

Internet Harms

We need a Regulator, not a Censor

[The December 2019 Queen’s Speech](#) promises ‘legislation to improve internet safety’ building on *the Internet Harms White Paper* published earlier in 2019. And *Carnegie UK* almost simultaneously published [a draft Online Harm Reduction Bill](#) and explanatory notes which may in many respects be quite similar to the government’s eventual proposals.

There will of course be lively debate about what harms are to be caught by the legislation, and how it is to be enforced. (My own view is that there should be a dedicated regulator, rather than ask Ofcom to take on yet more demanding duties.) But the broad approach – *duty of care regulation*, not censorship - seems to be becoming clear. As it happens, Cambridge’s *Bennett Institute for Public Policy* has recently published an article written by me which drew heavily on the work of [the Carnegie UK Trust](#) and on conversations with Lorna Woods, Professor of Internet Law at the University of Essex. I reproduce it below, with some additional material, in order to encourage wider understanding of the issues and of Carnegie’s proposals.

I will report further key developments via my [@ukcivilservant Twitter feed](#) and on the [Understanding Regulation website](#) – specifically the [Online Safety & Harm web page](#).

[illegible]

Hate speech, harassment, false accusations and baseless conspiracies cause huge harm. But free speech is a vitally important right in any democracy.

How should this tension be resolved when addressing the challenges presented by big social media platforms? The key is to recognise that the harm is amplified or otherwise by its context. An otherwise provocative argument, or a powerful but distressing image, can do huge harm if taken out of context and amplified by thoughtless algorithms or cruel attention seeking.

It would be quite wrong – and probably totally impractical – for a regulator to act as a censor and be required to decide whether particular items should be posted on social media platforms.

Instead, the regulator should be tasked with ensuring that the platforms' services and processes are, so far as reasonably practical, structured and designed so as to reduce the risk of harm to users.

Platforms may, for instance, be expected to ask themselves:

- Have we considered the risks associated with the service we provide?
- Are we aware of the ways users are engaging with our systems?
- Are we responding appropriately and proportionately to the unintended (and sometimes intended) consequences arising out of the use of our systems?
- Are we following best practice when deploying tools etc. intended to reduce harm?

Platforms should not be forbidden from making available material that some would find objectionable - as long as it is published in such a way as to reduce the damage to those who might be harmed.

It should be for the platforms – not the regulator – to decide how best to minimise the harms that might result from their services, and to demonstrate that they have done so. They have the necessary technical knowledge and resources, and they are best placed to understand the needs and vulnerabilities of their users¹. They also need to decide how best to fund their services, including through clicks/advertising, whilst minimising resultant harms. And a number of tools and approaches might be brought to bear, including:

- Adjusting the impact of recommender algorithms, targeted advertising and clickbait
- 'Age gates' – even if imperfect
- Transparency, including about complaints and the platform's responses to those complaints
- Giving users access to blocking tools
- Giving users access to correction tools
- Aggressive content moderation²

So how might it work in practice? There are at least five separate sets of issues.

¹ (See for instance [Facebook's impact assessment of their presence in Myanmar](#).)

² Facebook, for instance, ensures that some links and words immediately trigger an algorithm that prevents the item from being posted, but most moderation takes place only after problematic content is reported by users. This is often far too late.

- (1) Platforms are already prohibited from carrying obviously illegal content – adverts for drugs and the like. So no great change is needed here, although the regulator would need to be assured that the platform had taken steps to reduce illegal content as far as reasonably possible.
- (2) Platforms would become responsible, so far as possible, for restricting access to particularly dangerous or sensitive content. This might include:
 - Inflammatory and false material of the sort that inflamed the violence against the Rohingyas in Myanmar
 - Live streaming of crimes such as terrorist activity
 - Breaches of user privacy, such as allowing access to genetic or financial data or other information people want to remain private, and
 - Scams, such as adverts for dangerously unregulated financial and other services, and such as rip-off websites that pretend to be official government sites but then overcharge for a service that could otherwise be accessed for free or more cheaply.
- (3) There would be then be a number of areas where discussion would be permitted amongst those interested in the subject, but proselytising and evangelising to others might be prohibited. Such specified areas might include blasphemy, or anti-abortion or anti-vax messaging³- but they would need to be defined by politicians, not the regulator, aiming to balance freedom of speech against:
 - individuals' right to choose not to hear certain messages, and
 - society's need to safeguard public or individuals' health and safety.

The web would therefore retain dark and interesting corners for those interested in going into them, but platforms would be responsible for ensuring that such dark material was seen only by those who wished to see it.

- (4) Platforms would need to consider the extent to which their services were accessed by vulnerable users and children, and take any necessary steps to ensure that those users were not easily able to access material that would be harmful to them – or indeed driven towards such material via the site's algorithms. Popular public services such as Facebook, Snapchat and Instagram would in particular need to ensure that they offer a safe public

³ [The National Audit Office has reported](#) that there are several potential causes for the decline in uptake of pre-school vaccinations, but there is only limited evidence of any major impact on vaccination uptake rates from anti-vaccination messages. So limiting ant-vax messaging may be an over-reaction.

place for families. [Instagram has already made some steps in this direction](#) by prohibiting graphic images of self-harm. And [Pinterest has added a way to reach a suicide prevention helpline](#) in just one tap from a search or a Pinner's board.

- (5) Political Campaigns:- It has become all too clear that the misuse of social media can do great harm during election campaigns. Social media manipulation campaigns have taken place in 70 countries, up from 28 countries in 2017. Facebook and Twitter have attributed foreign influence operations to seven countries (China, India, Iran, Pakistan, Russia, Saudi Arabia, and Venezuela) who have used these platforms to influence global audiences⁴.

But a requirement that platforms should ban all political messaging could also do great harm. Where does politics end and campaigning begin – for action to combat climate change for instance? Disinformation ('fake news') is hardly a new phenomenon in politics and elsewhere⁵. What has changed in recent years has been the drastically increased levels of untrue or twisted information online which is directly accessible to billions of users⁶.

[Twitter has decided not to carry paid-for political advertising](#) and [Google has made a similar announcement](#).⁷ But such transparent and clearly owned communication is not the main problem. Indeed, shouldn't a democracy welcome such campaigning in all available media? It would also seem dangerous to expect sites to censor polite debates about climate change, for instance, or abortion – as long they as they use facts which were verifiable.

But there are problems with micro-targeting. It is surely important that we know, in a political debate, what is being told to someone else as well as being able to rely on the information with which we are provided.

This in turn leads to a separate concern that platforms can currently be paid to tell absolute lies – that a politician has done or said something that they have not, for instance. This seems wrong – but who is to judge the boundary during a fast-moving and highly charged political battle? Some

⁴ [The Global Disinformation Order, Samantha Bradshaw and Philip N. Howard](#)

⁵ Oliver Cromwell was greatly influenced by untrue stories that the 1641 Irish Rising had been accompanied by a general massacre of English men women and children, some dying out of starvation and exposure as they tried to make their way half naked towards the English-held enclaves such as Dublin. This encouraged his harsh treatment of the Irish some years later, for which he is well remembered to the present day.

⁶ [European Parliament elections: The Disinformation Challenge, Dimitar Lilkov](#)

⁷ Facebook's policy seems to be that they prohibit commercial advertisements that contain lies certified as such by authorised fact-checkers. But they don't apply this policy to political adverts.

use the word 'lie' to describe everything someone might gently take issue with. One commentator noted that:

“If I look down the barrel of a camera and say “A year has 380 days in it,” I am clearly lying, because everyone knows it doesn’t. If the Prime Minister in an interview says “We have the lowest Corporation Tax rate in Europe,” is that a lie, a mistake, an error or an error by omission? The truth is that there are four countries in Europe with a lower Corporation Tax rate than the UK. If the Prime Minister didn’t know that, he probably ought to have. He might have meant to say “among comparable countries in Europe”. He might have meant to say “one of the lowest rates...” To say with 100% certainty that he deliberately intended to tell an untruth is difficult to sustain.

In a similar vein, Adam Price, the Plaid Cymru leader wants to introduce a law to make it a criminal offence for a politician to lie. Is he really suggesting that a politician should be sent to prison, or fined, if he or she makes a campaign promise which a court finds that they couldn’t possibly have delivered on? It’s preposterous. Enough people are put off going into politics already, without a silly measure like this.”

One possibility might be for the regulator at least to require digital companies to stop accepting advertisements which spread disinformation and also make sure that this content is downgraded by their algorithms. It could also require a wider network of fact-checkers to be employed by the platforms, and require them to allow independent researchers access to private company data of past disinformation attempts in order to understand how they beat company’s algorithms.

More generally

Some platforms, though not all, will need to implement an age/ID verification service if they are to allow responsible adults access to their services, whilst denying access to certain services to particularly vulnerable users. This service should be entirely independent of the platforms, and act as an agent of their users.

Nothing in this approach creates a tortious duty – i.e. a duty that can lead to those who have been harmed claiming damages in court.

Could the platforms not be trusted to self-regulate, perhaps under pressure from advertisers? It would appear not. The tech platforms have made more than 125 announcements describing how, through self-regulation, they will solve the

manipulation of their platforms by bad actors but there is as yet no clear sign that the algorithmic changes made by platforms have significantly altered digital marketing strategies⁸.

The regulator should be responsible for deciding whether platforms are taking reasonable and proportionate steps to reduce harm to its users. Given the size of the companies which own the platforms, legislators will need to give the regulator real teeth, including a range of enforcement mechanisms. These might include licensing, enforcement orders, fines, directors' liability, and directors' disqualification.

The most important point though is that regulation is feasible and practical. There is no need to be resigned to the harms evident on social media platforms, nor to go to the other extreme and insist on the unpalatable step of requiring censorship. Neither is acceptable in a democracy, and neither is inevitable as long as regulatory measures like those suggested here are implemented.

Martin Stanley

Editor - the [Understanding Government](#) and [Understanding Regulation](#) websites.

December 2019

⁸ [The market of disinformation, Stacie Hoffmann, Emily Taylor & Samantha Bradshaw October 2019](#)