Reducing Online Harms through a Differentiated Duty of Care:
A Response to the Online Harms White Paper

Damian Tambini

The Foundation for Law, Justice and Society
in association with the Centre for Socio-Legal Studies
and Wolfson College, University of Oxford
Executive Summary

- The government is correct to propose a new regime for online content regulation. Evidence points to significant ongoing harms and a failure of the old regime. New considerations such as foreign interference in democracy and a business model based on exploitation of engagement via behavioural data call for a new system of regulation.

- The government is correct to propose a new institution with discretion to use a range of powers. Whilst regulatory plurality is advisable to protect freedom of expression and prevent censorship, it will be most efficient to centralize functions such as research and corporate relationship management in one institution, and maintain plurality through multiple standards codes on content that is legal but potentially harmful.

- There are potential dangers in the approach of the White Paper however, and these could have a significant impact on freedom of expression.

- There is a danger of regulatory uncertainty regarding what is within scope and the relationship of existing law to new de-facto regulatory standards derived from excessively wide regulatory discretion. The extension of pseudo-liability for harmful but not illegal content is problematic and, together with harsh sanctions, could lead to significant chilling of freedom of expression.

- Harms are insufficiently defined, and there is a blurring of the boundary between illegal and harmful content.

- Many of the problems with the government’s approach could be addressed by a clear distinction between the illegal/clearly defined and the legal/less clearly defined categories of content: (i) in the case of illegal content, the sanctions that the regulator will apply will be more direct and include civil fines and other penalties on a sliding scale; (ii) in the case of legal but harmful content, the monitoring, advice, and transparency function will be more prominent.

- It is for Parliament to decide if new offences and categories of content require new laws and liabilities and set standards for blocking or filtering the most dangerous content. Given the dynamic nature of online harms, the process for introducing new laws to reflect harms should be made more efficient and evidence-based, with advice from the new regulator.

- Censorship-like functions must meet the European Convention on Human Rights test of proportionality, legality (parliamentary oversight), and necessity in a democratic society.
Introduction: A Response to the Government’s Proposals

The UK government’s proposals, set out in its Online Harms White Paper, reflect growing frustration about a perceived lack of effective action by large internet companies to deal with online hate, incitement and harassment, disinformation, terrorism, and child abuse online. Since 2017, Parliament has conducted a series of inquiries, on ‘fake news’; on online hatred and incitement; and on internet regulation. These inquiries identified a number of problems with the approach of large social media platforms such as Facebook, Twitter, and YouTube (i) in providing protection or redress relating to content or conduct that is legal but breaches their own community guidelines and terms of service, and (ii) in removing or blocking content that is illegal. In short, harmful content was not effectively dealt with through self-regulation, and there was an enforcement problem with illegal content.

The government proposes a new regulator to enforce a ‘duty of care’ that platforms owe to their users. This constitutes a comprehensive new regulatory framework to cover all these various harms. In adopting this approach, the government has signalled that it is rejecting principled arguments against ‘state’ regulation per se, and adopting a pragmatic approach to online harms. In so doing, they are reflecting a long-term acceptance that some regulation of online speech may be necessary in a democratic society, and rejecting US-style First Amendment absolutism that rejects all forms of interference with speech by state bodies or law.

The White Paper points out that because certain voices, in particular the voices of women, children, and minorities, may be silenced by hatred and harassment online, there may be a positive obligation on states to intervene in speech in order to regulate online service providers in order to ensure that they do protect such people and ensure they are able to enjoy their right to communicate.

The call for regulation of online communication reflects a growing consensus that the internet is different to other spaces, in ways that justify an approach to regulation that differs to the offline world. In the past, the view was taken that online should be free from regulation because it was the realm of pure speech. More recently, the view has become widespread that online speech should be regulated differently to offline speech, for the following reasons.

- Online speech is disinhibited (and can be anonymous). There is evidence that many of the social and informal constraints on speech do not work for social media, for example.
- The business model and software design of social media often rewards and amplifies hatred, misogyny, and a range of noisy, anti-social behaviour because it rewards and amplifies any form of ‘engagement’. There is evidence that business interests based on high levels of engagement and ‘noise’ are benefiting from negative social externalities.
- [Related] As an open network, the internet can facilitate communication by machines and sophisticated uses of artificial intelligence (AI) to

Reducing Online Harms through a Differentiated Duty of Care
control/modify human behaviour.10 It can be difficult to identify and separate machine speech and human speech.

‘Bad actors’, including foreign states, can also feed into and benefit from the network features and communicative effects associated with internet communication. In so doing they can breach national security and conduct low-level permanent information warfare with long-term and widespread impacts on national interests.11

Related to this, there is a difficulty of enforcing existing law online. This has been evident since the early days of the internet with, for example, difficulties of regulating speech where there was a clear and agreed justification (as, e.g., with child abuse content, defam ation, and intellectual property12) but difficulty agreeing the relevant jurisdiction. The case of WikiLeaks showed that whilst the internet is not regulation-proof, it does have affordances that enable those who seek to do so to create difficulties for those aiming to regulate it.

The White Paper therefore calls for a single integrated regulatory structure for internet intermediaries. There will be principled objections to this on the basis of ‘internet freedom’ principles or the idea that internet independence from state institutions is inherently superior, but given issues of disinhibition, algorithmic amplification, and foreign exploitation of these, there is a need for some kind of intervention.

This is a major assertion of national sovereignty over internet content and conduct that occurs on the internet. It challenges some accepted notions of freedom of expression and the ways that it has been applied online. In many ways, the White Paper is an attempt to change the incentive structure and ensure that new incentives are introduced that will reinforce a ‘race to the top’ and boost the quest for a new ethical framework for social media, whilst also ensuring that the search for truth, rather than simply engagement or noise, is an overriding design principle of social media. As the proposals currently stand, there remains some doubt about whether they will sufficiently incentivise a virtuous cycle of responsibility among service providers.

The macro question to ask is whether this framework will lead to genuine behaviour change and reverse the trend toward growing online harms. In its current form, it is difficult to say. One of the most likely scenarios is that, as a result of the new duty of care, the costs of compliance may become so great that smaller or medium-sized companies, if within scope, may decide that rather than comply they will simply cease to provide services that fall within the definition of scope for the White Paper. A very real danger is that YouTube and Facebook are in fact the only services with the resources to provide the necessary levels of moderation. The framework uses only a few of the incentives that are available, and does so in a rather sudden and blunt, rather than an iterative way.

If it is going to be effective, and not lead to market exit of social media companies, the intervention needs to be more targeted, deeper, and longer term. Industry codes should be developed in accordance with a clearly defined and agreed framework, and with long-term policy development in mind.13

Another objection to the White Paper is that the ‘harm s’ are ill-defined. There have been many criticisms of the White Paper along these lines: the harms, it is claimed, are amorphous, subjective, and attempt to ‘roll-in’ collective issues such as harm to society. Cyberlaw expert Graham Smith suggests14 that this vagueness carries with it the danger that ministers will use the framework to outlaw particular views according to their own policy preferences, and there is a danger of legal uncertainty undermining the rule of law.

This is premature. It is certainly the case that such a framework, if implemented badly, could entail these dangers. But it is not the case that this general framework must necessarily lead to such outcomes. The next stage of policy development is therefore crucial: as currently defined, the new regulatory framework may or may not impose a chilling effect on speech. Depending on the detail of proposals, and the process of implementation, the framework could, but will not necessarily, result in a significant chilling of speech online.
Policy frameworks for online speech regulation should meet the following requirements:

- They should help rather than hinder the continuity of freedom of expression and open public debate, even if a more authoritarian government is in power. We should set a policy framework that is robust and constitutes part of the overarching framework of checks and balances in our democracy. This means that the framework should pass the test of a simple thought experiment: in a situation in which an incumbent government wished to quell significant areas of public debate or close down opposing voices, would this structure enable them, or prevent them doing so?

- Regulation should provide certainty about who is in scope and what standards apply. The White Paper has been rightly criticized for vagueness about this, but it is after all a White Paper. Decisions will now have to be taken about how to implement bright line rules.

- New regulatory frameworks should not undermine the economic viability of a plurality of free spaces for online public debate.

- They should be potential models for other countries elsewhere in the world that wish to uphold fundamental rights such as freedom of expression, and to support democracy and the rule of law.

In summary, proposed legislation on online harms could be a proportionate and effective response to a wide range of negative outcomes associated with internet use, but it could also be significantly damaging for freedom of expression and pluralism. Which of these outcomes emerges depends on decisions which are to be taken at the next stage of policy development. I am going to focus on the following key areas:

- regulatory architecture and independence
- definition of ‘harms’
- companies and services within scope
- market structure/business model
- implications for freedom of expression

Regulatory structure and independence

One of the reasons the UK proposals are radical is that they propose a new regulator for social media with wide reaching objectives, powers, and sanctions. Critics will claim that this is effectively a licensing regime, and with some justification. In the light of the potential for selective chilling of speech, there should be strict institutional separation and operational independence of any institution that has any potential to shape the formation of public opinion that could be subject to capture by the state or any interest group. There is also an advantage in having a plurality of standards, at least as regards speech that is legal but may be considered harmful.

The Council of Europe has made a number of recommendations on independence of regulatory authorities in the communications sector. In the media sector regulatory independence is a key requirement in order to protect democratic self-government, freedom of expression, and media independence. The European directive on audio-visual media services in its most recent version establishes a legal obligation on states to support the independence of regulatory authorities. A regulator without clarity on the legal basis of its interventions or relevant standards may require additional guidance from government, the provision of which could breach standards of independence of regulatory authorities.

The siting of the regulatory function with Ofcom would ensure that structural independence and separation from government is secured under the Communications Act (2003). Government has set out some broad definitions of harms already in the White Paper, and it is assumed that these would be further refined by Parliament in the light of codes written by the social media companies, but it is companies themselves that will write detailed codes of conduct required by the regulatory regime. This would permit at least some independence of code-setting and even plurality of standards, in a manner analogous to the broadcasting regime under the Communications Act, whereby both Ofcom and the broadcasters themselves develop programme codes and editorial...
guidelines, under the terms of general licence conditions. The structure does not envisage any form of prior censorship or approval by a regulatory body. In general, such a structure does follow established UK and European good practice in setting institutional frameworks for regulation that are, if not immune, at least resistant to censorship power grabs by the executive. The White Paper perhaps does not pay sufficient attention to the importance of the rules of engagement between regulators and regulated. In section 5.6 for example, there are many dangers of open-ended dialogue and horse-trading, which would compromise principles of regulatory independence and could lead to opaque reciprocities.

Definition of harms

The notion of regulation against harm has noble philosophical lineage: it references the ‘harm principle’ according to which restrictions on liberty are only justified to protect other rights or the rights of others, which was advanced by John Stuart Mill. The proposed framework has been criticized, however, for being too broad and ill-defined. It could also be criticized for being too narrow in that it leaves out many ‘harm’ that are associated with social media platforms.

Setting out the regulatory objectives of an independent body is a subtle art, particularly in areas where regulatory independence is so crucial. The government’s approach has been to describe a number of harms, rather than, for example, to set out specific targets or quantifiable objectives, or to define broader public values, specific protected groups or interests, or public interest objectives. One consequence of this is that the framework is overwhelmingly negative in the sense that broadcast licensing tends towards a definition of positive objectives to be realized (such as public service, quality, and the public interest). There are some aspects (for example, with regard to quality news and quality journalism) where positive objectives creep in.

The harms that the proposed regulator will be responsible for reducing consist in both illegal and legal forms of content, and include harms that the White Paper acknowledges have a ‘less clear definition’. This is problematic: the underlying legal basis of intervention is fundamentally different in the case of the White Paper’s second list of harms. The central legal and constitutional problem here is that establishing new standards in a code of conduct, and introducing sanctions and fines for ‘harm with a less clear definition’ and that are also legal, does not pass the European Convention on Human Rights free speech test according to which restrictions have to be prescribed by law, and necessary, for a legitimate aim. The requirement that restrictions should be ‘prescribed by law’ is a safeguard against a slippery slope to censorship. Constraints on speech should not be imposed on the basis of opaque agreements between platforms and politicians—a scenario arguably left open by the White Paper—they should be subject to the constraint of parliamentary debate.

Companies/services in scope

Companies subject to the duty of care are all those that provide opportunities to share user-generated content and for users to interact with one another, but there is a somewhat ill-defined general carve-out for ‘private communication’ (s4). This latter will create some definitional challenges.

The White Paper proposes, rather than defining a size threshold triggering regulation, locating discretion with the regulator to decide where to place regulatory emphasis: the regulator will have to decide, on the basis of evidence, what the harm reduction priorities are at any one time. This could lead to challenge of decisions but should result in an evidence-based decision-making and planning process.
There may be a need to protect pluralism and choice, including, for example, through regulatory plurality, which would be an institutional safeguard against capture of a regulatory authority by state power or other interests.

It might be useful to specify that companies within scope are those whose ‘primary purpose’ is the provision of services under the definition, rather than merely companies that provide such services. This is the current suggestion, which could entail the excessively wide categorization of service providers.

**Sanctions and enforcement**

The current draft outlines a wide range of potential powers and sanctions, including fines, disruption of business activities, measures to impose liability on individual members of senior management, and measures to block non-compliant services. The imposition of these remedies on service providers could have a significant impact on the provision of open public sphere services. In short, the sanctions could be too drastic, and rather than a deterrent effect that would increase responsibility it will simply deter service providers from providing social media services, leaving the market to a few large ‘surveillance capitalism’ companies based on a scale model and brokering of personal data.

Sanctions should be directly linked to the nature of the harms. For example, through ensuring that regulators apply legal remedies appropriate to the general law in the case of illegal content, and agreed tariffs and fines, together with traditional reputational/self-regulatory sanctions in the case of the self-regulatory range of sanctions.

**Market structure/business model**

One implication of establishment of a new regulatory apparatus is the potential that it may, through raising costs of compliance for everybody, favour companies that can achieve significant cost savings through economies of scale and therefore raise barriers to entry and consolidate existing players. This danger is acknowledged in the White Paper, which contains a number of initiatives, for example, for the regulator to support innovation and SMEs in particular. The regulator will also have duties to take into account the impacts on competition and small businesses. The proposals outlined in the White Paper would arguably do something to mitigate this impact, but the costs of compliance are likely to be so great, and the impact on risk calculation of service providers so huge, that it remains to be seen whether they can do anything to mitigate this problem.

A central challenge for this regulatory framework is the integration of competition and other regulatory objectives. Whilst there is a need for a general regulatory action to set new standards, and ensure that sanctions are engaged if companies repeatedly fail to reach them, behaviour change will also result from the competitive discipline exerted on companies by informed consumers switching service providers. The proposed regulator should have a clear remit to monitor and improve competition and switching through the provision of timely information on harms and protections.

**Implications for competition, freedom of expression, and innovation**

Freedom of speech groups such as Article19 have already come out against the proposal. They favour a more pure self-regulation approach they are calling ‘social media councils’. In the view of many freedom of expression groups, the lack of clear definitions of harms, the harsh sanctions, and the blurring between illegal and legal content categories under the ‘duty of care’ scheme will lead to a slippery slope toward censorship. In opposition to those views, other stakeholders have stressed that online manipulation, information warfare, disinformation and hate, for example, provide ample justification for measures that would be considered illegitimate restrictions of freedom of expression.

The White Paper does seek to address these issues. The regulator is charged with a number of ‘due regard’ get-out clauses — so in response to the objection that compliance may raise barriers to
entry and stifle innovation, the regulator will have duties to promote free speech, competition, and innovation. These may be incompatible. It is likely that the costs of complying with the proposals as outlined in the White Paper would be so huge as to tip the risk balance against all but the biggest and richest companies providing social media services free to users.

The differentiated model proposed by this paper has the advantages of creating the potential for a lower level of compliance costs for some companies and some categories of harm. Smaller companies could, for example, be exempted from some of the obligations to monitor or remove content, which could be triggered by a size threshold. A plurality of types and standards of regulation may present some challenges of clarity and predictability of regulatory standards (it would have to be closely monitored), but relying on self-regulation and the market to achieve as many regulatory objectives as possible would be hugely attractive from the point of view of continuing the proud tradition of speech liberty, and innovation in communication technology in the United Kingdom.

It is important to note also that there is an important nexus and relationship between freedom of expression and the market structure. Obligations on dominant companies to take down content can be equivalent to censorship functions (particularly if they are imposed through a single code).

According to the White Paper: ‘7.2.7. Companies will need to take proportionate and proactive measures to help users understand the nature and reliability of the information they are receiving, to minimize the spread of misleading and harmful disinformation and to increase the accessibility of trustworthy and varied news content.’

Many of the measures that will be imposed on companies include ‘real name policies’ to prevent misinformation through anonymity and misrepresentation of identity, flagging and tagging (e.g., through fact-checking and obligations to surface quality news — ‘the news quality obligation’), and downgrading unreliable news, such as news that has failed an independent fact-checking test.

The White Paper proposes that such standards and expectations should be written into the code provided by the regulator. There are also hugely far-reaching clauses such as 7.30, according to which, ‘companies will be required to ensure that algorithms selecting content do not skew towards extreme and unreliable material in the pursuit of sustained user engagement.’ How precisely this is to be done is not set out in any detail, but the White Paper acknowledges that ‘7.31 the code of practice … will ensure the focus is on protecting users from harm not judging what is true or not.’

**Clarifying the duty of care**

Bold action to reform social media regulation is welcome. However, the current proposals could create regulatory uncertainty and undermine competition, innovation, and free speech. In terms of the standards set out at the start of this policy brief, they raise significant questions about all four:

- The White Paper proposals could undermine long-term institutional protection of freedom of expression.
- The proposals could lead to significant uncertainty about scope and who is regulated.
- The proposals could undermine the economic viability of provision of online public spaces.
- The proposals should not be advanced as a model of regulation to be recommended in other countries.

The following main objections to the White Paper underlay this assessment:

- The harms are vague and ill-defined.
- The policy approach erodes the distinction between illegal and harmful.
- The regulator is required to impose sanctions for dissemination of content that is both difficult to define, and would previously be considered legal speech.
- It will therefore lead to potential mission creep and regulatory uncertainty.
- The proposed sanctions are too harsh and would lead to chilling of speech and significant reduction in provision of free, open spaces for public debate.
- The proposed framework will not lead to behaviour change or enlightened self-interest based on voluntary regulation, but may lead to a reduction in the plurality of sites for free and open exchange.
- It is likely to raise barriers to entry/lead to market exit of all but the dominant players.
- The quality of information provided and transparency reporting will not empower consumer switching if there is a reduction in number of providers and very high barriers to entry.

These objections can be addressed by careful clarifications and regulatory design. There is a need for a single regulator in order to centralize information, transparency, research, and expertise in one organization. However, the legal and constitutional basis of regulation of illegal content and legal content are distinct and should remain distinct, otherwise there is a risk of chilling of free speech, regulatory uncertainty, and a failure of the regulatory model. The new regulator should work closely with industry to create a new code on illegal online harms. For the category of legal/’difficult to define’ content, the regulator should provide research, transparency, good practice, and oversight of a plurality of self-regulatory codes, and work to promote good practice in self-regulation.

**Clarified online harms regulatory scheme**

The central challenge is to clarify the process of writing codes of conduct and adjudicating on their application. I propose a framework that provides for a much enhanced research and advice function, and a bifurcated approach that differentiates between illegal and harmful content. As regards illegal content, Ofweb would provide a unified single code outlining procedures for monitoring transparency and takedown. As regards the category of harmful but not illegal content, which is more subjective, Ofweb should support the development of multiple codes and foster consumer switching, competition, choice, and enhanced media literacy.
A RESPONSE TO THE ONLINE HARMs WHITE PAPER

The differentiated duty of care

For illegal content requiring urgent action to be removed or blocked, such as terrorism, child abuse, and hate speech that meets the legal threshold of incitement, an increased pressure to enforce should be ensured by a regulator through a code of conduct that clearly sets out the standards and expectations regarding procedures for removing or blocking content and other relevant responsibilities. These categories of harm are such that they may justify some form of prior restraint, and where the urgency of content removal by platform providers would justify effective deterrent sanctions for their failure effectively to do so. A code would ensure that this process is transparent and standards are supported by Parliament.

For legal but harmful content, Ofweb should promote best practice and provide guidance to improve self-regulation, and promote, through monitoring of self-regulation, improved consumer awareness, and competitive pressure for a culture of responsibility. This category of content includes political speech and other sensitive areas where prior restraint would traditionally be regarded as highly undesirable, and where too great a deterrent effect on speech would be socially undesirable. State control of this type of content would too easily tip the balance toward censorship and the restriction of free speech.

Therefore, both categories of content should be part of the regulator’s remit, but the regulator should adopt a ‘one country two systems’ approach and not confuse the distinction between illegal and harmful content. In effect, each company would be responsible to observe a general regulator-derived code for dealing with speech that clearly meets a standard of illegality, and a company-specific or sector-wide voluntary code that reflects the terms of service of the company. Both the code itself, and company records of complaints and compliance, will be audited on an annual basis by the regulator. The regulator will also monitor consumer awareness and consumer switching rates.

Hence, the duty of care should incorporate both (1) the external, public regulatory regime derived from the general Ofweb code and (2) the new, internal framework for best practice, advice, and self-regulatory audit and oversight that provide an additional layer of ethical restraint and responsibility-enhancing rule-making.

Ofweb would perform a multiple, centralized role:

- Standard setting through development of a code on content that is both illegal and harmful.
- An ombudsman function as an appeals forum for complaints should consumers not be satisfied with adjudication within a given time period.
- A research and advice function.
- Monitoring compliance with the code on illegal content.
- Monitoring and advice role in relation to harmful (legal) content.
- Advising companies seeking to develop their self-regulatory codes.
- ‘Name and shame’ reporting on companies’ codes and procedures on legal content, and levels of enforcement of their terms and conditions.
Notes


2 The White Paper (S2: Table 1) sets out three lists of online harms within scope. Those that are clearly defined are also clearly illegal. Less clearly defined tend to be those that are not illegal. This is a clear distinction that should be maintained.

3 See the Centre for Law and Democracy’s Guide to Article 10 for the explanation of international standards on freedom of expression, and p. 39 in particular for the test of necessity: https://rm.coe.int/16806f5bb3

4 See: https://www.gov.uk/government/consultations/online-harms-white-paper


8 According to UN Special Representative John Ruggie’s Principles on Business and Human Rights, states have a positive duty to protect and promote human rights, as well as a negative one not to infringe them.

9 See J.P. Barlow’s ‘A Declaration of the Independence of Cyberspace’, at: https://www.eff.org/cyberspace-independence


11 The justification for this was set out in the report of the LSE Commission on Truth, Trust & Technology, at: http://www.lse.ac.uk/media-and-communications/truth-trust-and-technology-commission/The-report

12 See, e.g., celebrated cases such as Godfrey vs Demon Internet Ltd (2001).

13 The justification for this was set out in the report of the LSE Truth, Trust & Technology Commission: http://www.lse.ac.uk/media-and-communications/truth-trust-and-technology-commission/The-report


15 See for example the Council of Europe Recommendation Rec (2000) 23 of the Committee of Ministers to member states on the independence and functions of regulatory authorities for the broadcasting sector.

16 Such approaches can be found in the licensing regimes for public broadcasters established by the Broadcasting Act (1996) and the Communications Act (2003).

17 The News Quality Obligation proposed by the Cairncross Review.

18 White Paper Table 1. (p. 31).

19 White Paper Table 1. (p. 31).

20 For a developed discussion of this approach, see the report of the LSE Truth, Trust & Technology Commission 2018: http://www.lse.ac.uk/media-and-communications/truth-trust-and-technology-commission. See also, D. Tambini et al., Codifying Cyberspace. London: Routledge, 2008.
The Foundation

The mission of the Foundation is to study, reflect on, and promote an understanding of the role that law plays in society. This is achieved by identifying and analysing issues of contemporary interest and importance. In doing so, it draws on the work of scholars and researchers, and aims to make its work easily accessible to practitioners and professionals, whether in government, business, or the law.

Damian Tambini is Associate Professor in the Department of Media and Communications at LSE, where he also serves as Programme Director for the MSc Media and Communications (Governance). He is an expert in media and communications regulation and policy, and active in policymaking as well as academic research. He is frequently called to give evidence to parliamentary committees and provide formal and informal policy advice to government. From 2014 to 2015 he served on the UK Government Expert Panel, advising on the value of electromagnetic spectrum. He was called to give evidence to the Leveson Inquiry in 2012, and from 2009 to 2010 he served on the Communications Consumer Panel, a non-executive role at the communications regulator Ofcom.

For further information please visit our website at www.fljs.org or contact us at: