[COR0147]

Written evidence submitted by Clean Up The Internet (COR0147)

1. About Clean Up The Internet

Clean Up The Internet is an independent, UK-based not-for-profit organisation concerned about the degradation in online discourse and its implications for democracy. We campaign for evidence-based action to increase civility and respect online, and to reduce online bullying, trolling, intimidation, and misinformation.

Clean Up The Internet was founded in March 2019. There is detailed information about the founding team, members of our Advisory Board, and how we are funded, on our website: https://www.cleanuptheinternet.org.uk/about-us

After a period of research and consultation, we decided that our first priority would be identifying and campaigning for measures to tackle the abuse of anonymity on social media platforms. We recognise that there are circumstances where anonymity can be a force for good, for example when used for whistle-blowing. However, at present anonymity is frequently being abused by trolls and bullies, and to spread misinformation. As such it is a major factor fuelling the degradation of online discourse.

We've identified practical steps which social media companies could take to restrict abuse of anonymity for trolling and disinformation, whilst at the same time safeguarding freedom of expression. These steps could be implemented by the social media platforms voluntarily. However, given their continued inaction on this issue, we believe that regulation is probably required to secure meaningful progress. We believe that anonymity is an important example of the kind of “risk factor” which the UK government's propose new Online Harms regulator could require social media platforms to address, under a general “duty of care”.

We published a report in April 2020, detailing our research into the role of anonymity and pseudonymity on social media, some polling we commissioned from YouGov into public attitudes, and our recommendations for how to restrict its abuse. That report can be viewed here:

https://www.cleanuptheinternet.org.uk/post/new-opinion-poll-83-of-brits-thinks-anonymity-makes-people-ruder-online

This submission touches on similar themes and draws on the same research, but is focused upon the period of the pandemic. It is structured around the three headings, related to Online Harms, listed in the call for evidence.

2. The nature, prevalence and scale of online harms during the Covid-19 period

Social distancing and “stay at home” measures adopted during the pandemic have made the UK more dependent than ever on online communication. As such the pandemic has obviously highlighted the importance and benefits of digital connectivity, and of social media.

However, a greater reliance on online communication has also led to increased incidence of a wide range of online harms. Some of these harms, such as child abuse and exploitation, fraud and scams, fall outside Clean Up The Internet’s current areas of expertise. We will focus on two types of online harm where we do have relevant expertise, and where we have seen a pre-existing problem brought into sharper relief during the pandemic: misinformation, and abusive and divisive online political discourse.

i. The spread of misinformation

There has been an explosion of false and misleading information on Social Media surrounding the pandemic. The World Health Organisation talks of an “infodemic of false information”[1]. Ofcom found that within the first week of “lockdown”, 46% of the public had encountered false or misleading information.[2] A study by the Reuters Institute of Journalism and Oxford University analysed 225 items of COVID-19 misinformation and found that 88% appeared on social media[3].

Some of misinformation appears to have developed organically, but much appears to be fuelled by concerted disinformation campaigns by bad actors. This appears to include both nation states seeking to divert blame[4], or to undermine their geopolitical enemies[5], and extremist groups seeking to peddle conspiracy theories which undermine faith in government, or blame ethnic minorities[6][7].

The spread of misinformation relating to the current pandemic follows similar patterns to previous waves of disinformation, and in many cases involve the same groups/agendas (e.g. the Russian state, anti-vaxxers, the far-right). It has been well documented how previous disinformation campaigns exploit the lack of identity verification, and the ease with which anonymous/pseudonymous accounts can be created[8]. Networks of anonymous and deceptively pseudonymous accounts act to spread and amplify misleading information, as well as contributing to a general culture of indeterminacy and polarised debate, in which moderate and credible sources find it harder to cut through.[9]

Clean Up The Internet is currently conducting more detailed analysis of the precise role of anonymous accounts in the spread of coronavirus misinformation. We expect this to be completed within the next few days and will be keen to share this with the committee.

Misinformation is not a new problem, but the potential for it to undermine public health messages and divide society during a pandemic has brought it into sharper relief. There have been real-world attacks on people and property with 5G mobile phone masts and broadband engineers targeted[10] and an upsurge in hate crime towards the Chinese community[11].

ii. Abusive and divisive online political discourse

[1] https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19

[2] https://www.ofcom.org.uk/__data/assets/pdf_file/0031/193747/covid-19-news-consumption-week-one-findings.pdf

[3] https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

[4] https://www.politico.com/news/2020/04/21/russia-china-iran-disinformation-coronavirus-state-department-193107

[5] https://www.theguardian.com/world/2020/mar/18/russian-media-spreading-covid-19-disinformation

[6] https://www.isdglobal.org/isd-publications/covid-19-disinformation-briefing-no-3/

[7] https://www.hopenothate.org.uk/2020/04/13/anti-5G-conspiracy-theories-are-dangerous-and-spreading-fast/

[8] https://smpa.gwu.edu/sites/g/files/zaxdzs2046/f/2019-07-22%20-%20Suspicious%20Election%20Campaign%20Activity%20White%20Paper%20-%20Print%20Version%20-%20IDDP.pdf

[9] https://www.cleanuptheinternet.org.uk/post/academic-research-about-anonymity-inauthenticity-and-misinformation

[10] https://www.wired.co.uk/article/5G-coronavirus-conspiracy-theory-attacks

[11] https://news.sky.com/story/coronavirus-hate-crimes-against-chinese-people-soar-in-uk-during-covid-19-crisis-11979388

Much of online political discourse is toxic. An October 2019 report from the Joint Committee on Human rights noted appalling levels of online abuse suffered by MPs, with most MPs affected to some degree but female and BAME MPs hit disproportionately[12]. Abuse is not just directed at MPs – insults, abuse and trolling characterise much online political discourse on the major social media platforms. A 2017 report by the Home Affairs Select Committee noted that abuse is targeted "particularly towards women and minority groups". Amnesty International describes Twitter as a "toxic place for women"[13], and noted that 1 in 3 women in the UK affected by online abuse reported having changed the way they express themselves online in response[14]. Online abuse risks deterring women and ethnic minorities from expressing their views or standing for elected office.

It is often noted that political discussion online is far worse than offline. Whilst there have been some terrible incidents of offline threats and political violence, conversations between parliamentarians and constituents conducted face to face, at a surgery or a public meeting, are much more likely to be constructive. Anonymity is widely recognised as a factor here. Oliver Dowden acknowledged this in his evidence to the DCMS Select Committee on 22 April this year, observing, "clearly, if people feel that they can act anonymously, they will act in a more aggressive fashion. I certainly see it in respect of correspondence I receive and engagement I receive on social media in my capacity both as a Minister and as a Member of Parliament. Frequently, certainly in respect of constituents, when you then confront them face to face, they have a completely different attitude″

The pandemic means such "face to face" political discussion is currently extremely limited, and seems likely to remain limited for some time to come. We are therefore more reliant than ever on a digital public sphere which is widely acknowledged to be toxic. This pre-existing problem has therefore become even more acute. We believe this poses two major threats to democracy:

a) **Inclusiveness of the public sphere:** Abuse and harassment during conversations about politics on social media are disproportionately directed at groups which are already otherwise vulnerable or under-represented in public life and in democratic debates. It is clearly corrosive to democracy for the public sphere to be a place where certain groups are systematically excluded or under-represented because of characteristics like ethnicity or gender. A greater reliance on online debate risks exacerbating these problems.

b) **Quality of democratic debate:** Even amongst those who aren't amongst the groups more likely to be excluded from the public sphere, the quality of online political discourse is often far lower than that which can be achieved offline. Constructive debate and discussion are frequently derailed due to the prevalence of misinformation, trolling, and abuse. A greater reliance on online political discourse therefore risks a further deterioration in the quality of political debate – and at a time of crisis when constructive debate and effective scrutiny are more important than ever.

iii. Overarching considerations for policymakers

The prevalence of disinformation and toxic political debate during the pandemic reinforce two important considerations for policy makers in the development of future Online Harms legislation:

[12] https://publications.parliament.uk/pa/jt201919/jtselect/jtrights/37/3707.htm#_idTextAnchor066
[13] https://www.amnesty.org.uk/press-releases/toxic-twitter-failing-women-letting-online-violence-thrive-new-research
[14] https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/

a) **Large social media platforms constitute an important part of the public sphere.** It was increasing difficult to sustain an argument that this was not the case even before the pandemic. But there's now, for the foreseeable future perhaps, very little "public sphere" which isn't online. Large scale platforms like Facebook, Twitter and YouTube are where citizens will share information, discuss ideas, and explore and advocate political positions. This means the quality of information and the health of political debate on such platforms should be an important concern for policy makers.

b) **A definition of "Online harms" needs to include societal harms, as well as harms to an individual.** The spread of coronavirus misinformation has the potential to undermine critical health initiatives, and therefore to increase the length and severity of a pandemic. When someone shares a coronavirus 5G conspiracy theory on a Facebook group, there is not one obvious individual "victim" - nonetheless it is manifestly harmful. Similarly a toxic political environment, where debate is low quality, and where many already disadvantaged groups are more likely to be excluded, is not only harmful to the individual victims of specific hateful remarks. Online Harms regulation should seek to address societal harms, such as a threat to public health, or the exclusion of minorities from democratic debate, alongside harms to specific individuals.

3. Steps that could be taken to mitigate these concerns

Clean Up The Internet has so far focused its efforts on exploring one specific cluster of issues which act as a significant driver of online harms: anonymity, pseudonymity, and the lack of robust option for identity verification on large social media platforms. We will limit our comments to our area of expertise, but should make it clear that we do not believe that tackling anonymity would be a panacea. We think anonymity is a useful focus, both because of the significant problems it causes in and of itself, and as an illustrative example of how Online Harms legislation could begin to regulate the design of social media to mitigate risks and reduce harms.

i. The prevalence of anonymous, pseudonymous and unverified accounts on social media

At present a tiny number of users of the major social media platforms have had their identities verified. Twitter offers a "blue tick"[15] to an extremely limited number of users, estimated to be under 0.05%[16]. Facebook has introduced a process of identity verification for users who wish to pay to run adverts on subjects which Facebook has judged to be "political"[17]. On Twitter some users are obviously anonymous, with handles, usernames and profiles which indicate as such whilst other use a real-looking, but unverified, name which may or not be their actual name. Facebook purports to have a "real name policy", but as there is no verification mechanism this is easily and commonly flouted.

Therefore on Twitter, some users are anonymous, some are using a pseudonym, some are using a deceptive/concealed pseudonym, and some are using their real name. On Facebook, some users

[15] https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts
[16] https://medium.com/@Haje/who-are-twitter-s-verified-users-af976fc1b032
[17] https://www.facebook.com/business/help/1838453822893854

are using their real name, and some are using a deceptive/concealed pseudonym of some sort. It's very hard for the average user to judge which of these different categories another user may fall – unless that other user is one of those 0.05% of twitter users with a blue tick.

ii. The role on anonymous, pseudonymous and unverified accounts in fuelling Online Harms

There's overwhelming evidence that social media users who feel unidentifiable are more likely to engage in rude and abusive behaviour[18]. A major reason for this is that feeling unidentifiable, through use of an anonymous or pseudonymous account, makes a user feel disinhibited.[19] In one recent academic experiment, participants were randomly assigned either an anonymous twitter account or one which identified them. Anonymous participants were far more likely to create and retweet misogynistic content.[20]

There's also overwhelming evidence that the use of inauthentic accounts is an important tool for those wishing to create and amplify misinformation.[21] A recent study found that the dominance of the AfD, the German far-right party, on Facebook during the 2019 European Elections was fuelled by a "dense network of suspect accounts", with tens of thousands of pro-AfD accounts displaying "multiple features commonly seen in fake accounts though rarely in real ones".[22] As mentioned above, Clean Up The Internet is currently conducting primary research into the specific role of anonymous and pseudonymous accounts in current coronavirus misinformation, and will share this research with the committee shortly.

Additionally, if a platform lacks robust identity verification, all other rules against abuse or disinformation are less meaningful. A "lifetime ban" for a persistent abuser is easier to evade if they can simply create a fresh account with a new false name, using a new email address or phone number – either of which can easily be acquired in a matter of seconds.

iii. Steps which could be taken to restrict abuse of anonymity – without banning it

Our research so far indicates that it would be possible to significantly reduce these negative effects of anonymity and pseudonymity, without resorting to an outright "ban" - and so would protect important and legitimate uses of anonymity, such as whistle-blowing. This is because the use patterns for these benign uses of anonymity differ significantly from those abusing anonymity for toxic purposes.

Regulation should require social media platforms to demonstrate how they have designed an approach to managing anonymity with a view to minimising its abuse. Key components of that would be:

a) Offering all users the option of a robust means of verifying their real name and location

[18] https://www.cleanuptheinternet.org.uk/post/some-useful-scholarly-articles-about-online-disinhibition-anonymity-and-online-harms

[19] https://doi.org/10.1089/1094931041291295

[20] https://doi.org/10.1016/j.chb.2015.06.024

[21] https://www.cleanuptheinternet.org.uk/post/academic-research-about-anonymity-inauthenticity-and-misinformation

[22] https://smpa.gwu.edu/sites/g/files/zaxdzs2046/f/2019-07-22 - Suspicious Election Campaign Activity White Paper - Print Version - IDDP.pdf

b) Offering all verified users the ability to choose whether or not unverified users are able to interact with them

c) Making it immediately transparent to everyone whether an individual's identity has been verified or not

Users should be free to choose to continue unverified. However verified users should also be free to choose whether or not they want to hear from unverified users. All users should be able to see who is verified and who isn't, and judge for themselves what this might mean for a user's credibility. A whistle-blower, with a manifestly good reason for remaining anonymous, would continue to be able to build trust, and a following, through the credibility of their content.

iv. Likely impact these measures would have had if already implemented during the pandemic

The extensive evidence as to the negative role played by anonymity at present suggests that had measures such as these been implemented before the pandemic, they would have been likely to:

a) remove a key tool currently used by bad actors for purposes of spreading disinformation, thus reducing its spread and influence

b) reduce levels of disinhibition amongst social media users, making them less likely to behave abusively towards other users, or to recklessly spread rumours and misinformation.

c) make it easier for individual users to manage whom they interact with and therefore reduce their exposure to abuse and harassment and make fewer users feel bullied out of debates.

d) make possible more meaningful enforcement of other terms and conditions on social media e.g. bans.

4. The adequacy of the Government's Online Harms proposals to address issues arising from the pandemic, as well as issues previously identified.

It is impossible to give a definitive assessment of the adequacy of the government's proposals, given continuing uncertainty as to what the government's proposals currently are. It is however possible to offer some commentary on progress to date, and on recent statements given by ministers.

i. The original Online Harms White Paper.

Clean Up The Internet broadly welcomed the direction of travel of the original Online Harms White paper. Our response, submitted in June 2019, can be accessed here:

https://www.cleanuptheinternet.org.uk/post/online-harms-white-paper-response

Requiring platforms to mitigate the negative effects of anonymity, as we propose, is compatible with the "Duty of Care" approach to regulation, first developed by Carnegie UK and at least tentatively proposed in the Online Harms White Paper[23]. The "duty of care" approach would provide systemic regulation, with an independent regulator requiring platforms to demonstrate how they were mitigating harms through the design and management of their platforms. Anonymity and user verification would be one such design question – another might be, for example, the way a platform's algorithms prioritise certain forms of emotional or inflammatory content over more measured views.

ii. Timing

We accept that ministers and civil servants have had more urgent priorities when developing an immediate response to the pandemic. Delays in bringing forward full proposals and legislation during the current period of crisis are perhaps inevitable.

However, it's regrettable that more progress hadn't been made before the pandemic started. And it's crucial that the pandemic isn't now also allowed to be used as cover for further, more avoidable, delays. With an even greater reliance on online communication for the foreseeable future, the pandemic should if anything give the government a greater sense of urgency and make it even more important for legislation to be introduced during this session of parliament. We don't see any reason why a desire to have some form of pre-legislative scrutiny need be incompatible with introducing legislation during this session.

iii. Scope and ambition of recent ministerial comments

Oliver Dowden MP's evidence to the DCMS Select Committee on 22 April 2020 raised a couple of areas of concern. Firstly, he appeared to suggest that the main the function of a regulator would be to tell the tech companies, "just stick by your terms and conditions". Such a limited approach would not give a regulator scope to define harms or determine what a satisfactory level of harm reduction would be. Secondly, he seemed to lay quite a lot of emphasis on a binary distinction between "underage harms and illegal harms" (his "focus") and "legal adult harms" (the "trickiest area"). We think this distinction is of limited value because it obscures the complex interplays between the two, and the potential for systemic regulation to reduce both criminal and non-criminal negative behaviour. Legislation focused purely on "underage and illegal harms" would explicitly not seek to tackle disinformation, or the degradation of online discourse.

On the other hand, statements made by Caroline Dinenage MP, including in evidence to your committee on 13 May 2020, were more detailed and suggested more potential for an adequate. She has referred to the Online Harms Bill as "an urgent piece of legislation", talked about the need for a "duty of care", a "regulator with a set of sanctions", and stated that a regulator should be able to "look at the design choices that some companies have made and be able to call those into question". These would all all be key components of adequate Online Harms Legislation.

iv. Essential components of a future approach to regulation

[23] https://www.carnegieuktrust.org.uk/publications/online-harm-reduction-a-statutory-duty-of-care-and-regulator/

Given the difficulties in discerning precisely what the approach of the current government is, perhaps because it it still deciding, it's perhaps simpler to summarise what Clean Up The Internet believes some necessary components of adequate regulation might be:

a) **A systemic approach to regulation, underpinned by a general "duty of care".** A regulator should be mandated to identify harms and require platforms to demonstrate how they are addressing those harms through the design and operation of their platforms. The scope should be broad enough to include societal harms as well as individual harms; to consider design questions such as anonymity or the design of content algorithms; and to consider the role of a particular platform in the "public sphere".

b) **Strong enforcement powers.** A regulator should be given strong enforcement powers. Past experience suggests that with global companies of this size fines alone are unlikely to be effective. Recent polling commissioned by Clean Up The Internet indicates strong public support for regulation to be backed by criminal sanction[24].

c) **Improved transparency and access to data for independent researchers.** This would help ensure social media companies' performance on mitigating online harms can be independently evaluated.

May 2020

[24] https://www.cleanuptheinternet.org.uk/post/new-opinion-poll-83-of-brits-thinks-anonymity-makes-people-ruder-online