Centre for the Study of Media, Communication and Power

King's College London

Submission to:

Consultation on Online Harms White Paper

Department for Digital, Culture, Media and Sport

Written by Martin Moore

About the Centre

The Centre for the Study of Media Communication and Power is an academic research centre based in the Department of Political Economy at King's College London. The Centre conducts research and analysis on the news media, the civic functions of the media and technology, the relationship between media and politics, and the extent to which technology is changing politics.

The author of this submission, Martin Moore, is director of the Centre for the Study of Media Communication and Power and a Senior Lecturer in Political Communication Education. He has had over a decade's experience working on projects related to media regulation, news standards and the influence of technology platforms on politics. This includes academic research, technological innovation, and policy proposals. He was the founding director of the Media Standards Trust in 2006. This submission reflects the views of the author based on previous experience and on research done since joining King's College London in 2015. The author is an employee of King's College London.

Democracy Hacked: How Technology is Destabilizing Global Politics (OneWorld, 2018) provides relevant context for this submission, particularly with regard to identifying some of the individuals and states that are responsible for some of the societal online harms of the type described in the White Paper, and for outlining some of the structural reasons why such individuals and states have been able to act in the way they have. These structural reasons, and comments made regarding the importance of platform scale, are developed further in Martin Moore and Damian Tambini's edited volume *Digital Dominance: the Power of Google, Amazon, Facebook and Apple* (OUP, 2018).

TABLE OF CONTENTS

Summary

This submission **supports**:

- The need for intervention to address online harms, given the evidence of significant and growing online harms that are not currently being adequately addressed (most notably by large digital platforms)
- The creation of an independent regulator that helps to define the processes these large digital services ought to have in place to address illegal harms online, and to monitor the adequacy and effectiveness of these processes over time

This submission **challenges**:

- The breadth of scope to which the new digital duty of care will apply
- The inclusion of legal with illegal legal harms within the new statutory duty of care
- The evidence provided for legal harm – particularly harm that 'threatens our way of life in the UK'

This submission **proposes** that any future legislation:

- Limit the scope of the duty of care to the largest digital platforms
- Constrain the responsibilities of an independent online regulator to the process by which illegal harms are dealt with
- Exclude legal online harms from the statutory duty of care
- Encourage differentiation between digital services on the basis of how they self-regulate legal harms

Many of the problems cited in the White Paper are structural and systemic and will not be solved by statutory regulation. Separate interventions are needed to address them

The need for intervention to address online harms

This submission supports the need for intervention to address online harms, given the evidence of significant and growing online harms that are not currently being adequately addressed (most notably by large digital platforms).

The largest technology platforms – such as Facebook and Google – are now of such a size and scale, and the evidence of their misuse so extensive, that the need for some form of intervention is manifest. The question then is what form of intervention is necessary, sufficient, and proportionate.

The individual and societal harms committed on these services, many of which have only been identified recently, are partly a consequence of opportunities presented by the transformation of communications environment (e.g. that allow malign actors to avoid accountability); partly a result of the way in which digital platforms and services are designed and run; and partly because these services are simply convenient conduits for existing human misbehaviour.

The author of this submission, along with the Centre for the Study of Media, Communication and Power at King's, has documented some of these harms, particularly those affecting democratic systems. *Democracy Hacked: How Technology is Destabilizing Global Politics* (Oneworld, 2018) illustrates the ways in which individual users, plutocrats, and State actors, are using the major digital platforms to manipulate and distort the democratic process in elections and referendums, and how their activities have been enabled by, and in some cases encouraged by, the platforms themselves. In 'Weaponising News: RT, Sputnik and targeted disinformation' (Centre for the Study of Media, Communication and Power, 2019), Gordon Ramsay and Sam Robertshaw detail Russian efforts to sow disinformation and discontent in the UK, and analyse the methods they use to do this.

The author and Centre are therefore aware of the nature and extent of certain harms – particularly societal, and of the failure of the largest digital platforms adequately to address them.

Other online harms that are described in the White Paper have been detailed elsewhere, for example in the Rapid Evidence Assessment conducted on behalf of DCMS by the University of East London and the LSE.[1] Further evidence of the activities of malicious or manipulative actors online has been published by the Institute for Strategic Dialogue, the Atlantic Council, the Oxford Internet Institute, Tactical Tech and by other university centres, think tanks and by numerous news outlets.

[1] Davidson, Julia et al (2019) Adult Online Hate, Harassment and Abuse: A Rapid Evidence Assessment, UEL and LSE

The creation of an independent regulator

This submission supports the creation of an independent regulator that helps to define the processes these large digital services have in place to deal with illegal harms online, and to monitor the adequacy and effectiveness of these processes over time

The largest digital platforms have ample experience of self-regulation. Indeed, regulation is synonymous with running a large digital platform. Amazon has to regulate the relationship between buyers and sellers on its services. Apple has to regulate the apps that are made available at its AppStore. Google and Facebook have to regulate the provision and evaluation of advertising via their platforms.

These platforms have, for the most part, chosen to regulate very lightly. Amazon, for example, sets few limits on the books available on Amazon (leading to the charge that many books are counterfeit).[2] YouTube has, until recently, set few limits on the type of videos that it allows people to upload (as long as they are legal).

This decision – to regulate lightly – was made in order for the platforms to 'scale'. In other words, to increase their number of users, and the content and services they can provide access to, and to do so very quickly.

Yet, by choosing to regulate lightly and to prioritise scale these platforms have allowed – and sometimes encouraged – harmful content and/or behaviour. Facebook, for example, gave third-party app developers access to considerable amounts of personal data in lieu of monetary payment. Some of this personal data was then used as part of political campaigns.[3]

Now that these platforms have successfully scaled (partly as a consequence of their limited self-regulation), they have found they have neither the principles nor the practical means necessary to police their own services. This has led one of their heads, Mark Zuckerberg, to call for intervention and regulation. 'I believe we need a more active role for government and regulators' Zuckerberg wrote in *The Washington Post* in March 2019.[4]

While many rightly criticized the Facebook chief executive for trying to promote a solution to the problem that he helped to create, his call provided a clear illustration of the limits and shortcomings of large platform self-regulation. Should democratic governments ignore the calls of Zuckerberg and others for regulation, then there is little reason to believe that harms identified over the last few years will not get worse.

Even in the US, which has long been antipathetic to regulation, there is a growing consensus, amongst politicians and the public, that the largest technology platforms ought to be regulated and/or broken up.[5]

However, as this submission will outline, such regulation ought to be restricted to the process by which digital services address illegal harms online, targeted at the largest technology companies, and limited to identifiable and well-defined online harms.

---

[2] Streitfeld, David (2019) 'What Happens After Amazon's Domination Is Complete? Its Bookstore Offers Clues', *New York Times*, 23 June 2019

[3] DCMS Committee (2019) Disinformation and 'fake news': Final Report, HC1791, 18 February 2019

[4] Zuckerberg, Mark (2019) 'Mark Zuckerberg: The Internet needs new rules. Let's start in these four areas', *The Washington Post*, 30 March 2019

[5] Pew Research Centre (2018) 'Roughly half the public thinks major tech companies should be regulated more than they are now', June 26 2018, https://www.pewinternet.org/2018/06/28/public-attitudes-toward-technology-companies/pi_2018-06-28_tech-companies_0-05/

The risk of excess scope

This submission challenges the breadth of scope to which the new digital duty of care will apply

The White Paper states that a duty of care will apply to all '**companies that allow users to share or discover user-generated content or interact with each other online**'. This submission contends that this scope is too broad and will have damaging – if unintended – repercussions. These may include:

- The restriction or closure of comments, user reviews and user ratings on internet sites
- The restriction or blocking of access to UK users of non-UK services built around user-generated content
- Greater censorship of user generated content – notably controversial, satirical and political content

It is now quite rare to find an online service that does not 'allow users to share or discover user-generated content or interact with each other online' – often of a personal nature. This can include football fan forums, online safe spaces for people who suffer from mental health difficulties, local community bulletin boards, news sites, or Wordpress blogs.

To put a broadly defined 'duty of care' on all of these services, to oblige them to police not only illegal but legal harms (as yet to be well defined), and to impose statutory regulation on them is, this submission contends, disproportionate and liable to lead to a significant reduction in opportunities for speech online.

At a minimum, it will induce many sites to reduce or close down their comments sections, to limit the ways in which users can interact with one another, and reduce or remove ways in which users communicate at online services and with one another. Certain non-UK sites, whose services are built around UGC – such as Reddit (currently in the top twenty most popular website in the UK, and top ten in the US) are likely to curtail access to their services in the UK.

News sites and smaller community forums will, inevitably, be affected. Despite assurances from the government that this legislation is not designed to capture news sites it is very hard to envision how they can be entirely excluded. As the UK government discovered when it tried to define 'news' for the purposes of the Crime and Courts Act (2013) after the Leveson Inquiry, it is extremely difficult to do this without either being highly specific about what constitutes news, or creating a definition so broad that it includes almost all published information.[6]

In addition, if the legislation chooses to draw the scope this wide, it will give the regulator excessive and impracticable responsibility. A regulator would be responsible for any perceived harms committed against UK citizens across the internet and, for each one it became aware of, assess the degree of harm it represented to the individual(s) concerned as well as to society more generally, and – on the basis of this assessment – decide whether to take action, what action to take, and how to make such action effective (given the global nature of the net). All this despite there being limited evidence of harms enabled by medium and smaller digital services.

This submission supports instead the French proposals which recommend that only the largest social media platforms should be subject to statutory regulation with respect to illegal online harms. The French proposals recommend two thresholds, stating that:

[6] Crime and Courts Act (2013) Provision 41: Meaning of "relevant publisher". See also Schedule 15, exclusions from definition of "relevant publisher"

- "The regulation would automatically apply for services for which the number of monthly users rises beyond a certain percentage of the population of the Member State (between 10% and 20%).
- The regulatory system would also be applied only following a reasoned decision by the regulator in the event of an identified and persistent breach for services with a monthly number of users lying between 0% and 5% of the population of the Member State. It should be noted that the lower this second application threshold, the more stringent and demanding the impact test must be in order to comply with a general principle of proportionality.
- The regulation is not applicable below these thresholds, but the common law provisions of the LCEN remain in force, allowing action for civil and criminal liability of the operators in case of breaches."[7]

Although these thresholds may not be suitable to the UK context, they create justifiable principles for the limitation of statutory regulation.

[7] 'Creating a French framework to make social media platforms more accountable: Acting in France with a European vision' (May 2019), Report commissioned by the Secretary of State for Digital Affairs, France

The inclusion of legal with illegal legal harms within the new statutory duty of care

This submission challenges the inclusion of legal with illegal harms within the new statutory duty of care

Large online platforms have, historically, done a poor job of monitoring and policing certain illegal harms carried out on their services. These include harassment, hate crime, incitement of violence, and certain types of criminal activity. Legislation that obliges them to take greater responsibility for addressing these illegal harms, which establishes a regulatory regime which directs and oversees the processes they have in place, and which ensures that these processes work effectively and expeditiously is, therefore, justified.

Large online platforms have also, historically, done a poor job of monitoring and policing certain legal harms carried out on their services. This is partly because most of the large platforms have not prioritised this function (focused, as they have been, on scaling quickly), partly because they have been relatively unaware of the nature and extent of the activities (unconsciously or by design) and partly because they have not believed that it was their responsibility to constrain legal speech or behaviour.

While it is right that these large online platforms be made aware of the nature and extent of legal harms committed on their services, and be encouraged to address these harms, there are significant risks to including these harms within the statutory duty of care.

Including legal harms in the statutory duty of care obliges digital services to:

1. Set out criteria for what is legal but harmful
2. Develop methods by which to identify content and actors that fulfil these criteria
3. Devise ways in which to deal with this content and these actors
4. Communicate their decisions and rationales

Investing them with such statutory obligations charges these services with considerable power and carries significant risks, particularly with regard to their control of free expression.

Setting out criteria for what is legal but harmful will be complex, subjective, and highly context specific. It is widely recognised, for example, that 'trolling' is difficult to define, is dependent on circumstance, and is not necessarily harmful. Anthropologist Gabriella Coleman, for example, writes of hacker collective Anonymous: 'What began as a network of trolls has become, for the most part, a force for good in the world'.[8] Similarly, most of the Facebook and Twitter posts published by employees of Russia's Internet Research Agency were legal, and would be considered acceptable political speech, but can be viewed as harmful when seen as part of a much wider effort to undermine the legitimacy of the US 2016 election and other political events within Europe.

Developing methods by which to identify content and actors that fulfil these criteria can be equally hard. Facebook, for example, is reported to be developing AI that identifies bullying on Instagram.[9] This includes looking for highly nuanced activities, such as what is referred to as 'intentional inducement of FOMO [Fear Of Missing Out]'. While it is the prerogative of an online platform to try to restrict such activities, making it obligatory through statutory regulation would inevitably encourage digital services to intervene regularly and to make subjective judgments about public conversations.

[8] Coleman, Gabriella (2015, 50) *Hacker, Hoaxer, Whistleblower, Spy: the many faces of Anonymous*, London: Verso

[9] Roose, Kevin (2019) 'Instagram Is Trying to Curb Bullying. First, It Needs to Define Bullying', New York Times, 9 May 2019, https://www.nytimes.com/2019/05/09/technology/instagram-bullying-teenagers.html

Devising ways in which to deal with this content and these actors that is both fair and proportionate, and avoids politicization, is similarly fraught with problems. Attempts by Twitter in the US to flag or remove content that it characterizes as extremist have led, for example, to accusations from Republicans, and from the President, of conservative censorship. 'Twitter should let the banned Conservative Voices back onto their platform, without restriction' President Trump tweeted on June 10th, 'It's called Freedom of Speech, remember. You are making a Giant Mistake!'.[10]

Not all legal online harms are the same. Some harms can lead to significant and irreparable damage, some can generate temporary mental or reputational damage, and some may not be harms at all (certain types of disinformation for example). Seeking to deal with these legal harms through statutory regulation, and all through one regulator that also deals with illegal harms, risks obscuring the differences between harms, assumes that certain legal activities are harmful when they are not, and ignores more effective remedies.

While there is growing evidence of legal harms online, there is currently not enough known about the nature, extent, causes and definition of such harms to warrant their inclusion in a statutory duty of care for digital services.

[10] @realDonaldTrump, Twitter 10 June 2019, https://twitter.com/realDonaldTrump/status/1137702218835136517

Case Study: Disinformation

Research by the Centre for the Study of Media, Communication and Power on Russia Today (RT) and Sputnik illustrates some of the difficulties inherent in defining disinformation as an online harm and obliging digital services to address it as part of their duty of care.[11]

Analysis of the news content of RT and Sputnik over a nine month period found that both Russian news services were systematically publishing news designed to promote Russian power and prowess, to increase scepticism about allegations of Russian interference in foreign countries, to emphasise the dysfunctionalism of western democracies, and to enhance distrust in democratic governments and authorities.

This could be seen clearly in the period following the attempted poisoning of Sergey Skripal and his daughter in March 2018. In the four weeks following the incident RT inserted 138 separate – and often contradictory – narratives to explain what might have happened. Most of these ran counter to the evidence which showed that two Russian GRU officers had committed the Novichok poisoning.

In addition, research by the Centre found that RT and Sputnik produced numerous news stories critical of NATO, alongside lengthy reports on the capabilities of new Russian weapons. These were within a broader narrative about the failure of western democracies, which focused 'on political dysfunction, institutional failure, social division and the negative effects of immigration'.

The actions of RT and Sputnik therefore appear to fit the description of the type of information that, in the words of the White Paper '**threatens these [democratic] values and principles, and can threaten public safety, undermine national security, fracture community cohesion and reduce trust'**.

However, when examined in more detail, it becomes clear how hard it would be for digital services to isolate these harms or address them. Most of the alternative explanations for the attempted poisoning of the Skripals were made by commentators on RT, or by sources interviewed by RT or Sputnik. As such, these news services could claim to be reporting commentary, views and speculation rather than promoting their own perspective. Stories on the dysfunctionalism of western political systems, though given particular priority or emphasis by RT and Sputnik, were not (for the most part) false. Indeed, many of the stories were also published in UK news outlets, though with different prioritisation and emphasis. Stories that were untrue – such as false claims about Russian military capabilities – were only discovered to be false after investigation and after having been reported elsewhere in UK media.

Therefore, to address the problem of disinformation – even on services where it appears to be promoted systematically – would require ongoing and complex editorial decisions by digital services which would require them to become, in the words of Mark Zuckerberg, 'arbiters of truth'. To identify and police similar legal online harms across all content produced by users would require these digital services either to become increasingly like news organisations themselves (and equally editorially selective), or to heavily censor any content not published by what they consider to be an 'authoritative source'. This latter approach would jeopardise the openness and accessibility of the internet and undermine the democratic principles the White Paper seeks to promote.

[11] Ramsay, Gordon and Sam Robertshaw (2019) 'Weaponising news RT, Sputnik and targeted disinformation', Centre for the Study of Media, Communication and Power

The evidence provided for legal harm – particularly harm that 'threatens our way of life in the UK'

This submission challenges the nature and extent of the evidence of legal online harms provided in the White Paper – particularly with respect to harms said to threaten 'our way of life in the UK'.

This submission contends that the evidence cited in the White Paper does not adequately show the consequences of legal harms and, as a result, does not provide a sufficient case for the inclusion of these harms under a statutory duty of care.

Echo chambers and filter bubbles: the White Paper states that '*Social media platforms use algorithms which can lead to 'echo chambers' or 'filter bubbles', where a user is presented with only one type of content instead of seeing a range of voices and opinions'*. The existence and effectiveness of echo chambers and filter bubbles on social media is highly contested, with recent research suggesting that frequent social media users are likely to be exposed to more news sources rather than less.[12] Other research indicates that the people in one's social media networks are considerably more influential in determining the news and information each person is exposed to rather than social media algorithms.[13]

Online manipulation: the White Paper gives online manipulation as an example of a legal harm and, by way of illustration, cites subliminal broadcast advertising. Putting to one side whether it has ever been proven that subliminal broadcast advertising was either effective or harmful (apart from being considered 'creepy' by the public), it is hard to see how this illustration translates to the contemporary online environment. Would, for example, retargeted advertising (when you are served an ad for a product from a site you have previously visited) be considered online manipulation and therefore a 'legal harm'? What if the unsubscribe button on a newsletter was insufficiently prominent – would that be considered online manipulation? Or if, an apartment rental website alerted someone that ten other people have already looked at a flat that they are considering, in order to put pressure on them to commit, would that be considered manipulative? There is a danger that most digital advertising could be considered online manipulation.

Disinformation: the White Paper states that 'Disinformation threatens these [democratic] values and principles, and can threaten public safety, undermine national security, fracture community cohesion and reduce trust'. Yet the examples cited (in Box 12) do not provide evidence for these harms but rather for:

- the public's desire for help to determine what is real news as opposed to fake
- the potential of AI for producing disinformation, and;
- examples of false narratives promoted by Russia

The references to 'Impact' (Box 12) do not provide any evidence of the impact of disinformation but rather to the limited public understanding of what determines the news people see. The public were similarly unaware of the process by which print news organisations determined which news appeared in the paper each day.

[12] Newman, Nic et al (2017, 10) *Digital News Report 2017,* Reuters Institute for the Study of Journalism, University of Oxford

[13] Bakshy, Eytan and Solomon Messing, Lada A. Adamic (2015) Exposure to ideologically diverse news and opinion on Facebook, SCIENCE, June 2015, pp 1130-1132

Proposed Changes

Based on the responses to the consultation made above, this submission suggests the following changes to the White Paper:

1. Limit the scope of the duty of care to the largest digital platforms

This submission proposes limiting the scope of the statutory duty of care to the largest digital services where there is clear evidence of significant harms that have not been adequately addressed. The size threshold could follow the principles of the recommendations recently made in France.

2. Constrain the responsibilities of an independent online regulator to the process by which harms are dealt with

This submission proposes giving an independent regulator responsibility for ensuring that large digital services deal adequately, expeditiously and proportionately with illegal harms online.

3. Exclude legal online harms from the statutory duty of care

This submission proposes excluding legal online harms from the statutory duty of care.

4. Encourage differentiation between digital services on the basis of how they self-regulate legal harms

This submission proposes that the legislation emphasise the opportunity for different digital services to differentiate themselves by the way in which they choose to self-regulate legal harms on their services.

Structural Change and Positive Interventions

There appears to be an assumption within the Online Harms White Paper that many of the harms identified on the internet can be solved through statutory regulation. Based on the research done by the Centre, this is a misguided assumption.

We are currently going through a period of significant structural change with regards to communication. Such periods of rapid structural change present many opportunities for illegal, unethical, malicious, vexatious, hazardous, hurtful and mischievous activities that may cause harm. These periods also present many opportunities for legal, ethical, constructive, creative, thoughtful, funny and generally benign activities that support healthy, considerate communication.

Many of the problems cited in the White Paper will not be solved by regulation – it is not a silver bullet. Many of the problems are structural and systemic. As such they will only be ameliorated by changes to the whole system. Such changes may involve reducing the size and scale of the large technology platforms, encouraging and supporting new and emerging civic spaces online, addressing the growing democratic deficit in news provision – particularly at a local level, and reforming democratic processes. To seek to solve so many problems through a single statutory regulator is, in the view of this submission, misplaced.

For further information or to contact to the author of this submission:

Martin Moore
Director, Centre for the Study of Media, Communication and Power
Senior Lecturer in Political Communication Education

King's College London
Department of Political Economy
Bush House (North-East Wing)

Email: martin.moore@kcl.ac.uk
Twitter: @martinjemoore