CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

House of Lords Communications Committee Inquiry. Response from Professor of Internet Law Lorna Woods, University of Essex and William Perrin – written evidence (IRN0047)

'The Internet: to regulate or not to regulate?'

Professor Woods gave oral evidence on 24 April 2018 to the Lords Communications Committee Inquiry 'The Internet: to regulate or not to regulate?'. During the evidence session Professor Woods touched upon work she was doing with William Perrin and Carnegie UK Trust on designing a regulatory system to reduce harm on social media. The Chair asked Professor Woods for a note about this work which follows in the form of a summary and a full first draft of our work prepared for the Committee

Summary

1) Professor of Internet Law Lorna Woods and William Perrin have made a proposal to Carnegie UK Trust (Carnegie Proposal) for a regime to reduce harm from social media services as a sub-set of internet intermediaries.

2) Social media service providers are not un-regulatable. We have faced far bigger and more profound issues before and have evolved a huge range of tools to correct corporate behaviours in the public interest. It has been policy since at least the 2000's, both at national and international level, that internet issues should be tackled wherever possible using 'physical world techniques' and social media is no exception.

3) Social media service providers should each be seen as responsible for a public space, much as property owners or operators are in the physical world. In the physical world, Parliament has long imposed statutory duties of care upon property owners or occupiers in respect of people using their places, as well as on employers in respect of their employees. A duty of care is simple, broadly based and largely future-proof. It focusses on the objective and leaves the detail of the means to those best placed to come up with context-appropriate solutions – those who are subject to the duty of care. We suggest this model for the largest social media service providers – a duty of care in respect of their users, enforced in a risk-based manner by a regulator. The duty of care would not apply to online services with their own detailed rules such as the traditional media.

4) A statutory duty of care to mitigate against certain harms be imposed on social media service providers with over 1,000,000 users/members/viewers in the UK in respect of their users/members. These categories of harm are to be specified in statute at a high level of generality. Those under a duty of care would be expected to identify the level of specified harms occurring through set-up and/or use of their respective platforms and take steps to reduce the level of harm, as set out below. This process would be monitored by an independent regulator. The regulator would be appointed and

funded by a share of the revenue from the tax on internet company revenues that the government seems about to introduce.

5) Central to the duty of care is the idea of risk. If a service provider targets or is used by a vulnerable group of users (e.g. children), its duty of care is greater and it should have more safeguard mechanisms in place than a service which is, for example, aimed at adults and has community rules agreed by the users themselves (not imposed as part of ToS by the provider) to allow robust or even aggressive communications.

6) We envisage the harm reduction cycle to look something like this:

   a) Each service provider works with the regulator, consulting civil society, to survey the extent and occurrence of harms, as set out by Parliament, in respect of the services provided by that provider;

   b) Each service provider then produces and implements a plan to reduce the harms, having consulted the regulator and civil society;

   c) Periodically, the harms are re-measured, the effectiveness of the plan assessed and, if necessary, further changes to company practices and to tools available to users introduced;

   d) after a period the harms are measured again as above, new plans are produced and the cycle repeats;

   e) progress towards harm reduction is monitored by the regulator, which may take regulatory action if progress is in the regulator’s view insufficient.

7) Action that a provider could take is not just about take down notices but could include:

   a) measures to empower users, for example pre-emptive blocking tools in the hands of the user; setting up sub-groups that have different toleration of certain types of language

   b) effective complaints mechanisms both in respect of other users but also the company itself

   c) transparency measures so that it is possible to see the number of complaints, the response, the mechanism by which the complaint was processed (human or automated) and the reasoning

   d) review systems of company processes that assess them for nudging users to certain sorts of behaviours.

8) The regulator would have the following responsibilities:

   a) producing, through a consultative process, a list of the qualifying social service providers with more than 1,000,000 users/members etc in the UK. Being on or off that list is challengeable by judicial review.

   b) monitoring the harm reduction processes run by the companies and supervises them into a continuous harm reduction cycle.

   c) Providing advice as to the scope of the harms, best practice on harm reduction;

   d) enforcing the duty of care using tools such as enforcement notices, prohibition notices and fines.

9) The list of qualifying social media service providers would likely include (but not necessarily be limited to):

a) Facebook
b) Twitter
c) YouTube
d) Instagram
e) Twitch
f) Snapchat
g) Musical.ly
h) Reddit
i) Pinterest
j) LinkedIn

10) The regulator would have a range of sanctions from adverse behaviour notices through to administrative fines on the scale of those found in the GDPR. Individuals may be able to bring court action but we emphasise that this should only be in respect of systemic failures and not as a substitute for a civil action in relation to specific items of content.

Notes on the summary

11) In our opinion this is compatible with EU law, in particular the e-Commerce directive. The immunity provisions relate to liability for the content of others and do not absolve providers from any duties of care.

12) The preventive element of duty of care will reduce the suffering of victims. It may also prevent behaviours reaching a criminal threshold.

13) A risk-managed approach only targeting the largest providers preserves freedom of speech. We envisage that platforms may take different approaches, and that a market could arise in which platforms develop aimed at particular groups. Content or speech patterns that are not acceptable on one platform may find a home elsewhere.

14) Harms represent external costs generated by the production of the social media service providers’ products. The duty of care, by requiring action to prevent harms internalises these costs to the provider. This makes the market function more efficiently for society on the polluter pays principle and ultimately drives a more effective market which also benefits providers.

About the authors

15) William Perrin and Lorna Woods 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.

16) William and Lorna approached Carnegie UK Trust in January 2018 with a proposal to undertake this work pro bono. Carnegie has a strong track record in public policy as well as technology expertise as part of its Digital Futures programme and wider work on national wellbeing. Carnegie has been publishing blog posts as drafts of a final report which will be published in the Summer.

17) The views expressed here are of the authors and not any other body.

Survey of regulatory regimes

Harms and market failure

18) The Government's Internet Safety Strategy Green Paper detailed extensive harms with costs to society and individuals resulting from people's consumption of social media services. Social media services companies early stage growth models and service design decisions appear to have been predicated on such costs being external to their own production decision. Effective regulation would internalise these costs for the largest operators and lead to more efficient outcomes for society.

19) There is a good case to make for market failure in social media services – at a basic level people do not comprehend the price they are paying to use a social media service – recent research by doteveryone[1] revealed that 70% of people 'don't realise free apps make money from data', and 62% 'don't realise social media make money from data'. Without basic awareness of price and value amongst consumers it will be hard for a market to operate efficiently, if at all. It would be interesting to see a full analysis of market failure in the sector.

Relevant regimes

20) Assuming that some sort of regulation (or self or co regulation) is necessary to reduce harm, what form should it take? We surveyed regulatory regimes for communications, the digital economy, health and safety and the environment.[2]

21) There are many similarities between the regimes we surveyed. One key element of many of the regulators' approach is that changes in policy take place in a transparent manner and after consultation with a range of stakeholders. Further, all have some form of oversight and enforcement – including criminal penalties - and the regulators responsible are independent from both Parliament and industry. Breach of statutory duty may also lead to civil action. These matters of standards and of redress are not left purely to the industry.

22) While the telecommunications model may seem an appropriate model give the telecommunications sector's closeness to social media, it may be that it is not the most appropriate model for four reasons:

a) the telecommunications regime gives the regulator the power of stopping the operator from providing the service itself, and not just problematic elements in relation to the service - we question whether this is appropriate in the light of freedom of speech concerns;

b) the telecommunications regime specifies the conditions with which operators must comply, albeit at a level of some generality – we feel that this is too 'top-down' for a fast moving sector and that allowing operators to make their own assessment of how to tackle risks means that solutions may more easily keep up with change, as well as be appropriate to the service;

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

1 Miller C, Coldicutt R and Kitcher H. (2018) People, Power and Technology: The 2018 Digital Understanding Report. London: Doteveryone, available: http://understanding.doteveryone.org.uk/files/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf

2 For more detail see 'Harm reduction in social media – what can we learn from other models of regulation?' May 4 2018 - https://www.carnegieuktrust.org.uk/blog/harm-reduction-social-media-canlearn-models-regulation/

d) the telecommunications regime is specific to the telecommunications context, the data and workplace regimes are designed to cover the risk entailed from broader swathes of general activity.

23) Although the models have points of commonality, particularly in the approach of setting high level goals and then relying on the operators to make their own decisions how best to achieve that - there are perhaps aspects from individual regimes that are worth highlighting:

a) the data protection and HSE regime highlight that there may be differing risks with two consequences;

b) that measures should be proportionate to those risks; and

c) that in areas of greater risk there may be greater oversight.

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

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

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

Duty of care

24) The idea of a "duty of care" is straightforward in principle[3]. 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 harm as a result then there are legal consequences. A duty of care does not require a perfect record – the question is whether sufficient care has been taken. A duty of care can arise in common law (in the courts) or, as our discussion of regulatory models above shows, in statute (set out in a law). It is this latter statutory duty of care we envisage. For statutory duties of care, as we set out above, while the basic mechanism may be the same, the details in each statutory scheme may differ – for example the level of care to be exhibited, the types of harm to be avoided and the defences available in case of breach of duty.

Social media services are like public spaces

25) Many commentators have sought an analogy for social media services as a guide for the best route to regulation. A common comparison is that social media services are "like a publisher". In our view the main analogy for social networks lies outside the digital realm. When considering harm reduction, social media networks should be seen as a public place – like an office, bar, or theme park. Hundreds of millions of people go to social networks owned by companies to do a vast range of different things. In our view, they should be protected from harm when they do so.

26) The law has proven very good at this type of protection in the physical realm. Workspaces, public spaces, even houses, in

3 For more detail see 'Reducing harm in social media through a duty of care' May 8, 2018 https://www.carnegieuktrust.org.uk/blog/reducing-harm-social-media-duty-care/

the UK owned or supplied by companies have to be safe for the people who use them. The law imposes a "duty of care" on the owners of those spaces. The company must take reasonable measures to prevent harm. While the company has freedom to adopt its own approach, the issue of what is 'reasonable' is subject to the oversight of a regulator, with recourse to the courts in case of dispute. If harm does happen the victim may have rights of redress in addition to any enforcement action that a regulator may take action against the company. We emphasise that this should only be in respect of systemic failures and not as a substitute for a civil action in relation to specific items of content. By making companies invest in safety the market works better as the company bears the full costs of its actions, rather than getting an implicit subsidy when society bears the costs.

A broad, general almost future-proof approach to safety

27) Duties of care are expressed in terms of what they want to achieve – a desired outcome (ie the prevention of harm) rather than necessarily regulating the steps – the process – of how to get there. This fact means that duties of care work in circumstances where so many different things happen that you couldn't write rules for each one This generality works well in multifunctional places like houses, parks, grounds, pubs, clubs, cafes, offices and has the added benefit of being to a large extent futureproof. Duties of care set out in law 40 years ago or more still work well – for instance the duty of care from employers to employees in the Health and Safety at Work Act 1974 still performs well, despite today's workplaces being profoundly different from 1974's.

28) In our view the generality and simplicity of a duty of care works well for the breadth, complexity and rapid development of social media services, where writing detailed rules in law is impossible. By taking a similar approach to corporate owned public spaces, workplaces, products etc in the physical world, harm can be reduced in social networks. Making owners and operators of the largest social media services responsible for the costs and actions of harm reduction will also make markets work better.

Key harms to prevent

29) When Parliament set out a duty of care it often sets down in the law a series of prominent harms, as can be seen in the 1974 Act, or areas that cause harm that Parliament feels need a particular focus, as a subset of the broad duty of care. This approach has the benefit of guiding companies on where to focus and makes sure that Parliament's priorities are not lost.

30) We propose setting out the key harms that qualifying companies have to consider under the duty of care, based in part on the UK Government's Internet Safety Green Paper. We list here some areas that are already a criminal offence –the duty of care aims to prevent an offence happening and so requires social media service providers to take action before activity reaches the level at which it would become an offence.

    a) Harmful threats – statement of an intention to cause pain, injury, damage or other hostile action such as intimidation. Psychological harassment, threats of a sexual nature, threats to kill, racial or religious threats known as hate crime. Hostility or prejudice based on a person's race, religion, sexual orientation, disability or transgender identity. We would extend

the understanding of "hate" to include misogyny.

b) Economic harm – financial misconduct, intellectual property abuse,

c) Harms to national security – violent extremism, terrorism, state sponsored cyber warfare

d) Emotional harm – preventing emotional harm suffered by users such that it does not build up to the criminal threshold of a recognised psychiatric injury. For instance through aggregated abuse of one person by many others in a way that would not happen in the physical world (see Stannard[4] on emotional harm below a criminal threshold). This includes harm to vulnerable people – in respect of suicide, anorexia, mental illness etc.

e) Harm to young people – bullying, aggression, hate, sexual harassment and communications, exposure to harmful or disturbing content, grooming, child abuse (See UKCCIS Literature Review[5])

f) Harms to justice and democracy – prevent intimidation of people taking part in the political process beyond robust debate, protecting the criminal and trial process (see concerns expressed by the Attorney General[6] and the Committee on Standards in Public Life[7])

31) We would also require qualifying social media service providers to ensure that their service was designed in such a way to be safe to use, including at a system design level. This represents a hedge against unforeseen developments as well as being an aggregate of preventing the above harms. We have borrowed this idea from risk based regulation in the General Data Protection Regulation and the Health and Safety at Work Act which both, in different ways, require activity to be safe or low risk by design[8]

32) People would have rights to sue eligible social media service providers under the duty of care; for the avoidance of doubt, a successful claim would have to show a systemic failing rather than be deployed in case of an isolated instance of content. But, given the huge power of most social media service companies relative to an individual we would also appoint a regulator. The regulator would ensure that companies have measurable, transparent, effective processes in place to reduce harm, so as to help avoid the need for individuals to take action in the first place. The regulator would have powers of sanction if they did not.

Which social media services would be subject to a statutory duty of care towards their users?

33) Parliament would set out in law characteristics of social media services that

4 J E Stannard, 'Sticks, Stones and Words: Emotional Harm and the English Criminal Law' (2010) 74 Journal of Criminal Law 533, available: http://journals.sagepub.com/doi/abs/10.1350/jcla.2010.74.6.668

5 Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650933/Literature_Review_Final_October_2017.pdf

6 Attorney General The Impact of Social Media on the Administration of Justice: call for evidence, 15 September 2017, available: https://www.gov.uk/government/publications/the-impact-of-socialmedia-on-the-administration-of-justice

7 See Committee on Standards in Public Life Intimidation in Public Life, A Review by the Committee on Standards in Public Life, December 2017, Cm 9543, available: https://www.gov.uk/government/publications/intimidation-in-public-life-a-review-by-the-committeeon-standards-in-public-life

8 8 The Network and Information Systems Regulations 2018 have a similar risk based approach http://www.legislation.gov.uk/uksi/2018/506/contents/made. On this generally, see https://www.carnegieuktrust.org.uk/blog/harm-reduction-social-media-can-learn-models-regulation/

could be covered by the regime. There are always difficult boundary cases and to mitigate this we propose the regulator makes a list of qualifying services[9].

a) Have a strong two-way or multiway communications component;

b) Display and organise user generated content publicly or to a large member/ user audience;

c) A significant number of users or audience – more than, say, 1,000,000;

d) Are not subject to a detailed existing regulatory regime, such as the traditional media.

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

36) In order to maintain an up to date list, **companies which fall within the definition of a qualifying social media service provider would be required in law to notify the regulator after they have been operating for a given period**. Failure to do so would be an offence – as it is a number of existing regulatory regimes. Notification would be a mitigating factor should the regulator need to administer sanctions.

37) **The regulator will publish a list based on the notifications and on market intelligence, including the views of the public.** The regulator's decision to include a service on the list could, as for any such type of decision, be subject to judicial review, as could the decision not to include a service that the public had petitioned for. Services could be added to the list with due process at any time, but the regulator should review the entire list periodically, perhaps every two years.

38) Broadly speaking we would anticipate at least the following social media service providers qualifying, we have asterisked cases for discussion below.

a) Facebook
b) Twitter
c) YouTube
d) Instagram
e) Twitch*
f) Snapchat
g) Musical.ly*
h) Reddit
i) Pinterest*
j) LinkedIn

Managing boundary cases

39) Providing a future proof definition of a qualifying social media service is tricky However we feel that giving the independent regulator the responsibility to draw up a list allows for some future-proofing rather than writing it in legislation. The fact that it is the regulator which makes this list by reference to objective criteria also reduces the risk of political interference. It is quite proper for the government to act to reduce harm, but in our view there would be free speech concerns were the government to say who was on the list. An alternative would be for the regulator to advise the

9 Which social media services should be regulated for harm reduction? May 8, 2018 https://www.carnegieuktrust.org.uk/blog/social-media-services-regulated-harm-reduction/

Secretary of State and for them to seek a negative resolution in Parliament but in our view this brings in a risk to independence and freedom of speech.

40) Internet forums have some of the characteristics we set out above. However hardly any forums would currently have enough members to qualify. The very few forums that do have over one million members have, in our opinion, reached that membership level through responsible moderation and community management. In a risk based regime (see below) they would be deemed very low risk and would be unlikely to have to change their processes significantly. We do not intend to capture blog publishing services, in our view the conversational interaction about a single blog, let alone a whole blogging service, is not on the scale of a social media service and they would not qualify.

41) Twitch has well-documented abuse problems[10] and has arguably more sophisticated banning regimes[11] for bad behaviour than other social networks. Twitch allows gamers to stream content that the gamers have generated (on games sites) with the intention of interacting with an audience about that content. Twitch provides a place for that display, multiway discussion about it and provides a form of organisation that allows a user to find the particular content they wish to engage with. We therefore feel that Twitch falls within scope. Other gaming services with a strong social media element should also be considered, particularly with a strong youth user base.

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

Risk based regulation – not treating all qualifying services the same

43) This regime is risk based. We are not proposing that a uniform set of rules apply across very different services and user bases. The regulator would prioritise high risk services, and only have minimal engagement with low risk services. Differentiation between high and low risk services is common in other regulatory regimes, such as for data in the GDPR and is central to health and safety regulation. In those regimes, high risk services would be subject to closer oversight and tighter rules as we intend here.

44) Harmful behaviours and risk have to be seen in the context of the platform. The regulator would examine whether a social media service operator has had particular regard to its audience. For example, a mass membership, general purpose service should manage risk by setting a very low tolerance for harmful behaviour, in the same way that some public spaces take into account that they should be a reasonably safe space for all. Specialist audiences/user-bases of social, media services may have online behavioural norms that on a family-friendly service could cause harm but in the community where they originate are not harmful. Examples might include sports-team fan services or sexuality-based communities. This can be seen particularly well with Reddit: its user base with diverse interests self organises into separate subreddits, each with its own behavioural culture and moderation.

10 See e.g. Steffan Powell, "Twitch and YouTube 'taking misogynistic abuse in gaming seriously'" BBC Newsbeat, 28 Sept 2016, available: http://www.bbc.co.uk/newsbeat/article/37485834/twitch-andyoutube-taking-misogynistic-abuse-in-gaming-seriously

11 Twitch Community Guidelines Update, available: https://blog.twitch.tv/twitch-community-guidelinesupdates-f2e82d87ae58?sf181649550=1

45) Services targeted at youths are innately higher risk – particularly where youth services are designed to be used on a mobile device away from immediate adult supervision. For example, teen focussed lip synching and video sharing site musical.ly owned by Chinese group Bytedance according to Channel 4 News[12] has 2.5 million UK members and convincing reports of harmful behaviours. The service is a phone app targeted at young people that also allows them to video cast their life (through their live.ly service) with as far as we can make out few meaningful parental controls. In our opinion, this appears to be a high risk service.

Regulation and enforcement

46) Legislation should set the framework within which the regulator will act, allowing it some flexibility and to respond appropriately in a fast moving environment. Our proposal is that the regulator is tasked with ensuring that social media services providers have adequate systems in place to reduce harm. The regulator would not get involved in individual items of speech. The regulator must not be a censor[13].

Harm reduction cycle

47) We envisage an ongoing evidence based process of harm reduction. For harm reduction in social media the regulator would work with the industry to create an on-going harm reduction cycle that is transparent, proportionate, measurable and risk-based.

48) A harm reduction cycle begins with measurement of harms. The regulator would draw up a template for measuring harms, covering scope, quantity and impact. The regulator would use as a minimum the harms set out in statute but, where appropriate, include other harms revealed by research, advocacy from civil society, the qualifying social media service providers etc. The regulator would then consult publicly on this template, specifically including the qualifying social media service providers. Regulators in the UK such as the BBFC, the ASA and OFCOM (and its predecessors) have demonstrated for decades that it is possible to combine quantitative and qualitative analysis of media, neutral of political influence, for regulatory process.

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

50) The companies would then be required to act to reduce these harms. We expect those actions to be in two groups – things companies just do or stop doing, immediately; and actions that would take more time (for instance new code or terms and conditions changes). Companies should seek views from users as the victims of harms or NGOs that speak for them. These comments – or more specifically the qualifying social media service providers respective responses to them (though it should be emphasised that companies need not adopt every such suggestion made)

12 F Manji, 'Children bombarded with sexually explicit chat on Musical.ly and Live.ly' 8 Jun 2017, available: https://www.channel4.com/news/children-bombarded-with-sexually-explicit-chat-onmusical-ly-and-live-ly

13 13 For more detail see 'How would a social media harm regulator work?' May 10, 2018 https://www.carnegieuktrust.org.uk/blog/social-media-harm-regulator-work/

– would form part of any assessment of whether an operator was taking reasonable steps and satisfying its duty of care. Companies would be required to publish, in a format set out by the regulator:

a) what actions they have taken immediately;

b) actions they plan to take;

c) an estimated timescale for measurable effect; and

d) basic forecasts for the impact on the harms revealed in the baseline survey and any others they have identified.

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

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

53) The baseline will reveal four likely outcomes – that harms:

a) have risen;
b) stayed the same;
c) have fallen; or
d) new harms have occurred.

54) If harms surveyed in the baseline have risen or stayed the same the companies concerned will be required to act and plan again, taking due account of the views of victims, NGOS and the regulator. In these instances, the regulator may take the view that the duty of care is not being satisfied and, ultimately, may take enforcement action (see below). If harms have fallen then companies will reinforce this positive downward trajectory in a new plan. Companies would prepare second harm reduction reports/plans as in the previous round but including learning from the first wave of actions, successful and unsuccessful. Companies would then implement the plans. The regulator would set an interval before the next wave of evaluation and reporting.

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

56) At this point we need to consider the impact of the e-Commerce Directive which gives immunity from liability to neutral intermediaries under certain conditions. Although we are not convinced that all qualifying social media companies would be neutral intermediaries within the meaning of the directive, there is a question as whether some of the measures that might be taken as part of a harm reduction plan could mean that the qualifying company which was neutral would lose its immunity, which would be undesirable. There are three comments that should be made here to mitigate this concern:

a) Not all measures that could be taken would have this effect;

b) The Commission has suggested that the e-Commerce Directive be

interpreted – in the context of taking down hate speech and other similarly harmful content[14] as not meaning that those which take proactive steps to prevent such content should be regarded as thereby assuming liability;

c) After Brexit, there may be some scope for changing the immunity regime – including the chance to include a 'good Samaritan defence' expressly.

57) This harm reduction cycle is similar to the techniques used by the Commission in a series of documents as it works with the social media service providers to remove violent extremist content.[15]

Other regulatory techniques

58) Alongside the harm reduction cycle we would expect the regulator to employ a range of techniques derived from harm reduction practice in other areas of regulation. We draw the following from a wide range of regulatory practice rather than the narrow set of tools currently employed by the tech industry (take down, filtering etc). Some of these the regulator would do, others the regulator would require the companies to do.

59) For example, each qualifying social media service provider could be required to:

a) develop a statement of risks of harm, prominently displayed to all users when the regime is introduced and thereafter to new users; and when launching new services or features;

b) provide its child protection and parental control approach, including age verification, for the regulator's approval;

c) display a rating of harm agreed with the regulator on the most prominent screen seen by users;

d) work with the regulator and civil society on model standards of care in high risk areas such as suicide, self-harm, anorexia, hate crime etc; and

e) provide adequate complaints handling systems with independently assessed customer satisfaction targets and also produce a twice yearly report on the breakdown of complaints (subject, satisfaction, numbers, handled by humans, handled in automated method etc.) to a standard set by the regulator.

60) The regulator would:

a) publish model policies on user sanctions for harmful behaviour, sharing research from the companies and independent research;

b) set standards for and monitoring response time to queries (as the European Commission does on extremist content through mystery shopping);

c) co-ordinate with the qualifying companies on training and awareness for the companies' staff on harms;

14 Commission Recommendation on measures to effectively tackle illegal content online (C(2018) 1177 final) 1 March 2018, available: https://ec.europa.eu/digital-single-market/en/news/commissionrecommendation-measures-effectively-tackle-illegal-content-online

15 See e.g. Commission Recommendation (n 14) and Communication Tackling Illegal Content Online (COM (2017) 555 final), available: https://ec.europa.eu/digital-singlemarket/en/news/communication-tackling-illegal-content-online-towards-enhanced-responsibilityonline-platforms

d) contact social media service companies that do not qualify for this regime to see if regulated problems move elsewhere and to spread good practice on harm reduction

e) publish a forward-look at non-qualifying social media services brought to the regulator's attention that might qualify in future;

f) support research into online harms – both funding its own research and co-ordinating work of others;

g) 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

h) maintain an independent appeals panel.

Consumer redress

61) We note the many complaints from individuals that social media services companies do not deal well with complaints. The most recent high profile example is Martin Lewis case against Facebook.[16] At the very least qualifying companies should have internal mechanisms for redress that meet standards set by an outside body of simplicity (as few steps as possible), are fast, clear and transparent. We would establish, or legislate to make the service providers do so, a body or mechanism to improve handling of individual complaints. There are a number of routes which require further consideration – one route might be an ombudsman service, commonly used with utility companies although not with great citizen satisfaction, another might be a binding arbitration process or possibly both.

62) Publishing performance data (specifically in relation to complaints handling) to a regulatory standard would reveal how well the services are working. We wish to ensure that the right of an individual to go to court is not diluted, which makes the duty of care more effective, but recognise that that is unaffordable for many. None of the above would remove an individual's right to go to court, or to the police if they felt a crime had been committed.

Sanctions and compliance

63) 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. We bear in mind the Legal Services Board paper on Regulatory Sanctions and Appeals processes:

'if a regulator has insufficient powers and sanctions it is unlikely to incentivise behavioural change in those who are tempted to breach regulators requirements.'[17]

64) Throughout discussion of sanctions there is a tension with freedom of speech. The companies are substantial vectors for free

16 M. Lewis, 'Martin Lewis: Suing Facebook left me shaking - it's now admitted 1,000s of fake ads, here's the latest', 1st May 2018 updated 2nd May 2018, available: https://blog.moneysavingexpert.com/2018/05/martin-lewis--suing-facebook--left-me-shaking--it-snow-admitted/

17 Legal Services Board, Overseeing Regulation: The LSB's Approach to Its Role, June 2013, available: http://www.legalservicesboard.org.uk/news_publications/LSB_news/PDF/2013/20130611_LSB_Sets_Out_Its_Approach_To_Overseeing_Regulation.pdf

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.

65) All processes leading to the imposition of sanctions should be transparent and subject to a civil standard of proof. By targeting the largest companies, all of which are equipped to code and recode their platforms at some speed, we do not feel that the argument that 'the problem is too big' is adequate. There may however be a case for some statutory defences.

66) Sanctions would include:

a) Administrative fines in line with the parameters established through the Data Protection Bill regime of up to €20 million, or 4% annual global turnover – whichever is higher.

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

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

d) Adverse publicity orders – the company is required to display a message on its screen most visible to all users detailing its offence. A study on the impact of reputational damage for financial services companies that commit offences in the UK found it to be nine times the impact of the fine[18].

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

Sanctions for exceptional harm

67) The scale at which some of the qualifying social media services operate is such that there is the potential for exceptional harm. It is not impossible to imagine a social media service being exploited to provoke a riot. Imagine people were severely injured or died and widespread economic damage was caused as a result. The regulator had warned about harmful design features in the service, those flaws had gone uncorrected, the instigators or the spreaders of insurrection exploited deliberately or accidentally those features. Or sexual harm occurs to hundreds of young people due to the repeated failure of a social media company to provide parental controls or age verification in a teen video service. Are fines enough or are more severe sanctions involving the criminal required, as seen elsewhere in regulatory schemes?

68) 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 extreme circumstances which often involve a death[19]

18 Armour et al, 'Regulatory Sanctions and Reputational Damage in Financial Markets' (2017) 52 Journal of Financial and Quantitative Analysis 1429-1448, available: https://www.cambridge.org/core/journals/journal-of-financial-and-quantitativeanalysis/article/regulatory-sanctions-and-reputational-damage-in-financialmarkets/462D1A709D61F3B94605A64E626A3DEE

19 e.g. L Applebey 'Site manager jailed following fatal fall' Health and Safety Practitioner, 19 July 2016, available: https://www.shponline.co.uk/site-manager-jailed-following-fatal-fall/

or repeated, persistent breaches[20] to seek a custodial sentence for a director. The Digital Economy Act contains power (Section 23) for the age verification regulator to issue a notice to internet service providers to block a website in the UK. In the USA the new FOSTASESTA package apparently provides for criminal penalties (including, we think arrest) for internet companies that facilitate sex trafficking. This led swiftly to closure of dating services and a sex worker forum having its DNS service withdrawn in its entirety.

69) **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.**

Who should regulate to reduce harm in social media services?

70) We now address two linked questions:

a) why a regulator is necessary, as we have already implied it is; and

b) the nature of that regulator.[21]

The Need for a Regulator

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

72) Is the fact that individuals may seek redress in relation to this overarching duty (by contrast to an action in relation to an individual piece of content) in the courts not sufficient? At least two pieces of profound legislation based on duties of care do not have 'regulators' as such – the 1957 Occupiers Liability Act and the 1973 Defective Premises Act. By contrast, the 1974 Health and Safety at Work Act does rely on a regulator, now the Health and Safety Executive (HSE). A regulator can address asymmetries of power between the victim and the harm causer. It is conceivable for a home owner to sue a builder or a person for harm from a building, or a person to sue a local authority for harm at a playground. However there is a strong power imbalance between an employee and their boss or even between a trade union and a multinational. A fully functioning regulator compensates for these asymmetries. In our opinion there are profound asymmetries between a user of a social media service and the company that runs it, even where the user is a business, and so a regulator is required to compensate for the users' relative weakness.

What Sort of Regulator?

73) Assuming a regulator is needed, should it be a new regulator from the ground up or an existing regulator upon which the powers and resources are conferred? Need it be a traditional regulator, or would a self or coregulator suffice? We would not at this stage rule out a co-regulatory model, although our preliminary conclusion is that a regulator is required. As we shall see below, instances of co-regulation in

20 e.g Health and Safety Executive, 'Four Receive Suspended jail Sentences for Health and Safety Failings', 16 November 2016, available: http:// ress.hse.gov.uk/2016/four-receive-suspended-jailsentences- for-health-and-safety-failings/

21 See Who should regulate to reduce harm in social media services? May 10, 2018 https://www.carnegieuktrust.org.uk/blog/regulate-reduce-harm-social-media-services/

the communications sector have run into problems. Self-regulation works best when the public interest to be served and those of the industry coincide. This is not the case here.

74) Whichever model is adopted, the important point is that the regulator be independent (and its members comply with the Nolan Principles[22]). The regulator must be independent not only from government but also from industry, so that it can make decisions based on objective evidence (and not under pressure from other interests) and be viewed as a credible regulator by the public. Independence means that it must have sufficient resources, as well as relevant expertise.

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

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

77) The BBFC seems to have its hands full with the age verification regulator from the Digital Economy Act 2017. The launch date has been missed for reasons that are unclear and in our view this removes them from consideration. This also raises the question of how well delegated responsibilities work; Ofcom has recently absorbed responsibilities in relation to video on demand, rather than continue to delegate them to ATVOD. While the ASA regulates some content online including material on social media platforms, but this is limited to advertisements (including sponsorship and the like). Overall the ASA focusses quite tightly on advertising; this may test its expertise. Adding in the substantial task of grappling with harm social media services more broadly could damage its core functions. The HSE has a strong track record in running a risk based system to reduce harm in the workplace, including to some extent emotional harm[23]. It has a substantial scientific and research capability, employing over 800 scientists and analysts. However our judgement is that harm reduction in social media service providers require a regulator with deep experience of and specialism in online industries, which is not where the HSE's strengths lie.

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

22 Committee on Standards in Public Life, The Seven Principles of Public Life, 31 May 1995, available: https://www.gov.uk/government/publications/the-7-principles-of-public-life

23 HSE, Work-related stress and how to tackle it, available: http://www.hse.gov.uk/stress/what-todo.htm

on this work without unbalancing the organisation.

79) The regulator could be funded by a small fraction of the revenue planned to be raised by the Treasury from taxing the revenues of internet companies[24] , of which this would be but a tiny percentage. The relative costs of large regulators suggest that the required resource would be in the low tens of millions of pounds.

Simple legislation to pass quickly

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

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

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

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

84) We are considering drafting such a bill to inform debate and test our estimate.

11 May 2018

24 K. Ahmed 'Tech giants face new UK tax clampdown' BBC News 22 February 2018, available: http://www.bbc.co.uk/news/business-43161736