Digital Safety Commissioner Bill[34] to create a Commissioner to:

> *'ensure the oversight and regulation, in accordance with this Act, of a timely and efficient procedure for the take down, that is, removal, by digital service undertakings, of harmful digital communications.'*

42. Further information has emerged on the impact of persuasive design in technology leading to overuse and potential harm to children. The 5 Rights Foundation report, "Disrupted Childhood" (June 2018)[35], listed the following areas where overuse of technology as a result of persuasive design could adversely impact children: anxiety and aggression, diminution in the quality of social interactions, creativity, autonomy, memory, reduced sleep and increased sleep deprivation and reduced educational performance. 5Rights Foundation's new report "Towards an Internet Safety Strategy"[36] looks at both risk and harms and offers a useful itemised list of harms to children from digital media.

43. The statutory duty of care is intended to bite at a systems level, which would include harmful aspects of design. The duty would cover not just harmful persuasive design, but also careless service design that leads to harm. We continue to work closely with 5Rights on this matter.

44. Both the 5Rights report and Baroness O'Neill (speaking in a House of Lords debate) show that harms are not only harms to an individual, but that there are harms to a community or society as a whole.

30 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) [2010] OJ L95/1

31 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU 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) in view of changing market realities [2018] OJ L303/69, Article 28b https://eur-lex.europa.eu/eli/dir/2018/1808/oj

32 Ofcom/ICO internet harm research (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

33 Biography: http://www.sinnfein.ie/donnchadh-o-laoghaire

34 Digital Safety Commissioner Bill text https://www.oireachtas.ie/en/bills/bill/2017/144/?tab=bill-text

35 Report June 2018 https://5rightsfoundation.com/in-action/disrupted-childhood-the-cost-of-persuasive-technology.html

36 "Towards an Internet Safety Strategy" (January 2019) https://5rightsfoundation.com/

Baroness O'Neill said[37]:

> *'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.'*

45. The NSPCC has throughout 2018 highlighted many cases of harms to children that occur online, criminal and not. We are studying the NSPCC work with Herbert Smith on a regulatory regime for social media that draws upon our work a duty of care.

46. Parent Zone has continued to issue balanced, informed news and updates[38] for parents and others of the latest issues affecting children in digital media.

47. Work on harms during 2018 reinforces our view that under this regime Parliament should set out the primary or heads of harms as a non-exclusive list to give the regulator initial direction. Parliament should ask the regulator to proceed on the precautionary principle[39]. The regulator should make an annual report on trends, research and state of harms in the UK much like OFCOMs' market reports and also examine international developments working with other national regulators.

From harms to codes of practice

48. In our original work, we set out a harm reduction cycle of industry measurement of harms and action to reduce them, overseen by the regulator working with civil society. The approach of: agree how to quantify harm, carry out measurement, invest to reduce harm, repeat as necessary is similar to that which one would use to reduce pollution. Providers would be under a transparency obligation, in a format set by the regulator, to ensure an accurate picture of harm reduction was available to the regulator and civic society organisations. We now judge that an output of this cycle would be 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.

We propose: the regulator should have the power to draw up codes of practice with industry and civil society or to approve already existing codes.

37 (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

38 https://parentzone.org.uk/latest

39 As set out for instance by the Inter Departmental Group on Risk Assessment. http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.html

Who can act on harms?

49. We have thought further on whether the statutory duty of care should enable an individual right of action to allow someone to sue a company personally rather than – or in addition to – allowing the regulator to act. The statutory duty of care is rather aimed to be preventative, monitored/enforced by a regulator focussing on systemic issues in companies. This means that some of the difficult questions that arise in the context of an individual tortious action – notably causation and evidence that the harm in an individual case caused the injury – fall away. 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.

50. Given the motivation of our proposal, 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; it may be that other mechanisms (e.g. a form of super-complaint mechanism where nominated advocacy groups can bring a complaint to the regulator about aspects of the regulated services that cause harm[40]) involving a designated organisation could be an appropriate safeguard in the event of a dilatory regulator. We re-iterate that the statutory duty of care would not displace the existing causes of action that individuals may have against users or the service providers.

We propose: there should not be an individual right of action under the statutory duty of care though any existing individual rights under other causes of action should not be displaced

We propose: that a super complaint mechanism be introduced.

Interaction with criminal law

51. We set out a series of 'key harms' in our original work. 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[41] mechanisms in the context of proven criminal material[42], and early warning tools for some categories of crime (e.g. patterns of communication in re grooming).

52. This is important in protecting the victims of crime and preventing the ongoing commission of crimes (through continuing to distribute criminal content) especially as 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 bolster the measures set out in the Law Commission's review.[43]

40 What are super complaints: https://www.gov.uk/government/publications/what-are-super-complaints/what-are-super-complaints

41 Stay down mechanisms: where technical measures are taken to ensure that a piece of illegal material is not simply reposted after having been taken down.

42 This is not to suggest that take-down mechanisms should not be available in the context of civil claims.

43 https://www.lawcom.gov.uk/abusive-and-offensive-online-communications/

53. 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.

54. 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.

Penalties and Sanctions

55. In our original work, we set out a range of potential penalties and sanctions for non-compliance[44], ranging from light touch interventions to significant fines, but still encounter scepticism about whether a regime can be 'made to bite' on some of the world's biggest companies. We wonder though if that scepticism is wholly justified – where there is regulatory enforcement or the credible threat of such, companies do by and large comply. It is possible that scepticism arises in part due to the design of existing regulation lagging behind public expectations. We note that some large service providers are themselves calling for regulation. 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 CNIL decision[45] against Google in France will be an early indicator of the effectiveness of the GDPR regime in modifying corporate behaviour.

56. We have considered some options for stronger penalties and enforcement mechanisms that sit between fining companies, the largest of which have tens of billions of dollars cash at bank, and the extreme penalty of S23 of the Digital Economy Act which effectively cuts off a service in the UK.

57. We examine corporate responsibility and director responsibility.

Corporate responsibility

58. The UK government has explored new models to get laws to bite on large companies since the Fraud Act 2006, including – in particular – the Bribery Act 2010 which creates a strict liability. The most recent approach is the Corporate Criminal Offences[46] (CCO) set out in the Criminal Finance Act 2017 (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.

44 See second half of this blog post https://www.carnegieuktrust.org.uk/blog/social-media-harm-regulator-work/

45 Google fined 50m euro: https://www.legifrance.gouv.fr/affichCnil.do?oldAction=rechExpCnil&id=CNILTEXT000038032552&fastReqId=2103387945&fastPos=1

46 What are CCOs https://www.gov.uk/government/publications/corporate-offences-for-failing-to-prevent-criminal-facilitation-of-tax-evasion

59. For large companies, this means they have to make a risk assessment and follow a process that sounds much like enacting 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. In our view, this approach, although largely untested, should drive systems-level compliance.

60. The CCO does not solve the problem of levying an enforceable fine which has sufficient deterrent effect for the largest companies. 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. We note that emerging understanding of the new, somewhat blunt instrument of FOSTA-SESTA in the USA creates a criminal offence for some online services, but also we understand from media reports that it carries a jail term for directors. We do not think that this is the case for the CCOs.

Director responsibility

61. Observation suggests that charismatic founders in tech companies continue to involve themselves in service design and detail long after the company reaches significant size.

62. 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'[47]. The Financial Conduct Authority says:

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

In the extreme case of a financial institution failing, senior managers could be charged with a criminal offence.

63. Could such a regime work with the service providers covered by the statutory duty of care regime? We have strong reservations about the power of a state to arrest a director of a social media service because of our concerns about freedom of expression – any interference with speech should be proportionate. Given that the impact on a company director could be felt across a platform as a whole risking collateral censorship, it could therefore only be justified in extreme cases. With a different penalty though, such as a personal fine, would change be driven by giving directors specific responsibility and standards of conduct as in the financial services sector? Many types of fines, however, are routinely insured against[48]. It might be difficult though to apply to companies not established in the UK without a licensing regime (which we are not proposing).

47 Senior Managers and Certification Regime https://www.fca.org.uk/firms/senior-managers-certification-regime

48 Directors' and officers' insurance is commonplace see https://www.hiscox.co.uk/business-insurance/directors-and-officers-insurance

64. The regime we discuss here focuses on harm. In many countries, responsibility for health and safety is a director responsibility or falls to a nominated senior officer. The Health and Safety at Work Act 1974 has an uncompromising approach to directors' liability when set against its duties of care.

> *37 Offences by bodies corporate.*
>
> *(1) Where an offence under any of the relevant statutory provisions committed by a body corporate is proved to have been committed with the consent or connivance of, or to have been attributable to any neglect on the part of, any director, manager, secretary or other similar officer of the body corporate or a person who was purporting to act in any such capacity, he as well as the body corporate shall be guilty of that offence and shall be liable to be proceeded against and punished accordingly.*

65. The HSE prosecuted 46 directors in 2015/16. We have noted our reservations about imprisonment above.

66. The Company Directors Disqualification Act 1986 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[49] 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.

67. More generally, the Government amended the Privacy and Electronic Communications (EC Directive) Regulations 2003, 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.

68. 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.

49 https://www.gov.uk/government/news/nuisance-marketing-calls-lands-company-director-6-year-ban

We propose: our preliminary view is that directors should be liable to fines personally but we intend to continue our work on sanctions and penalties to reflect on whether any of the other mechanisms discussed would offer additional strategies to drive compliance. We welcome views on this point.

Implications for the regulator

69. We have expanded the scope of our proposed regime which will have implications for the regulator. Despite this broader scope, the regulator would be dealing with far fewer companies than the ICO or the HSE and it would still be able to carry out its task.

70. If the regulator has new responsibilities, it will require more resources. Resources should still be provided on a "polluter pays" basis – either from the proposed internet services tax or from a new industry levy.

71. The regulator would need to prioritise its work based on risk management. In setting priorities the regulator would consult civil society, industry the public and parliament, much as existing regulators do. The regulator would not be able, nor would it wish to implement the entire regime from day one but would require a judicious phasing based on risk.

Next steps

72. We shall continue to discuss these issues with a wide range of stakeholders and welcome feedback on these proposals to comms@carnegieuk.org We shall produce a synthesis of the above and our original proposal for ease of reference.

Professor Lorna Woods
William Perrin
Maeve Walsh

Carnegie United Kingdom Trust
Incorporated by Royal Charter 1917
Registered Charity No: SC 012799 operating in the UK
Registered Charity No: 20142957 operating in Ireland