HOUSE OF LORDS

Select Committee on Democracy and Digital Technologies

Corrected oral evidence: Democracy and Digital Technologies

Wednesday 26 February 2020

3.55 pm

Watch the meeting

Members present: Lord Puttnam (The Chair); Lord Harris of Haringey; Lord Lipsey; Lord Lucas; Lord Mitchell.

Evidence Session No. 14 Heard in Public Questions 169 - 177

Witnesses

I: Professor Sarah Roberts, Co-Director, Center for Critical Internet Inquiry, UCLA; Dr Safiya Noble, Co-Director, Center for Critical Internet Inquiry, UCLA (via Skype).

Note: This evidence session was held via video link. As such, some sections were inaudible due to technical difficulties. We have contacted the witnesses to request clarification.

Examination of Witnesses

Professor Sarah Roberts and Dr Safiya Noble (via Skype).

Q169 **The Chair:** Thank you very much indeed for joining us. This is a formal questioning session, so I am afraid I have to read this rather boring police caution beforehand. The session is open to the public. It is being broadcast live on the parliamentary website. A verbatim transcript will be taken of your evidence and put on the parliamentary website. You will have the opportunity to make minor corrections for the purposes of clarification or accuracy. For the record, would you introduce yourselves?

Professor Sarah Roberts: I am an associate professor in the Department of Information Studies in the Graduate School of Education and Information Studies at the University of California, Los Angeles.

Dr Safiya Noble: Good morning. I am an associate professor in the Department of Information Studies and the Department of African American Studies at the University of California, Los Angeles.

Q170 **Lord Harris of Haringey:** Hello. We are very pleased to see you. You cannot see me but that does not matter. The first question is central to the work that we are doing in this inquiry. How do content moderation policies and algorithmic designs of major technology platforms shape democratic discussion online?

Professor Sarah Roberts: Thank you for having us this morning. I will probably address the content moderation issue and then cede to my colleague, Dr Noble, on algorithmic decision-making. Just by the nature of your question, you have indicated a key facet here, which is that those things work in concert. It is not down to one or the other.

When we think about commercial content moderation as I think about it, one of the key things to surface is the human processes that are engaged there. Of course, even when those processes are automated, the evidence of human activity, values, politics and other kinds of motivations are present in those systems. In some ways, they are rendered less tangible as we abstract into these computation tools.

To the point of your question, the issue of democracy, these systems work to create the ecosystem and the landscape which users around the world engage in. One of the important things to understand about the systems is that they are typically not tangible or visible to the user. It is conceivable and reasonable for the user to believe that they are seeing content that is somehow the best, or that is somehow there because it is meant to be there. They have not given a great deal of thought to the ways in which it has been curated to get there.

Dr Safiya Noble: One thing that we need to understand about the impact of algorithmic decision-making on democratic processes is that other platforms are primarily organised as advertising platforms. This is a really important misnomer. The public migrate to a variety of different types of platforms — a public commons, a shared commons, a knowledge commons, a news commons — but they are in fact designed to optimise their clients who pay.

When we think about the early internet narrative we think of the internet being a space of democratic participation and freedom, and it is possible to think about some of those conversations in the pre-platform age, but, in the age we live in now, algorithmic decision-making is optimised by paid advertising and those who are able to pay.

There is a grey market of people who are invested in optimisation. Those might be people with deep technical skill or who work in boutique kinds of situations where they are looking to game the system without necessarily having to pay as much as others. The research certainly bears out that those who have the most capital can influence the kind of content that comes from these systems. This is what we are seeing in modern democracies and the way in which these platforms are working now.

Lord Harris of Haringey: Could I just follow that up? Dr Tarleton Gillespie, in his book on content moderation, has suggested that it is those moderation processes — deciding what stays, what is removed and so on — that decide things. Are those moderation policies, in your view, driven solely by commercial interests, or is there an active intent — a moral and ethical component, if you like — to that, or is it only moral and ethical in so far as it will drive people away if you are not moral and ethical?

Professor Sarah Roberts: I appreciate the specific question, because this is a point where Dr Gillespie and I part ways in our belief system for what is going on on the platforms. I have argued for some time now that the primary motivation under the burden of content moderation practices on commercial sites is, in fact, the brand management process which Dr Noble alluded to, which also goes on algorithmically.

Therefore, when we see benefit to users — of course, there is benefit — the primary logic that propelled the development of this kind of activity into industrial-scale, for-pay, professionalised type of work was, in fact, the platform’s need to manage their own brand. This is a nuance that is especially important to understand in the context of the United States, which has a particular provision, Section 230, under the 1996 Communications Decency Act, which allows platforms the sort of discretion that I described. In large part, it allows for immunity from perhaps illegal content that might flow over their channels.

At the same time, and maybe even more importantly, it gave them the discretion to decide to remove something if, for example, it impacted their brand management negatively. In fact, they did not have to give any reason at all, and we might think of that and its relationship to democratic free expression as being rather contrary to it.

So in 2020, of course, the general public, and maybe more importantly regulatory bodies, are now asking the very kinds of questions you are asking, which means that companies can no longer simply and solely address their brand management issue. It so happens that, for mainstream firms, not having abhorrent content on their site or trying to remove some of it because it affects their advertisers has the knock-on effect of placating users. I do not think that is the direct audience they had in mind in the first place, especially because so many sites have propagated the myth that they exist for user free expression.

The last thing that I would say to introduce more complexity into the state of affairs today, is about the fact that in the United Kingdom, but also in the context of nation states within the European Union — both individual members and at the European Union level — there is more and more demand on these platforms that the jurisdictional norms of those places be respected in the practices of content moderation undertaken by the platforms. This is not something that the platforms have been interested in responding to, because again it removes their own discretion and means that they potentially have to set up different sets of forums for different parts of the world.

I give you the example of Germany and its NetzDG law. Now, platforms of two million users or more have to respond in a particular way in relation to their content moderation practices within the boundaries of Germany, where German law is in force. That is a different kind of state of affairs from them saying, "We will just employ our own discretion to our own profit and benefit in our relationship management with our advertising partners when we keep things up and when we remove things".

That is where we are today, but I would fundamentally argue that this practice was born out of the need for brand management. That was the primary motivating factor. I will tell you anecdotally that, when I have encountered people in the industry, they have essentially co-signed that viewpoint.

Lord Harris of Haringey: Could I pursue another, slightly different issue? Obviously we are concerned with the extent to which these things drive democratic debate and discussion in the democratic context. We received evidence yesterday that highlighted the fact that algorithms are often designed to keep people on platforms and take them down the rabbit hole towards more and more extreme content. Is that something that you recognise? If you do, how could one combat or address that? Should one combat and address it?

Dr Safiya Noble: It is a great question. Certainly, we have seen a great body of research. There is the research from Professor Jessie Daniels at the City University of New York and from Dr Joan Donovan at the Shorenstein Center at Harvard. We all see that there is a bit of use of what we will call titillating content that might have a different flavour in different contexts and at different moments. That titillating content, whether racist, xenophobic, homophobic or misogynist, is certainly the type of content that is highly engaged with on platforms.

Makers of that content are also interested in and understand content moderation practices, such that they are incredibly sophisticated in the ways that they gradually nudge people along a path. They use an ecosystem of content and make these ecosystems of racist content or misogynist content, or both, to edge people down a path. In big-tech terms, we would call that engagement. As long as there is engagement with content, it is likely to stay up. It is really when the more egregious content becomes a public-relations liability or is not caught by screeners of content and then has other types of consequence in the world, that that content will then come down or there will be a reflection upon the logics of recommending.

One of the things we should remember with the more egregious types of content is that the algorithm also depends upon users to flag content. That helps inform machine-learning algorithms. If you are circulating content in certain types of communities where people enjoy that content, it will not be flagged and it may not make it to a content moderator for some time. These are the complexities of the interplay between content moderation and the machine-learning intent of engagement.

This is one of the challenges. We are certainly seeing a lot of important research here in the United States that is signalling the mechanisms and methods that are used to pull people into content that seems like it might have a little bit of a sympathetic bent towards, let us say, white nationalism, white supremacy or neo-Nazi-oriented content. It does not start out that way. It starts slowly, and we might think of it is a grooming type of process that happens on platforms with regard to that kind of content.

Professor Sarah Roberts: I want to add to something that Dr Noble brought up in her opening remark, which is that there is a clear motive at play here. Think of the most notorious platform that has come into the public view for this rabbit-hole function that you describe; not only is YouTube known for this particular functionality and this particular issue of a rabbit hole, but it is also the most obvious platform for picking apart profit motive for why people produce and circulate this material in the first place.

On YouTube, of course, we know that there can be a revenue share for people who produce the material. It is based on the number of shares or the number of views: how popular a piece of content is. One thing that YouTube has done in the past couple of years is to go through a process of demonetisation for producers who were producing abhorrent content or distasteful content — racist, xenophobic, homophobic, et cetera — but content that is not necessarily illegal, at least in some jurisdictions.

This goes to the point you made in your question about the extent to which morality or an ethical concern is invoked in this grey area between Section 230 immunity from prosecution or liability, essentially, for anything you carry, vis-à-vis takedowns. In the space between is that appeal to morality, but a cynic might say that appeal to morality is also based on a public relations management strategy, which of course has monetary implications.

Q171 **The Chair:** This is very much, I am afraid, a House of Lords-type question, but with regard to the piece of Clinton-era legislation that you referred to — Section 230 — were there people at the time of the original congressional debates warning of this? Is this totally out of the blue, or are there people who can genuinely say, "Look, I told you so. You did not account of these issues and you are now paying the price"?

Professor Sarah Roberts: I am sure there are some, because there were plenty of us online already in that era who, by virtue of our own identities online, might have experienced a variety of harms from those identities, whether it had to do with sexual orientation, race or ethnic identity, or national origin. However, largely, if we time-track back to the mid-1990s, there are a couple of things to note when that legislation went into a hat.

First, the notion of the existence of something like Facebook was purely science fiction. It could not have been fathomed at the time. I will tell you anecdotally that I was on a bulletin-board system that was considered a smashing success; we had something like 10,000 active users at any given time. A certain limitation was imposed at that time, based on computational power, on bandwidth and on full access to computers, which were hardly ubiquitous; now, we carry around supercomputers in our pocket in the form of smartphones, which well outstrip the power of my desktop computer in the mid-1990s.

The internet was still a glorified space that had yet to be fully commercialised—and this was right about 1995-96, with the advent of the world wide web and of graphical internet, which I famously quipped to someone would never take off because we all know that the internet is task-based. I will never live it down. It was the total elasticity of those features, in essence — bandwidth, computational power, graphical interfaces — and the rise of commercial inputs beginning to colonise those spaces that meant that the problems that we have in 2020 exist in a way that in 1996 were really unfathomable.

The legislation talks about things like internet intermediaries and internet service providers, which in that era would have meant quite literally a space that looked something like a warehouse with a rack of modems set up for people to dial in and reach the rest of the internet. At best, it would have been something like America Online, Prodigy or a for-pay service like that. I would argue, and others are now arguing, that Section 230 is an anachronism. It just does not make sense in the context of the modern internet.

The other issue, of course, is that it is American. I have had arguments with legal scholars over the years who have said that Section 230 is a non-starter and that it will soon be superseded by something else because it is an American bit of statute. Now it is global and other countries will make laws that address issues within their borders.

However, I have recently become aware that the major platforms — Google, Facebook and others — are actively engaged in ensconcing language similar to Section 230 into some of the most anti-democratic kinds of agreements that exist; I am talking about trade agreements between nation states.

It has come to my attention that Google, for example, has lobbied heavily to put something that looks very much like Section 230 into a trade agreement with Japan. Just when we are starting to have a debate on this in the United States and questioning whether or not Section 230 still fits, and other legal scholars have told me that it will soon be a thing of the past, these things are getting into other kinds of realms. They are also described in a more surreptitious way, which is in the cultural fabric of the makers of these products in Silicon Valley, who then package up their value systems and export them to the whole world. I hope that was not too professorial.

The Chair: Not at all. I found it fascinating.

Q172 **Lord Lucas:** What do best practices in moderation look like? Do moderators have enough time and contextual information to make reliable decisions? Do the labour conditions that moderators face impact the decisions that are made? Can it be right that moderators are, at times, required to sign NDAs, the effect of which is to encourage suspicion of malpractice?

Professor Sarah Roberts: Thank you for these incredibly pertinent questions. To put it simply, the moderation practices as they have come to exist in the commercial space — I call that commercial content moderation — have grown up largely as an afterthought to industry. They have been not a profit centre but a cost centre for firms. They have been thought of as the unfortunate necessity and not something that was given a great deal of forethought, which means that things like best practices are largely ad hoc and haphazard.

Also, just as Professor Noble described those algorithmic processes by which companies develop and keep user interest, so too are content moderation processes treated as trade secrets, so rather than get all the heads of operations into one room from the top five — the FAAMGs, as they are often called — and say, "Let's broker a best-practices mandated charter around content moderation, which will benefit the labourers", everybody is off doing their own thing, and because this is seen as a significant source of cost and not upside revenue to firms, they are constantly chasing sites of labour that will provide the labour at the lowest cost.

It should come as no surprise to members today that those are places that tend not to enshrine the rights and well-being of workers. In fact, it is quite the opposite. The Philippines, of course, is one of the major players in commercial content moderation, and they have an entire government ministry dedicated to soliciting business to business from transnational corporations, on the grounds of the fact that there have been only three strikes in a decade in special economic zones in the Philippines. That does not bode well, in other words, for the workers who labour under those conditions.

Your question about the time allotted for decision-making is also important. I often think of the old adage, "Time is money", and in this case it truly is. When I was in the Philippines in 2015, I interviewed some workers there who talked about the fact that, when they had begun working as contract labourers in the call-centre environment there, they had something like 32 seconds to evaluate and respond to a given piece of content. Essentially, they were making rather low-level decisions and the outcome was going to be one of two things: leave it up or delete it.

In a sense, the decisions were not profoundly sophisticated, but it took evaluation and it took perhaps looking through contacts or other kinds of clues to indicate if there was other value in having a piece of content up. I might give the example of the famous Vietnam-era photo of a young child burned by napalm, which was deleted not because of the obscenity of violence towards children but because of child nudity — it is an absurdity, right? — because someone in the Philippines who did not have the context looked at it for some period of seconds.

Thirty-two seconds already seems like a pretty short period of time, but, as these workers informed me, during their period of being on this particular contract, which was at most for two years, the amount of time they were allotted to make a decision went from 32 seconds to 10 to 15 seconds. If we do the math, that means that what they were asked to produce was essentially doubled. It meant that they were not allowed to give as much time and thought to a decision that they were rendering. Another way to look at it is that their wages were essentially halved in terms of productivity. The discipline that was meted out around this change, which in other places might have resulted in a work stoppage or some kind of action, was, "If you do not do this, this contract is going to leave the Philippines and go to India, where people will do the work for less money and they will produce more".

This issue of productivity metrics and how they come to bear, both on the quality of decisions, which is a user-facing issue, but also how that might affect workers, is key here. Again, if we think about the orientation that companies that solicit this work have towards it as being trade secret—your point about disclosure agreements comes into play again here—it becomes very difficult to obtain and ascertain truthful information about the system in which this operates.

In the last couple of years, in no small part due to regulatory pressure and other forms of pressures, firms have become a bit more open about this, but many of them, because of the scale of the amount of content needed to be adjudicated, are using a patchwork approach to getting the work done. They will have people under a particular set of labour conditions in their branded headquarters in Silicon Valley, or perhaps in Dublin or Barcelona, who labour under a given set of parameters and conditions. However, just as in the textile industry and other industries, as we move further and further away from the locus or point of origin of the content which companies have solicited, the more difficult it becomes to control the labour conditions.

Cynics such as me argue that that kind of distancing is by design, because when there is a gap, or a PR issue, or for example a factory collapse in Bangladesh in the textile industry, it becomes apparently more plausible for the soliciting firm to say, "We didn't even know our garments were manufactured here". There is a similar orientation to this outsourcing of distance organisationally to content moderation.

Dr Safiya Noble: I just wanted to add one layer to this. When we talk about vulnerable communities, in some cases we might be talking about ethnic minorities, immigrants and refugees — people who might be part of societies where these platforms are being used. We often see that the moderators do not understand the social context of the kind of content they are looking at. Something that might be blatantly racist in the West, for example, might be imperceptible to them in their given context and make it impossible for them to adjudicate. In Europe or the United States, we might have a clear understanding of the sophisticated nuance of that content and what it means, but it is imperceptible in other cultural contexts to people who are not oriented to or familiar with those histories or that type of communication.

Of course, it is explicit when we look at visual images in particular. I have been stunned to see content moderators not recognise what we would probably characterise or classify as blatantly racist or xenophobic propaganda, but just adjudicate it as being under the auspices of free speech.

This is where we also need to have a lot more nuanced specification of content, because this is about the level of training and depth of knowledge. Even in the country of origin where it might circulate, people might not fully understand what they are looking at when they are looking, for example, at stereotypes or other types of material. It is very important to mention this as well.

Lord Lucas: In some bits of the internet that I inhabit, particularly Twitter, moderation seems to get weaponised as a means of deleting views you oppose. Indeed, one of our colleagues in the House of Lords is currently banned from Twitter for this sort of reason. Are the sorts of systems that are being operated capable of resisting that, or will they inevitably be pushed so that they become, at least for one side of the argument, a suppressive system?

Professor Sarah Roberts: I appreciate the question. I will just share with you the results of a study by Sarah Myers West, a scholar who looked at some 500-plus cases of individuals who had had content removed from mainstream social media platforms, and she had assembled those users to find out, for example, how they felt about their content being removed. For my purposes, the most fascinating result from that study was the fact that, of the more than 500 individuals who had had material removed, presumably from every aspect of the political spectrum and from all walks of life, almost every person felt that they were being personally persecuted and targeted due to their political beliefs.

To some extent, there is an element of that baked into the platforms, in part because they have operated for so long under the auspices of offering all people access to fundamental, unfettered, democratic, free discussion, and they have been reneging, in essence, on that promise by putting an asterisk beside them and saying, “We are going to go ahead and adjudicate things at our discretion” — Section 230 — “to try to make it a more hospitable environment”. Each platform has developed a capacity for tolerance that looks different from the others, so in some cases you can do things on YouTube that you cannot do on Facebook.

However, is it a characteristic fundamentally of platforms let us say to encourage polarisation to this extent that might result in takedowns or being banned and so on? To some extent, I think it is a fact that that is baked into the platforms. I always have to say, before we get out the handkerchiefs on behalf of the platforms and say how difficult a problem this is for them, we have to remember that it is content that they have allowed to be posted [inaudible] that has led to this.

Hany Farid, who is a computer scientist at Berkeley now, developed a product by which there can be some automated removal of content in the important context of child sexual exploitation material. He wrote a paper he discussed the fact that no matter what is under discussion in terms of remediation around behaviour online on platforms, what never seems to be discussed by the platforms is limiting or in any way impeding upon the amount of content that is being uploaded endlessly, 24/7, or doing any more significant vetting of the people who are uploading and the reasons for which they are doing it.

The business logic that is at play is set up to lead to a certain series of outcomes, and there are other kinds of outcomes that are permanently under discussion. That has led to the logic of it starting to seem almost natural or incontrovertible or undoable. I would argue that it is none of those things. We would have to go much further back in our process and question the platforms and the decisions that they have made in order to get at some functional alternative outcomes.

I always say to students that there is a real upside here for someone who wants to come to the marketplace and offer a platform that is heavily moderated by sophisticated, taste-making moderators whose work is visible, which you can then debate or take under advisement. There are other people who want to be online and not be threatened with rape or homophobic or racist violence or to experience ethnic slurs to themselves and others.

It seems, however, that other notions have taken hold. I would argue that it is much simpler in that case to have a principle of “free speech” rather than one that has set very clear parameters and then constantly and tirelessly works towards enforcing them, earning the ire of a certain percentage of the populace.

Q173 **Lord Lipsey:** It is a fascinating description of moderation that you have been giving. It strikes me that it is rather like a football match in which one side is allowed to pay the referee, decide what rules the referee will apply and fire the referee if they do not like what he is doing.

That leads me to ask if you could speculate on alternative systems of moderation. There could just be a regulator who appointed the moderators, or there could at least be a regulator of moderators who looks over the rules that are employed — the terms and conditions, and so on — to see whether they allow the job to be done. Leaving it to the firms to moderate does not seem to me to be fulfilling the public purpose, even if it is quite good for the profits of the firms concerned.

Dr Safiya Noble: Certainly, there are people who are arguing now for alternatives and what some refer to as public-interest technologies or public-interest platforms, which are owned by the public and might have a different value system or logic where a profit motive may not be primary. This is where we think back to a Web 1.0 type of environment, in fact, where we did not necessarily have the commercial dimension of those approaches.

In some spheres, that may be of value. In the domain of things like search engines, for example, which the public relate to as vetted, rated, public commons of good information, those are indeed nothing but advertising platforms that can be optimised in a 24/7 live auction of content. That also intersects with social media, so that information or propaganda that might move powerfully through social media and up higher levels of engagement might be the index for your search engine, and it is credible. The interplay between different types of platforms is very important. As Professor Roberts suggests, there is very little communication or interplay between companies to think through the complexities of that.

We must talk about alternatives. No matter what these platforms are, it is the public's perception of these platforms that really matters and the way in which users engage them. In fact, I met a librarian a year ago who said, "I thought Google was a non-profit". I know, it is shocking.

Think about the variety in the spectrum in which people are engaged across platforms. For example, when users become aware that there is disinformation through social media, they immediately turn to the search engine and fact-check. Of course, this, again, is about the interplay between search engines and types of platforms. We should be thinking about that.

From a regulatory perspective, we are talking today about our platforms completely undermining analytics and predictive analytics that are used in major platforms also for a vast system of companies that are not well known or whose brand names are not well known, and where predictive analytics and approaches create tremendous public harm.

It is true that we would never let guys rent some space in a strip mall, make chemistry experiments, manufacture a drug, roll it out at the local drugstore and then nationwide or internationally, let people die or be damaged, and then say, "Maybe we should have thought about that before". But that is, in fact, what we have in the tech space, where people are allowed to make all kinds of predictive analytics and a vast array of different technologies that have zero oversight, for the most part, and it is not until after the fact that researchers like me, Professor Roberts and others can document the harm and regulators start getting interested in these conversations.

In the United States, we have the Federal Trade Commission, which looks at things like consumer harm. Certainly, that body needs to be advocated in the US. We have other federal agencies, such as the Food and Drug Administration, and other types of bodies that oversee pharmaceuticals or the quality and health of our air and water and so forth, and these might be models that regulators could look to.

Professor Sarah Roberts: To go back to the football analogy, not only do they set the rules, with the referee and so on, but they own the pitch and they call the game "football" but maybe it is another game. From top to bottom, this is completely optimised and owned by the firms themselves at every stage. One thing that is apparent to both of us and to many of our colleagues is that, now that these companies are finding themselves at the centre of public concern and regulatory inquiries, they are of course professing a desire to reform, but their ideas for reform, just like that football match, again have to do with all these things: identifying the people who ought to be involved in that, in setting up ethics in AI and other kinds of boards that are bringing in some very familiar characters who are very close to the firms themselves. In other words, I describe it as the fox guarding the henhouse.

It is, in fact, the same position of an ecosystem where the logic of the platforms prevails in remediation. We also have to be very sceptical of claims which all sorts of firms are making; we have to ask by whom and to what end. We have been concerned as we have watched what amounts to ethic-washing, just like greenwashing in relation to environmental concerns in an industry.

Q174 **The Chair:** Sarah, can I just make an observation? Last weekend, I saw a film called "The Cleaners", which I think you are interviewed in; you are very good. What was extraordinary to me when watching that was that one girl who was interviewed said that on her first day she saw an eight year-old girl being abused, and she had to stop and report it. All the guy did was to show her contract to her and tell her to go back to work.

My point really is this: once you have got over the shock of seeing your first eight year-old abused, when you have seen your thirtieth and fortieth eight year-old abused, your ability to decide whether it should be deleted or remain becomes dramatically altered. I come from the movie industry and I do know this stuff, yet what does not seem to be taken account of is that the sheer repetition of watching abuse alters the individual’s sense of what is or is not permissible. That is just an observation. Incidentally, the film is still only available in German, which is extraordinary.

Professor Sarah Roberts: That is absolutely correct. Just as a bit of background, I was the scientific adviser on that film and it is based on much of my research, but that is it; it is the vision of the directors. I know it intimately, but I too find something new in it every time I view it. Your point is something that I brought up in my book. If you do not have copies, we will make sure that those are made available to you.

On the point about the outsourcing of the commercial content moderation industry, in many cases in the United States, moderators were term-limited. Some of the workers that I spoke to in Silicon Valley were allowed to work only at a particular site for two years. Presumably, they worked for another company but they were limited at the major tech firm where they moderated to a term of two years.

There are only two plausible reasons for this. One is because, about three few years ago, Microsoft was very heavily sued and lost because it was employing so-called contractors who were classed as full-time, long-term employees, so the term limitation had to do with making it clear that these people were not long-term employees. Of course, why would they be? [inaudible].

The other reason is the one that you bring up, which is the efficacy of content moderation after a period of time. There are only two outcomes for workers who have done this for any significant period of time. One is that they are no longer good at their job because they are traumatised by what they have seen, and they do see those things every day, over and over again. That is a worrisome dimension, especially when you think about the nature of the workers and how they are just put back into the workforce and into the social fabric without any longitudinal support or follow-up.

Perhaps even worse, at least to my mind, is the point that you made, which is that we may be producing a legion of individuals for whom viewing the sexual exploitation or abuse of a child, or an unedited blog with a bunch of footage from a warzone where individuals are harmed, or any number of horrors, becomes quotidian, and those people become desensitised and inured to that particular type of content.

Of course, for the purposes of commercial content moderation, that renders them less effective, but there is an even bigger concern here about their psychological well-being going forward as they cycle back to society. One woman I know who used to work for Myspace, when that was a big concern, has returned to her work life as a bookkeeper. She has nothing to do with the tech industry. When I met with her, she had been out of it for a good decade. She said to me, “It was about three years after leaving Myspace before I could shake a stranger’s hand when I met them”. I said, “What do you mean?” and she said, “I have seen what people do. People are disgusting”. She just said it matter-of-factly.

We ought to be concerned collectively about both types of changes: the trauma as well as the desensitisation, which in essence is another form of trauma that is going on in the industry, and the lack of aftercare and the lack of longitudinal studies.

Q175 **Lord Mitchell:** Good morning. What role could civil society play in improving content moderation? Can we learn any lessons from Facebook and the third-party fact-checking network that could be generalised more widely to content moderation?

Dr Safiya Noble: Civil society has been playing a huge role in content moderation for a long time. The foundational framework for moderation stems from community guidelines that govern all kinds of pre-Facebook platforms, even loose federations, organisations and communities that are online.

We need to go back to the old days, to 25 years ago, when we were on the internet and trying to make sense of information. It is probably worth now pausing for a moment. Sarah often shares that it is important for us to nuance the word "content", because "content" is a word that really flattens out a lot of things that are happening on the internet. It makes it increasingly difficult to discern what is knowledge, information, fact, fiction, propaganda and disinformation, because it all gets flattened into this word "content". I just want to bookmark that, and maybe you will want to come back to that at some point.

If you think about that framework for the kind of information that moves around and circulates on the internet, you will see that subject matter experts play an incredibly important role in different kinds of information systems. Pre-Google, in the early days of trying to discern and find expert information and content, librarians used to play a significant role in organising information into very sophisticated decision trees, with the clustering of different forms of knowledge and their interplay and relationships. People who were subject-matter experts would hang out in chatrooms and communities with each other and share information and deeper knowledge; novices could enter those and ask a question of these experts, who might be teachers, professors, hobbyists or people who were self-educated in a topic.

That is a very different environment than the one we currently occupy, where there is so much capacity for knowledge, information, evidence or propaganda to flow through any one of the systems, from search engines to Facebook, that it becomes much more difficult. You have the same terrain when it comes perhaps to the BBC using Facebook to share views and information right alongside white nationalists, and those have equal visibility on some of the platforms because of the engagement with that content.

Of course, even people who are not interested in comparable content often look at it because they want to be aware of it, so even the intent of our interaction with content is not well understood by the platforms. It is just branded as engaged with and therefore it should be made more equal.

These are the kinds of challenges that we are looking at when we think about the public’s role in moderation and in engaging with information on the internet. These are valid questions that really require our time and our consideration. We have so many deep pools of expertise. You have some of the finest universities and schools in the world there in the UK, and those are bodies and places where deep knowledge, expertise, information and evidence can emanate from. When that type of knowledge is up against other types of propaganda, let us say, that becomes a threat to various publics.

Of course, we have nations in the not-so-distant past and currently that understand, for example, how the circulation of propaganda is a public health threat because it gives the conditions for things such as ethnic cleansing, genocide or harassment and targeting of various communities. So we may have to disaggregate what we are talking about when we talk about content, and then think about whether it is appropriate to force that public total responsibility for what happens in the platforms, which unfortunately is the current model; Facebook, YouTube and Google in its various companies rely upon us to flag content and to trigger awareness. Of course, that is uncompensated and incredibly hard work. What happens is that all kinds of things get flagged, not just egregious content.

Professor Sarah Roberts: We should to look at the funding of mainstream social media platforms, which have been a source of information of a variety of types and quality levels for the past decade and a half, and map that against that for public libraries. I know that, unfortunately, the public library system has been massively depleted over the past several years in the UK as well. If you were to map things up, we would see a massive rise in participation, while we see funding decrease. It is not that the public somehow decided that we ought to de-fund libraries because there is less interest in being informed.

There was a suggestion that there would be a one-to-one replacement — something like the Google search engine — for a library. In fact, the CEO of YouTube, Susan Wojcicki, is making me nervous, because she goes around the world making this pronouncement that “YouTube is a library”. It is important to note that, on this library called YouTube, there is no librarian. There is nothing that the user can interact with to get guidance and ask about the veracity of the origin of a particular piece of so-called content, which, again, I find to be, as Dr Noble indicated, a rather self-serving bucket that the platforms use.

Sometimes I think about solutions that exist in this space, especially when we are concerned about information quality and veracity. Maybe the platforms are not placed to cite the solution to the problem they themselves created. Would it not be extraordinary if we called upon these platforms to give back to communities and to entire society by offering some of their immense profit to some of the institutions that exist in the public but which have been systematically defunded. I suggested that to one CEO of one of these firms. So far, that has gone nowhere, I am sure you will not be surprised to learn.

Lord Mitchell: Just as a follow-up to that, do you think that Facebook's third-party fact-checking is better or worse than others?

The Chair: We will accept a yes/no answer.

Dr Safiya Noble: This attempt at fact-checking is important, of course. It is acknowledged that there is the potential and, quite frankly, a lot of evidence of the amplification of disinformation that moves through Facebook, particularly around election cycles such as the upcoming presidential election in the US. There is an incredible amount of attention on this and these attempts to flag. Google has also been engaged for many years in trying to flag or note where problematic or false content might be showing up.

The question a step before that is: how is it that we live in an information environment where false information can reach millions of people with the speed and scale that platforms allow for? Again, questions about fact-checking and the models of how, for example, news content is adjudicated and the veracity of it are political questions, quite frankly. I will leave it there, just to say that there might be other mechanisms outside platforms.

Professor Sarah Roberts: It is not quite clear where that kind of duplicitous or problematic information has produced this engagement. So it does not seem to be necessarily the bullseye of where the intervention ought to happen.

Q176 **Lord Lipsey:** Supposing I was head of Facebook, God came to me some morning and I determined that none of the algorithms run by my company were going to discriminate against women in future. What do you actually do to change the algorithm? How does this get built into the algorithm in the first place?

Dr Safiya Noble: If I had the answer to that, I would be a billionaire. It is very difficult for us to answer these kinds of questions, because these algorithms are trade secrets. Unfortunately, the only way we can even arrest what is happening on platforms is when researchers, for example, are able to document job ads that are not shown systemically to women, or housing ads that are systemically not shown to people of colour—this kind of micro-targeting of content or ads in particular. We cannot see inside these firms and their logic.

We do know the outcomes, however. They are regularly in the headlines, and the public are increasingly aware. I have spoken with a number of tech workers, for example, who have never considered that there might be federal, national or state discrimination laws on the books that should be considered. In fact, it is a novel idea to consider extant legislation, for example.

Part of this is about the lack of depth of knowledge about policy and protections and responsibility in the public sphere. There are legal requirements that all businesses must comply with. Again, this is where we cannot underscore enough the importance of Section 230, because the platforms put the onus upon advertisers who might be running targeted discrimination against women, for example, rather than the platform itself to be responsible for adjudicating whether in fact that is illegal activity. The range of the illegal activities is vast across the world and in different parts of the world.

These are some of the challenges that we are facing, and we are talking about a volume of content that is of a magnitude that is unfathomable for most people. This is one of the reasons why these platforms are so [inaudible].

Professor Sarah Roberts: I would add that, rather than thinking about the algorithm itself, which as Dr Noble said is very difficult for anyone outside the platform to understand, we might also think about calling some of these businesses to account for their practices, for example in relation to an advertisement. Perhaps we ought purposefully to say that before you are allowed to do that, or maybe now that you have done that, that there is evidence of discrimination or harm in those processes. We need an audit of those processes. We need an audit of microtargeting—by "we" I mean perhaps the United Kingdom—or an inquiry into microtargeting advertisements that determines whether or not the practices as offered by the firms are likely to comply with or contravene the law in the United Kingdom.

Rather than focusing on the algorithm, we might focus on the judicial status quo, ask that the legislation be enforced and put the onus on the company to write algorithms that will meet that particular onus, rather than coming up with "anything is possible" and then being surprised when laws are broken and other country norms are contravened.

I was in the room when this incident was described. We had this engineer in an AI lab who wanted a philosophical debate about what might constitute fairness. His project was to [inaudible]. He and I looked at each other and said, "You could just attend to the federal government regulation around what constitutes fairness [inaudible], because there are such things. In fact, your company is responsible for them, so [inaudible] want to do, that is more important [inaudible].

Q177 **The Chair:** Thank you very much. We are in danger of getting cut off from you, so I suggest is the following. You have already answered question 5 in a sense, and if there is any aspect of it that you feel you would like to add to, please drop us a note — and similarly with question 6.

I would like to jump to question 7, if I may. If technology platforms could do one thing to improve their recommendation algorithms, what should it be? If government could do one thing to regulate it, what should that be?

Dr Safiya Noble: The logic of recommending algorithms is predicated not only upon engagement but on helping the companies to make money. My vantage point on that is that maybe a different set of values or priorities come into play beyond just what will be most profitable for the companies. That is a really difficult challenge. Facebook certainly has a report on it, which was commissioned by Sheryl Sandberg and has many recommendations in relation to the protection of civil rights and civil society that could probably be taken on.

On the question of what government could do to regulate platforms, we might pull back a bit to sectors rather than thinking about companies. I will give you an example. In the United States, our economy was predicated on big tobacco and big cotton. We had a full egregious set of labour relations in the United States for that industry, and we also had deep investment and ideas from big tobacco that tobacco was good for you. In fact, an ad used to run that four out of five doctors preferred a Camel cigarette. I am sure the doctor had a cigarette hanging from his mouth when he delivered me.

It is a paradigm that has shifted and which we cannot imagine any more, but it cost harms to three generations or more of people in this world—harms that came from that industry and its practice and product, quite frankly, and it had its own investment in research to prop itself up. We need to pull back, quite frankly, look at the sector and say, "One hundred years from now, we will be differently oriented about the harms that came from the sector. Yes, there was the incredible economic boom globally from it, not unlike other industries that have led to huge economic boom. Yes, it would be difficult to shift the paradigm and reimagine differently, but that also can be done".

The question is what role you could all play in helping to create a paradigm shift rather than asking platforms which, if it is not too harsh to say, are deeply implicated in the collapse of democracy and in anti-democratic processes around the world. To fixate on tweaking an algorithm is mislocating and misdiagnosing the problem.

Professor Sarah Roberts: One of the issues at play is the fact that, until very recently, these firms themselves have been able to self-define as so-called tech companies. It is not just an existential turn of phrase; it has deep regulatory implications. If they were, for example, broadcast media companies, they would have to be responsive to hosts of regulations that exist and which other media firms must comply with.

When I think of something like a recommendation algorithm, I think about the ways in which children are endlessly exposed to a bevy of material that is frankly just advertising. In most countries, there are prohibitions on what kind of material might be accessible to children and to what extent they can be advertised to. The industry players that are self-defining outside the confines of traditional media have really been able to skirt much of the regulatory apparatus that has existed for the most part of the 20th century.

There is a way in which we might look at extant regulatory bodies and their mandates, and think about how we want to strengthen and apply them, because the notion that these so-called platforms exist somehow in a state of exception has worn its welcome. I am going to leave my comments there.

The Chair: That is very good of you. We have taken a huge amount of your time. I have a question to ask Sarah, if she can help me. I am in touch with the people who made "The Cleaners" and am trying to get hold of a clean version without German subtitles or a German voiceover. They have been very nice but a push from your end might help a great deal.

Professor Sarah Roberts: Yes.

The Chair: You also mentioned the doctors smoking, and you are absolutely right that there were lots of ads. I teach and I use little clips from movies. There is a scene from one of the early episodes of "Mad Men" where a gynaecologist is inspecting a woman, and he is smoking a cigarette. When I run that clip, people suddenly get it: "I see. The world has changed".

Dr Safiya Noble: You play such an important leadership role in the world with respect to tone-setting, and one of the things that we witnessed in our research is watching the regulatory bodies and policy leaders who drank the Kool-Aid. They are thinking about how to optimise or perfect rather than paradigm-shift. We are living in a moment when it is worth considering the body of evidence that is advancing and how many more information health crises we can endure, and how much our Governments and our societies endure that over the long haul.

The Chair: You have been very helpful and we hope you will like our report. If you would follow up with some written thoughts on the questions we did not get to, I know we would all be very grateful.

Dr Safiya Noble: We would be happy to.

The Chair: Thank you very much, both of you, honestly.

Dr Safiya Noble: Thank you.

Professor Sarah Roberts: Thank you so much.