CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

Submission to Science and Technology Committee Inquiry

November 2018

"Impact of social media and screen use on young people's health"

Background

1. Lorna Woods (Professor of Internet Law, Essex University) and William Perrin (Trustee of Carnegie UK Trust) have been working with Carnegie UK Trust (CUKT) to design a regulatory system to reduce harm on social media. The proposals have been published via a series of blog posts[1] and in detailed evidence[2] submitted to the ongoing Lords Communications Committee Inquiry ("The Internet: to regulate or not to regulate?").

2. We have vast experience in regulation and free speech issues. William has worked on technology policy since the 1990s, was a driving force behind the creation of OFCOM and worked on regulatory regimes in many economic and social sectors while working in the UK government's Cabinet Office. He ran a tech start up and is now a trustee of several charities. Lorna is Professor of Internet Law at University of Essex, an EU national expert on regulation in the TMT sector, and was a solicitor in private practice specialising in telecoms, media and technology law.

3. Our Carnegie work was catalysed by the harms set out in the government's Green Paper[3] and much reporting of harms by interest groups. Our background in regulatory policy gave us confidence that solutions existed that had not received popular attention. We published our work just before the government's May 2018 announcement that they would bring forward a White Paper (now expected in early 2019) that will:

   'set out plans for upcoming legislation that will cover the full range of online harms, including both harmful and illegal content. Potential areas where the Government will legislate include the social media code of practice, transparency reporting and online advertising."[4]

4. This submission primarily responds to the final three questions posed by the Committee's call for evidence:

   - What monitoring is needed, and by whom;
   - What measures, controls or regulation are needed;

1 "Harm Reduction in Social Media" (CUKT): https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

2 Written submission here: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/communications-committee/the-internet-to-regulate-or-not-to-regulate/written/82684.html

3 Internet Safety Strategy (11 October 2017): https://www.gov.uk/government/consultations/internet-safety-strategy-green-pape

4 18 September 2018 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf

- Where responsibility and accountability should lie for such measures.

5. We believe that our proposal for a "duty of care" provides a flexible, future-proofed and preventative approach in these three areas, protecting young people from emotional and psychological harms related social media use.

6. Before setting out our proposal in more detail, we will first look at the challenges around evidence in this area and how the "precautionary principle" should apply to consideration of policy and regulatory action in this area.

Evidence and a basis for action

7. The Committee asked the following questions in relation to the evidence of harms to young people's mental health from social media use and screen time.

- What evidence there is on the effects of social media and screen-use on young people's physical and mental well-being — for better and for worse — and any gaps in the evidence;
- The areas that should be the focus of any further research needed, and why;
- The well-being benefits from social media usage, including for example any apps that provide mental-health benefits to users;
- The physical/mental harms from social media use and screen-use, including: safety online risks, the extent of any addictive behaviour, and aspects of social media/apps which magnify such addictive behaviour.

8. There is a growing body of survey material and qualitative research reports examining the impact of social media on the mental health of children and young people – the Committee refer to a number of these sources in their Terms of Reference[5] – along with research into the potentially harmful cognitive effects, such as distraction and lack of concentration, arising from heavy social media use and "multitasking"[6].
We also note and welcome the two new pieces of research commissioned by the Department of Health and Social Care in this area too: the Chief Medical Officer's systematic review of international research in this area and NHS Digital's prevalence survey of children and young people's mental health.[7]

9. While such studies often demonstrate a correlation between excessive screen time and negative mental health outcomes in children and adults, proving causation is much more difficult, leading to a degree of caution in terms of proposed interventions and calls for more research.[8] Indeed, as the Committee's ToR also notes, other studies have found no harm to children's health from the use of various screens and there are undoubtedly positive impacts for many young people from the use of social media, which – as the Royal Society of Public Health have argued – need to be promoted and optimized at the same time as the significant risks and negative impacts are mitigated.[9]

5 https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2017/social-media--young-peoples-health-inquiry-launch-17-19/

6 http://sciencenordic.com/why-looking-social-media-work-disrupts-your-concentration; https://journalistsresource.org/studies/society/social-media/multitasking-social-media-distraction-what-does-research-say

7 Referred to in DCMS evidence to the Science and Technology Committee inquiry: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/impact-of-social-media-and-screenuse-on-young-peoples-health/written/81892.html

8 Centre for Mental Health: "Social Media, Young People and Mental Health (September 2018) https://www.centreformentalhealth.org.uk/sites/default/files/2018-09/CentreforMentalHealth_Briefing_53_Social_Media.pdf;

9 Royal Society of Public Health: #StatusofMind (May 2017): https://www.rsph.org.uk/our-work/campaigns/status-of-mind.html

10. However, at the other end of the spectrum, the explosion of social media use amongst young people has corresponded with evidence over a similar period of an increase in self-harm and suicidal behaviour. For example:

    - Between 2011 and 2017, a 68% rise in rates of self-harm was recorded among girls aged 13 to 16; [10] and
    - A US study in 2017 found suicide rates for teens rose steadily between 2010 and 2015 after they had declined for nearly two decades[11].

11. Looking more broadly at online harms, the recent OFCOM/ICO research paper[12] ('Internet users' experience of harm online') establishes an evidence base of harms independent of lobby groups. OFCOM's sample size of 1,600 is an order of magnitude better than most extant research but this is short of large-scale multi-annual randomised control trials, which are very difficult to secure in an area of innovative technology that suffers from waves of fashion in its user base. In addition, there is likely to be a gap between users' reported experience of harms experienced on social media and what they actually experience (or are willing to admit to experiencing) online. So, is there then sufficient evidence to act?

12. The traditional approach of not regulating innovative technologies needs to be balanced with acting where there is good evidence of harm but there has not been enough time to establish indisputable evidence of the existence of harm and its causation. We see this as a core challenge for establishing and operating a new regulatory regime.

13. A well-established approach to assessing the desirability of regulation in the face of a plausible but still uncertain risk of harm is the precautionary principle. In the UK, after the many public health and science controversies of the 1990s, the government's Interdepartmental Liaison Group on Risk Assessment (ILGRA) published its version of the precautionary principle aimed at decision makers.[13]

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

14. The ILGRA document advises regulators on how to act when early evidence of harm to the public is apparent, but before unequivocal scientific advice has had time to emerge, with a particular focus on novel harms. The ILGRA's work is still current and hosted by the Health and Safety Executive (HSE), underpinning risk-based regulation of the sort we propose.

15. While the Government has put forward a draft Code of Practice[14] for social media companies, as required under the Digital Economy Act 2017, we believe that such a voluntary Code is now no longer sufficient

10 https://www.nhs.uk/news/mental-health/worrying-rise-reports-self-harm-among-teenage-girls-uk/#where-does-the-study-come-from; ISER study: https://doi.org/10.1186/s12889-018-5220-4

11 https://nypost.com/2017/11/14/rise-in-teen-suicide-connected-to-social-media-popularity-study/)

12 See https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm-online

13 http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm

14 Annex B: Government Response to the Internet Safety Strategy Green Paper https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf

on its own to pre-empt and reduce the current level of harms that can be experienced by users of social media. We would, however, envisage a role for industry-generated best practice on standards or codes as part of our co-regulatory model to deliver the duty of care.

A duty of care

16. We believe that our "duty of care" proposal applies as much to the reduction of potential mental health harms for young people associated with social media use and screen time as it does to risks relating to abusive or illegal behaviour online, from bullying through to hate speech and intimidation. The high-level details below cover our answers to the Committee's questions around: monitoring; measures, control and regulation; and responsibility and accountability. Further detail is set out at the references above and we would be delighted to provide further information to the Committee, either in writing or as oral evidence.

17. Social media platforms are forms of public spaces. People go to such platforms for all sorts of activities and, while using them, should be protected from reasonably foreseeable harm as they would expect in any public place, such as an office, bar or theme park. While some places are subject to specific regimes (e.g. pubs), other rules apply more generally, for example the Occupiers Liability Act and the Health and Safety at Work Act 1974 each of which impose a statutory duty of care.[15] This concept is straightforward in principle and well-established. A person (including companies) under a duty of care must take care in relation to a particular activity as it affects particular people or things. If that person does not take care, and someone comes to a harm identified in the relevant regime as a result, there are legal consequences, primarily through a regulatory scheme but also with the option of personal legal redress. We propose imposing a statutory duty of care on social media platform providers.

18. Applying the "duty of care" approach to the social media sphere has a number of significant benefits:

- It is simple, broadly-based and largely future-proof – expressed in terms of outcome (the prevention of harm) not specifics of process.[16]
- The regulatory approach is essentially preventative, reducing adverse impact on users before it happens, rather than a system aimed at compensation/redress
- The categories of harm can be specified at a high level, by Parliament, in statute.[17]
- It would apply to social media service providers accessible in the UK, with a proportionate approach depending on levels of risk (eg vulnerability of users) and size of company. Online services from traditional media companies would be out of scope.
- A risk-based regulatory approach provides for safe system design (including operational and business choices)
- It is compatible with EU law including the eCommerce Directive and minimises collateral damage to freedom of speech.
- In micro economic terms returns external costs to the production decision and is efficient if applied in a manner proportionate to risk of harm.

15 Occupiers Liability Act 1957 S2 http://www.legislation.gov.uk/ukpga/Eliz2/5-6/31/section/2

16 The government recently confirmed that the 1974 duty of care in the Health and Safety at Work Act applies to artificial intelligence software employed in the workplace. https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Lords/2018-05-23/HL8200/

17 See Health and Safety at Work Act 1974 S2: http://www.legislation.gov.uk/ukpga/1974/37/section/2

How would it work?

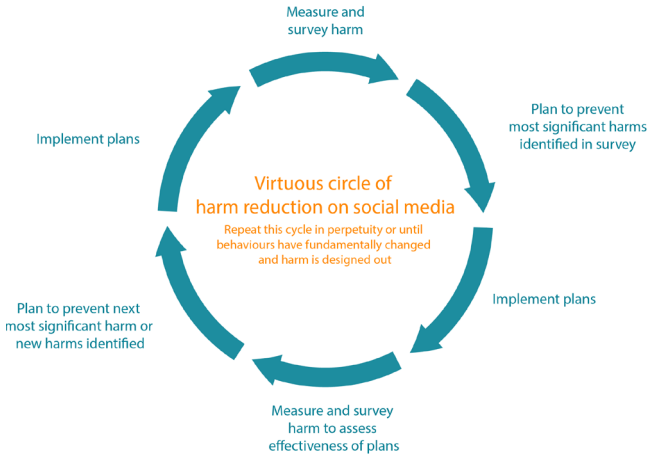
19. New legislation would set out the duty of care and identify the key harms Parliament wants the regulator to focus on. We suggest that those harms would be: the 'stirring up of hatred offences', national security, harms to children, emotional harm, harms to the judicial and electoral processes, economic harms. While these categories would be established in statute, work would need to be done on the scope of each of the harm. It could be that this work would be delegated by the act to an independent regulator which would operate in a transparent and evidence-based manner.

20. We envisage the tasks of such a regulator to be:

- provide guidance on the meaning of harms;
- support best practice (including by recognising good practice in industry codes);
- gather evidence;
- encourage media literacy;
- monitor compliance; and
- take enforcement action where necessary.

More detail can be found in the annex to this evidence.

21 We suggest that the regulator runs a harm reduction cycle,[18] as set out in the diagram below and, in more detail, in the annex, involving civil society as well as companies at each consultative step. The regulator would begin by requiring companies to measure and survey harm, produce plans to address these harms for public consultation and agreement with the regulator then the companies implement the plans. If the cycle does not reduce harms or the companies do not co-operate then sanctions could be deployed.



22. To aid the Committee's understanding of our proposal, with particular regard to its questions around monitoring, controls and responsibility, we have included further detail in the annex on how the harm reduction cycle would work in practice.

Who should regulate?

23. The government, OFCOM, the ICO and countless lobby groups have described serious harms occurring now apparently at a population scale. We can observe new networks being created at a rapid rate; many quickly generating new types of harms and threats, particularly to children and young people. [19] Barriers to entry are low, capital plentiful, technology costs falling still and incentives for self-regulation are non-existent. In considering structural regulatory options, weight should be giving to doing things quickly.

18 Detailed blog post on harm reduction cycle: https://www.carnegieuk-trust.org.uk/blog/social-media-harm-regulator-work/

19 "Children Blackmailed for sexual images in online video chats": https://www.bbc.co.uk/news/uk-45869178; "Kik chat app 'involved in 1,100 child abuse cases'": https://www.bbc.co.uk/news/uk-45568276

24. The duty of care proposal is based on a well-understood regulatory approach that could be legislated and deployed quickly. We see no need for a new regulator to implement it. A new super regulator will be complex to legislate and deploy.[20] In a turbulent climate, where Parliamentary time is focused on Brexit and political distractions are rife, it is likely impossible to achieve in anything less than five years.

25. Unlike the thousands of clauses required for a "from-scratch" regulatory regime with a new regulator, we speculate that a short Bill (overall length of six sections, c30 clauses) could do it. Such a vehicle could slot into the legislative timetable, even though it is crowded by Brexit legislation. We are considering producing such a draft Bill ourselves this autumn.

26. We argue in our detailed proposals that Ofcom's independence, expertise and size make it the ideal body to take on oversight of the new regulatory regime within a short timescale. This is a first-best route to quick effective action.

27. Harm reduction through a duty of care can certainly be delivered through a new regulator but would only be effective in the longer term and a second-best option.

From Professor Lorna Woods and William Perrin

ANNEX: THE HARM REDUCTION CYCLE

We envisage an evidence-based harm reduction cycle in which the regulator would work with the industry to create an on-going process that is transparent, proportionate, measurable and risk-based. It might look something like this:

i) Measurement of harms: the regulator would draw up a template covering scope, quantity and impact, using as a minimum the harms set out in statute. The service provider works with the regulator, consulting civil society on the template, and then surveying the extent and occurrence of harms, as set out by Parliament, in respect of the services provided by that provider.

ii) Service provider action: each service provider then runs a measurement of harm based on that template and produces and implements a plan to reduce them. The regulator would have powers in law to require the qualifying companies to comply. The companies would be required to publish for consultation their survey results and plans of action in a timely manner, establishing a first baseline of harm, and including details on:

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

iii) Re-measurement and assessment: periodically, the harms are re-measured, the effectiveness of the plan assessed and, if necessary, further changes to company

20 William Perrin (as a DTI civil servant in 2001) devised the paving Bill approach that created OFCOM: OFCOM was proposed in the Communications White Paper in December 2000, created in a paving act in 2002 but did not vest and become operational until December 29 2003 at a cost of £120m (2018 prices).

practices and to tools available to users introduced. The re-assessment process would provide the first progress baseline and would show four likely outcomes; that harms:

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

If harms surveyed in the baseline have risen or stayed the same, the companies concerned will be required to act and plan again; the regulator may also take the view that the Duty of Care is not being satisfied and, ultimately, may take enforcement action (see below). If harms have fallen then companies will reinforce this positive downward trajectory in a new plan.

The cycle then repeats, with harms measured and new plans produced by the service providers, while the regulator monitors progress towards overall harm reduction, taking action where necessary. It is important to emphasise that we do not envisage the harm reduction processes to necessarily involve take-down processes. Moreover, we do not envisage that a system that relied purely on user notification of problematic content or behaviour and after the event responses would be taking sufficient steps. Tools/ techniques that could be developed and deployed include:

- the development of a statement of risks of harm, prominently displayed to all users when the regime is introduced and thereafter to new users;
- an internal review system for risk assessment of new services prior to their deployment (so that the risk is addressed prior to launch);
- the provision of a child protection and parental control approach, including age verification, (subject to the regulator's approval/ adherence with standards);
- the display of a rating of harm agreed with the regulator on the most prominent screen seen by users (this is separate from any user-driven ratings system in relation to individual items of user-generated content, e.g. www.yourateit.eu;
- development – in conjunction with the regulator and civil society – of model standards of care in high risk areas such as suicide, self-harm, anorexia, hate crime etc; and
- provision of adequate complaints handling systems with independently assessed customer satisfaction targets and also produce a twice-yearly report on the breakdown of complaints (subject, satisfaction, numbers, handled by humans, handled in automated method etc.) to a standard set by the regulator.

The regulator would also:

- publish model policies on user sanctions for harmful behaviour, sharing research from the companies and independent research;
- set standards for and monitoring of response time to queries (as the European Commission does on extremist content through mystery shopping);
- co-ordinate with the qualifying companies on training and awareness for the companies' staff on harms;
- contact social media service companies that do not qualify for this regime to see if regulated problems move elsewhere and to spread good practice;
- publish a forward-look at non-qualifying social media services brought to the regulator's attention that might qualify in future;
- support research into online harms – both funding its own research and co-ordinating work of others;

- establish a reference/advisory panel to provide external advice to the regulator – the panel might comprise civil society groups, people who have been victims of harm, free speech groups; and
- maintain an independent appeals panel (in relation to complaints about the companies' responses rather than being a first port of call for content disputes).

Sanctions and compliance

Some of the qualifying social media services will be amongst the world's biggest companies. In our view the companies will want to take part in an effective harm reduction regime and comply with the law. The companies' duty is to their shareholders – in many ways they require regulation to make serious adjustments to their business for the benefit of wider society. The scale at which these companies operate means that a proportionate sanctions regime is required.

Throughout discussion of sanctions, there is a tension with freedom of speech. The companies are substantial vectors for free speech, although by no means exclusive ones. The state and its actors must take great care not to be seen to be penalising free speech unless the action of that speech infringes the rights of others not to be harmed or to speak themselves. The sanctions regime should penalise bad processes or systems that lead to harm and all processes leading to the imposition of sanctions should be transparent and subject to a civil standard of proof.

Sanctions would include:

- Administrative fines in line with the Data Protection Act regime of up to €20 million, or 4% annual global turnover – whichever is higher.
- Enforcement notices – (as used in data protection, health and safety) – in extreme circumstances a notice to a company to stop it doing something.
- Enforceable undertakings.
- Adverse publicity orders – the company is required to display a message on its screen most visible to all users detailing its offence.
- Forms of restorative justice – where victims sit down with company directors and tell their stories face to face.

Sanctions for exceptional harm

The scale at which some of the qualifying social media services operate is such that there is the potential for exceptional harm, where activity on its platforms (potentially as a result of design flaws that the regulator may have flagged) has, for example, provoked a riot or resulted in sexual harm to hundreds of young people.

In extreme cases, should there be a power to send a social media services company director to prison or to turn off the service? Regulation of health and safety in the UK allows the regulator in circumstances which often involve a death or repeated, persistent breaches to seek a custodial sentence for a director. The Digital Economy Act contains power for the age verification regulator to issue a notice to internet service providers to block a website in the UK.

None of these powers sit well with the protection of free speech on what are generalist platforms – withdrawing the whole service due to harmful behaviour in one corner of it deprives innocent users of their speech on the platform. However, the scale of social media services mean that acute large-scale harm can arise that would be penalised with gaol elsewhere in society. Further debate on this aspect is needed.