



Select Committee on Democracy and Digital Technologies

Uncorrected oral evidence: Democracy and Digital Technologies

Tuesday 25 February 2020

11.35 am

Watch the meeting

Members present: Lord Puttnam (The Chair); Lord Black of Brentwood; Lord German; Lord Harris of Haringey; Lord Holmes of Richmond; Baroness Kidron; Lord Knight of Weymouth; Lord Lipsey; Lord Lucas; Lord Mitchell; Baroness Morris of Yardley.

Evidence Session No. 13

Heard in Public

Questions 163 - 168

Witnesses

I: Dr Jennifer Cobbe, Co-ordinator, Trust and Technology Initiative, University of Cambridge; Christoph Schott, Campaign Director, Avaaz; Alaphia Zoyab, Senior Online Campaigner, Avaaz.

USE OF THE TRANSCRIPT

1. This is an uncorrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.
2. Any public use of, or reference to, the contents should make clear that neither Members nor witnesses have had the opportunity to correct the record. If in doubt as to the propriety of using the transcript, please contact the Clerk of the Committee.
3. Members and witnesses are asked to send corrections to the Clerk of the Committee within 14 days of receipt.

Examination of Witnesses

Dr Jennifer Cobbe, Christoph Schott and Alaphia Zoyab.

The Chair: Welcome to all of you. I am required to read out a caution, like a police caution; I had never thought of it that way. As you know, this session is open to the public. A webcast of it goes out live and is subsequently accessible via 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. If you would be good enough to introduce yourselves, we will then go to the first question.

Christoph Schott: Good day, everyone. It is an honour to be here. I am a campaign director at Avaaz, the biggest global citizens' movement across the world. I am here and therefore missed my four year-old daughter's birthday yesterday. I thought that maybe the best birthday present I could give her is being here and caring for democracy in the future that she will grow up in. I am honoured to be here and I hope she will be proud once she understands what I am doing.

Alaphia Zoyab: I am a senior campaigner at Avaaz. I have worked on our disinformation campaigns for the last year and a half, focusing particularly on hate speech in Assam, as well as investigating other aspects of disinformation. I have two kids, whose birthdays I did not miss.

Dr Jennifer Cobbe: I am a research associate and affiliated lecturer in the department of computer science and technology at the University of Cambridge. I work on law and regulation of new and emerging technologies, such as internet platforms, machine learning and automated decision-making. This topic is very much part of the things I have been working on for quite a while.

The Chair: Just to note, it is my birthday today.

Alaphia Zoyab: Happy birthday.

Q163 **Lord German:** Mr Schott, I am daring you to ask this question of your four-year-old daughter. I wonder if you could all tell us how the technology platforms are using recommendation algorithms to shape what users see. Perhaps you could distinguish between content focus and collaborative focus, and how that makes a difference. I am sure you do not actually talk to your four year-old daughter about this.

Christoph Schott: To simplify, the algorithm is the platform and business model of social media. As there is so much content out there, they are curating the information for each of us in a very specific, personalised way. To give one statistic, YouTube says that 70 per cent of the time you spend on the platform comes through it recommending videos to you. You might be searching for terms like climate change, and then you get drawn into YouTube, which recommends more and more of

what the algorithm deems interesting for you, to keep you on the site for as long as possible.

The main goal of pretty much all social media platforms is to keep you on their site for as long as possible, so they can collect as much data as possible about you, show you ads and make as much money as possible. That is where you see the danger: the goal of social media platforms might not always align with the goals we set for ourselves. It might draw more and more on the simple emotions, such as outrage and hatred, that we see going viral and spreading far and wide.

Right now, what scares me is that there are companies in the world that can reach over two billion people every single day with different targeted messages, so nobody knows what everyone is seeing. Giving any individual such power, without any regulation in place or any boundaries for how the content, especially its distribution, should be treated, is a real danger for democracy. That is why we are here today: to present solutions and ideas for how the recommendation algorithm can be made more ethical and smarter, so that it contributes to public good, not public menace.

Alaphia Zoyab: In the research that Avaaz has undertaken, one of our methodologies is to try to follow the algorithm. For example, if we have tracked disinformation on certain pages, by liking those pages we continue to follow the algorithm to see what rabbit hole it takes us down.

Adding to what Christoph has said, we are guessing what the algorithm does. A few whistleblowers from some of these tech platforms have told us what they do, but we do not actually know how they really behave, and I suspect neither do the tech platforms. That is the worrying aspect for us.

Dr Jennifer Cobbe: I would echo what Christoph has been saying. We have to consider these algorithms in the context of the business model of these companies and how they use them to drive profit. Fundamentally, the recommender systems play two main roles in the business models of these companies. The first is delivering behavioural advertising to users, which brings direct revenue to the companies. The second is presenting content to users to drive engagement and keep people on the platforms, so that they can be served with more advertising. You have direct and indirect revenue sources through these algorithms.

Your question talked about the content-based and collaborative-based filtering. Content-based algorithms are where the user has watched video X and might also be interested in videos Y and Z with a similar title. Collaborative filtering is where user A has watched this video, and similar users B, C and D have watched this video, so user A might like that video. That is how those two systems work — at a very high level, of course.

When it comes to the way these are used, looking at them at a non-technical level is more useful. We divide the way they are used into three

different forms. The first we call closed recommending, where the system recommends only content produced by the platform itself. If it is a news website, it just recommends its own content. The second kind we call curated systems, which are commonly used on platforms such as Spotify or Netflix. They deal with content that has been selected, licensed or chosen in some way by the platform itself, so there is a degree of editorial control over content. The third kind we call open recommending, where user-generated content is brought in without any editorial review. It might include content produced or licensed by the platform, but the distinguishing feature is the presence of user-generated content without editorial review.

Most of the problems for democracy and the public sphere are emerging and developing in open recommending. If you allow anybody to upload content to this system, to be disseminated really broadly, you risk that being gamed and used by people to spread disinformation, conspiracy theories and extremism.

When we talk about this stuff, the content itself is not the problem. If a conspiracy theory video on its own is seen by 10 people, that is not a public policy issue or something that we need to worry too much about. It becomes a problem when it has a larger audience and is put alongside a lot of similar content. Audience and context are really important for this and recommender systems give you both those things. Open recommending in particular takes user-generated content, potentially disseminates it to tens of millions of people, which gives you the audience, and places it alongside similar or related content on the platform, which gives you the context.

Those two things work to reinforce the potential harm caused by some of this content. Given the capacity of open recommending to amplify the audience and provide that context, I would argue that that is the real problem.

The Chair: I get a note each week, because I forget, reminding me to speak more slowly, because I am making life impossible for the person doing the transcript. You will never meet them, but they will be very grateful if we all speak more slowly.

Lord German: In the open area, which you think is the most difficult, is it possible to identify easily when that is happening? In which case, is it easy for people to flag that as something observable?

Dr Jennifer Cobbe: Generally speaking, yes. We can easily talk about which platforms use open recommending. Facebook newsfeed is open recommending. Twitter timeline is open recommending. TikTok, one of the newer ones, is open recommending. YouTube is open recommending, as are quite a few others. It is easy to tell, because if they include user-generated content, and if you can upload stuff and have it disseminated by the recommender system, it is an open system. The distinguishing point between open on one hand and curated and closed on the other is easy to spot. Yes, we can easily tell what they are using.

Lord German: Do you agree that this is the distinguishing mark between other forms of recommendation algorithms?

Alaphia Zoyab: Yes. That is the problem. For instance, our last study was on YouTube. We set out to discover what percentage of the recommendations and videos YouTube spits out when searching certain key words are disinformation. Our last study showed that even when putting in a neutral search term like "global warming", 16 per cent of the related videos YouTube threw up in the search results were clear disinformation. We would recommend that clear disinformation be pulled out of the recommendation algorithm, so that people are not exposed to that content and companies like YouTube, and the creator of that video, do not profit from clear disinformation. That is where the problem lies. These recommendation algorithms are not curating to take out disinformation.

Lord Lucas: Who makes the choice as to what is disinformation and not information?

Alaphia Zoyab: That is a great question, which we are always asked. The policy Avaaz is proposing, which we call "Correct the Record", relies on third-party, independent fact-checkers to determine what is false, verifiably false, partly misleading, harmful, et cetera. The determination of what is and is not true should not be left to Governments or tech platforms. We should leave it to professional fact-checkers.

Once third-party, independent fact-checkers have determined something to be misleading or partially misleading, the onus must be on platforms to correct the record, in the same way that a respectable newspaper would issue a correction if it made a mistake, or a TV channel might issue a correction if it aired something misleading. The onus must be on the platform to correct the record, which is not happening at the moment, at least not at scale. Avaaz has campaigned for this for the last year and a half. The platforms under pressure are finally moving now, because they see there is a category of legal content that can be extremely harmful.

Some of the most obvious examples in the recent past have been health disinformation campaigns, such as anti-vax or disinformation about the coronavirus. There is a lot of information that is harmful to democracy as well. We are really clear that it should not be in the hands of government to decide what is and is not true.

Lord Lipsey: You have gone some way to answering what puzzles me the whole time. Let me try to put it in a simple way so that even I can understand the answer. The algorithms themselves are mathematical models, so they say, "Lipsey, 71, is not likely to be very interested in dating sites".

At the same time, the platforms themselves have an interest in promoting certain kinds of content. It may be lying content that draws people in: flat earth, to take one. That is not done by the mathematics of the machine; it must be done through some form of conscious

intervention in the mathematics, by the companies themselves. Which is creating the problem? Is it the algorithms or companies manipulating the algorithms to promote their own interests?

Dr Jennifer Cobbe: They cannot really be distinguished. Algorithms are not neutral, purely mathematic things. Their design, deployment and use are inherently imbued with the values and priorities of the people who are designing, deploying and using them. When we ask, "Is it the algorithm or the platform?", they are two sides of the same coin. Platforms prioritise engagement with the goal of growing market position, revenue and profit. Because they prioritise engagement over anything else, anything controversial, shocking, extreme or divisive will get people's attention. That is a function of these algorithms used in this context for the purposes that the platforms have set for them. You cannot really distinguish between the two things.

Lord Lipsey: But there is no outside intervention or even transparency in the algorithms. Whatever the company chooses to feed into the mathematics comes out in the recommendations the algorithm makes.

Dr Jennifer Cobbe: There is not necessarily any outside intervention, although some people try to game these systems by using certain kinds of content or using bots to manipulate the metrics that underpin this. Fundamentally, whenever companies design algorithms, they design how the algorithms work and determine what they prioritise over other things. Their values and priorities are encoded in the algorithm, and from there into the platform more generally.

Lord Lipsey: That is helpful.

Christoph Schott: In most cases, the algorithm has a specific goal of engagement increase. Disinformation actors know that. For example, Facebook admits that, right now, it has over 100 million fake accounts on its platform and likely over 200 million duplicate accounts, so 10 per cent to 15 per cent of the accounts on Facebook are not of real people. They likely have a high activity. If 15 per cent or 20 per cent of the people in Britain had a very specific wish or idea about certain issues, imagine how much they would draw out the public debate. That is the way fake accounts are able to game the algorithm; by teaching the algorithm that specific things like outrage, hatred and false information are supported by many people across the world, which is not true. The algorithm is biased towards engagement and is gamed by actors who put up so many fake accounts to influence its behaviour that it gives people not what they want but what the disinformation actors want.

Q164 **Lord Knight of Weymouth:** From the answers we have had, I am interested in thinking about what should be regulated. A lot of the debate has suggested that we should regulate content and have fact-checkers that we decide are independent, so this is their job. Should we be regulating algorithms?

Alaphia Zoyab: The short answer is yes. In the context of disinformation, which is the work that Avaaz does, we are interested in

the distribution of disinformation. That is the piece of it that needs to be regulated, rather than every little piece of content. All our uncles are free to share whatever they want on social media, but the point at which it becomes viral is when we need intervention from a fact-checker. We need Government regulating to make sure that is happening.

To take an example, I know the Committee has had testimony from the editor of the *Yorkshire Post*, talking about the picture that it was claimed was false. When we investigated that, we found that just four Facebook accounts, over two days, had shared it 40,000 times, despite a fact-checker putting out full facts and saying, "This picture is real. It happened". There was no intervention and no onus on Facebook to correct the record. That is the point at which we want to see the distribution of disinformation regulated, which relies on the algorithm.

Lord Knight of Weymouth: In your model, the independent fact-checker flags a piece of content and then the regulator regulates the algorithm, because that is the distribution mechanism.

Alaphia Zoyab: The regulator says to Facebook, "This piece of content has gone viral. It is clearly disinformation, as the fact-checker has pointed out. This needs to be taken out of your recommendation algorithm". Our model protects free speech, so it is not taken down, but the algorithm is not offering that piece of disinformation free publicity, so it has to be taken out of the recommendation algorithm. If you search for it, you can find it, but it is not being boosted.

Christoph Schott: The key piece is that social media has the obligation to go back to all the people who have seen the false information and show them the corrected piece. The idea is not to say, "You have been lied to. It was wrong". It is to say, "Fact-checkers have found verifiably false claims in the news you have seen and here is a corrected piece from them". It is just to provide more information, rather than less. For example, over the last month, we have been working with George Washington University and Ohio State University to set up a social media polling tool, where people first saw disinformation pieces and then saw the corrections. Among those who saw the corrections, belief in the disinformation was cut in half.

People initially believed that Trump said, "Republicans are the dumbest voters ever", and then saw a correction that he had never said that, after which they were half as likely to believe it. We have seen that corrections work, but right now the problem is that a piece of false information goes viral, 10 million people see it, fact-checkers fact-check it after three days. They put up the fact check and nobody sees it, because Facebook does not show it to anyone. It does not go viral, because facts are often not as sexy as lies. The idea is that, to give the truth a chance to catch up with the lies, we say, "Facebook, if there is a viral piece of information and a verified fact-checker has said it is untrue, this correction also needs to be shown to these people who have seen the false information".

Lord Mitchell: In the case of the boy on the hospital floor in the heat of

an election, I am not quite sure how it can be regulated. When something like that has gone viral, the genie is out of the bottle. What is your thinking on that?

Alaphia Zoyab: Scale and speed with fact-checking is a problem and a challenge. That example was published in a reputable newspaper and a fact-checker said that it was real. If Facebook notifies users as they are sharing something over a two-day period, the genie is out of the bottle, but at least there are some corrective steps. At the moment, it is a complete wild west. Anybody determined to lie and use fake accounts to boost disinformation can have a free run. That is a problem.

Dr Jennifer Cobbe: On your point about content versus algorithms, part of the problem with regulating content is that at that point we are not really regulating platforms, we are regulating individual speech, which has a whole host of human rights problems. Particularly with content or individual speech that we think might be harmful but is not illegal, if the Government want to regulate that they should try to make it illegal and see whether that can survive challenge in the courts. I suspect that it probably would not. That is a serious problem with regulating content.

We can regulate algorithms. To some extent, there is a freedom of expression concern in deciding that certain communications should not be disseminated as far or as widely as other communications. Fundamentally, if you leave the content on the platform, where it can be searched, browsed for or shared by other people but it is not disseminated by the algorithm, it does not bring the same freedom of expression issues as intervening to remove it. As I said before, audience and context is the real problem for a lot of these things, rather than the content itself.

Lord Lucas: Is a suggestion that the fact check be available alongside the original the right way to do this, so people can make their own decisions rather than some person saying, "This should not be distributed, this should be downgraded"? Make platforms allow people to see both next to each other, so they are aware that there is a fact-checker who says, "This is rubbish", and can choose whether to read or believe them.

Christoph Schott: Yes. Generally, the big problem is that virality is super-fast. Within 24 to 48 hours, most of the sharing has been done and most of the harm has happened. Fact-checkers need time, because it is a hard job to fact-check something very well. After two days, 10 million people have seen it. It is very unlikely that those 10 million people go back to the person who shared the original post, where they liked it. If you just do what Facebook is already doing, as you suggest, and add the fact-check information to the old post, you may reach three per cent of the people who have seen the false information.

We see a massive gap there, where those people will never find out. We did research during the yellow vests protests, where verifiably false information that was fact-checked was able to reach about 100 million views. There are roughly 30 million people on Facebook in France, which

means that, on average, every user on Facebook in France has seen three pieces of that information. A very small percentage of them have seen the fact check. They all likely still believe that images of women bleeding on the streets of Paris, which were actually from Spain, were true, or that Macron was dancing in the Middle East while the protests were happening. They have never been informed, so their trust in democracy or institutions might be eroded bit by bit.

Our research shows that, had they seen the correction, at least half of them, which is a really big percentage, would have started to think, "Maybe that is not actually true", and been more critical. This is one of the best media literacy ideas out there: to go back to people who have seen false information and show them that it was not true. They will become more aware that this is happening.

Alaphia Zoyab: I want to underscore what the algorithms are doing right now. If I am a peddler of lies and I want to launch a disinformation campaign, the algorithm is letting me have free publicity. The more triggering and outrageous the content I post, the more it is going to boost it. It is not just the principle of letting people decide, but why is verifiably false information, debunked by fact-checkers, getting free publicity through the algorithm? That goes to the heart of the business model: they want to keep you on the platform for as long as possible.

We are saying, "Do not take down the content, but just take it out of the algorithm". If people search for it, they can find it, so you have free speech, but you do not get publicity with demonstrable lies. I also want to cite a study that YouTube did. It did a test in the US, taking out junk conspiracy theories from its recommendation algorithm. The views on those videos dropped by 50 per cent. Small tweaks can change the health of the information ecosystem.

Lord Lipsey: I want to come back to fact-checkers. You rightly said that it takes time to fact-check. I wonder if anybody has any sense of the scale of these things. I helped to found Full Fact, which is the most prominent British fact-checker. Last time I looked, we had 26 employees, half of whom were working on automated fact-checking, which is a promising role for the future. If you could check six facts to standard in a week, you were doing pretty well in that kind of organisation. When you compare the size of that with Facebook, Google, all these things, it is trivial. The number of things they could hope to fact-check under the present system is nearly nil. Somehow, someone surely has to tackle the problems of the resources of fact-checkers.

Christoph Schott: I will answer this question as best I can, because it is the right one to ask. First, fact-checking would have to be scaled up massively, if that was a requirement. We do not yet have the one solution for it, but there might be ways for that to be funded. Secondly, from what we have seen, there is a lot of disinformation out there, but the number of pieces that go viral is not as crazy as the amount of disinformation out there. If you could fact-check 100 pieces in a week, that would reach 10 million people, and you could go after the most

important stuff and start inoculating people by showing them that they have seen false information.

There is another step that we are exploring. I do not know if you have heard of the company NewsGuard. It is not rating false information but is rating news sites, websites, fake news websites, against nine criteria. Most of them are former journalists. They give ratings based on whether journalistic standards are adhered to; where they are not, it is mostly peddling misinformation. An independent group, like those journalists, could rate pages and websites that are peddling misinformation. You go one step further, because the content is harder each time, but after a while you build up a reputation. Highly reputable sites, such as the Guardian and the BBC, would likely have a high rating. That is something we are developing. You could work with social media to integrate that in how the algorithm makes decisions.

We have been talking about human intervention in the algorithm. Right now, the algorithm is built by humans and is making decisions. We are saying, "No, the machine has decided that this piece of content is okay to go viral". There should be an intervention, because that is as least as dangerous as thinking about wise ways to intervene or give boundaries to the set of algorithms that make the choices.

Q165 Baroness Morris of Yardley: You made a statement before about digital literacy. You said that when you told someone what the truth was, they adjusted their beliefs, so they took note of that. Do you know how long that good scepticism stays with them? Next time they see something that is not accurate, are they more likely to hold back and think, "I wonder whether this is true"?

Christoph Schott: We have not run a public study on that yet. We have done some with our own membership. Yes, it stays with them. They are better able to distinguish true from false. I can speak only for myself. Once I started to understand, I stopped clicking on content from Russia Today, for example, which I might not have done in the past because I did not care that much about the source. Now I always check the source. I believe that people who understand this issue will be more prone to doing that, but we do not have any studies on it.

Dr Jennifer Cobbe: We have answered two questions in one. The first is about the scale of these platforms and the challenge of fact-checking them. "Scale" is a very interesting word, because platforms and tech companies use it as a way of getting funding and investment, and building a profile. The second we try to do anything about potential harmful effects of these companies, they turn round and say, "We are too big now. We are at scale. We cannot do anything about it". They use it as a sword and a shield. My argument would be that if these platforms are too big to do this kind of stuff responsibly, they might just be too big. We need to think about addressing that problem at a structural level rather than tinkering round the edges.

The second point is on inoculation. Colleagues in Cambridge, Dr Sander van der Linden and Jon Roozenbeek, have done research on building a web browser game that helps to inoculate people against disinformation. They have done studies and peer-reviewed stuff on this. It seems measurably, significantly, to reduce perceptibility to this kind of disinformation. They have done work with tech companies, which I do not necessarily want to name, to roll out similar things in their platforms. That is potentially a route towards addressing those problems in at least some way.

Baroness Kidron: You have all touched upon this in certain ways, but it would be great to say directly, on the record for us, whether you feel that recommendation algorithms actually cause new behaviours or whether they just exacerbate behaviours that those users would otherwise display.

Dr Jennifer Cobbe: From my point of view, they generally exacerbate problems. We should not pretend that disinformation, extremism, conspiracy theories or radicalisation exist only because of recommender systems. There are deep-seated socioeconomic and political causes for these things. But these algorithms can cause feedback loops, where people who are susceptible to certain kinds of information get served up that information, which reinforces their beliefs, and it becomes a self-reinforcing cycle. It also works on content producers. It is not just on the side of the consumers. People who want to reach a bigger audience, either because they want to get more revenue or because they want to reach as broad an audience as possible, can have the same effects. They begin to create content to meet what they think the algorithm will disseminate more broadly.

We have multi-sided feedback loops going on. It is not about just the consumers. It is a multidimensional problem, but the algorithms are absolutely part of that problem. They are one of the dimensions.

Alaphia Zoyab: I would like to cite two studies. The first is a terrifying study done by Facebook in 2014, which tested the concept of emotional contagion. They tested it on 689,000 users: "If we flood one set of people's news feeds with positive news, do they respond by posting more positive stuff?", and similarly with negative news. That study would seem to suggest that you can make people feel a certain way.

On the exacerbation point, we detected that, taking anti-vax as an example, you may start off being sceptical and wondering whether it is true, and then be led down a rabbit hole. Avaaz has developed some of the methodology in our research by following the algorithm. The algorithm has always led us down a rabbit hole of a filter bubble.

Christoph Schott: Ahead of the European elections, we did major research into disinformation networks on Facebook. So you can see the scale, we got Facebook to take down some of those pages, which had racked up over three billion views in a year. We flagged many more. One main tactic they used was to start movie pages or book club pages. Slowly but surely, once they had enough support and membership, they

started to put in content that was more hateful, false or supporting one political party over the other. We have also found that, once you follow one of those groups, you get recommended similar groups.

In this case, someone who was interested in music might have been drawn into a far-right hate group environment and stayed there. Because of the algorithm but also because of how it is gamed, people who start with a specific issue or interest can be drawn into something that might be related, but might also be much more extreme.

Baroness Kidron: I was really interested in what you said: if they are too big to manage it, maybe they are just too big, period. I am interested in the other side of that. We are often told that making an intervention helps the incumbents. By the way, I thought your paper was fantastic. Have you have thought about whether it privileges the incumbents — Facebook, Google, et cetera — to have a system where they have to follow up bad news with good news; or, more accurately, bad facts with good facts? Have you thought about how that will affect people coming into the marketplace and whether it is affordable

Dr Jennifer Cobbe: Putting additional requirements on platforms for this kind of stuff could privilege the big companies. That is very possible. I said before that this is multidimensional problem; we need multidimensional solutions. Just regulating algorithms is not going to be enough. We also need intervention at a structural level in the tech industry, from the point of view of competition and many other aspects. If we are aware that this might privilege those companies, we can construct competition interventions in a way to help address that.

We should note that these algorithms, because they drive for engagement, are key parts in driving the growth and dominance of those platforms anyway. There is more than one thing going on here.

Christoph Schott: Speaking freely, when the EU was discussing the digital tax, it also examined new, smaller businesses. The obligation was mainly for bigger players in the market. Also, the idea is that the fact-checkers would be independent from the platform, so if they issue a fact check on a specific issue all the platforms could take the fact check and send it to their users. For smaller ones, there should not be an extra hurdle to step over.

Q166 **Lord Lucas:** Should we have an official body in the UK charged with analysing these algorithms, looking at their effects and making that public, so we can all participate in the discussion of whether they do what we want them to do?

Dr Jennifer Cobbe: We need some kind of oversight body. I would not want to suggest what that would look like and who it should be, because there needs to be quite a lot of work looking at existing regulation, oversight bodies and frameworks, and what does and does not work, perhaps in other jurisdictions. But we need some kind of oversight body, whether it is a new regulator or one of the existing regulators. Of itself,

that will not solve the problem, because a lot of regulators are already underresourced. For example, the ICO is massively underresourced for dealing with data protection issues. We could put this on an existing regulator or construct a new regulator, but either way they need to be given significant amounts of money and resources to address this. For specifics, I would not want to say.

Lord Mitchell: How do programmatic advertising and recommendation algorithms interact? Does this cause legitimate advertisers to find themselves unwittingly funding material they would never wish to be associated with? I am very interested in the connection between it all.

Dr Jennifer Cobbe: On the second question, whether advertisers end up unwittingly sponsoring content they do not want to sponsor, that happens. For example, we have seen advertisers withdrawing from YouTube after discovering that their adverts have been placed alongside extremist, far-right material. YouTube has created a set of policies for demonetising, as it is called, videos and channels with extreme material. That happens and it is a valid concern.

When it comes to programmatic and recommending crossover, a lot of the time the dissemination of the advertising comes through the same systems. If you look at your newsfeed and you see sponsored posts, that is disseminated through the recommender system in the same way as any other content.

On the programmatic point, there is a major question mark over whether that is legal at this point. We should probably wait and see on that one, but it interacts with the recommending quite closely.

Christoph Schott: To add a data point from our study on planet misinformation on YouTube, we found about 100 advertisers running ads ahead of misinformation on climate, which was denying climate change and saying, "It is a hoax. The scientists have not got it right". We called up many of those brands, from Samsung to L'Oréal, from Greenpeace to WWF, and none of them was aware that it was shown on those videos. They likely had put in something like "climate change" as an interest, so people who are interested in seeing videos about climate change see their advertising, but they were unable to opt out of misinformation videos. They were unwittingly funding them.

This is one of the main problems: 55 per cent of the money from advertising on YouTube goes to the creator. The creators behind the misinformation are making money out of sharing this misinformation with many people. There is a monetary incentive to share this information, because it has so many of the characteristics of information that is shared far and wide. That is the danger. Big brands are unwittingly supporting climate deniers, who in many cases likely did it on purpose

Lord Black of Brentwood: We have talked a lot this morning about transparency in algorithms. Indeed, the Committee has had a lot of evidence about it. It would be interesting to know whether you see

disadvantages in increasing algorithmic transparency. For instance, would it degrade the experience for users, or simply allow spammers and others to come in and manipulate the algorithm and search results?

Dr Jennifer Cobbe: It strongly depends on the kind of transparency you are talking about. If it is transparency to users, you can very quickly overwhelm users with information. That essentially becomes a waste of time, because you are not giving them anything meaningful. If it is transparency to oversight bodies or regulators, there is less risk of people using that information to game the system, because they would not necessarily have access to it.

Transparency is useful, but talking about transparency generally does not necessarily move us forward. Transparency to regulators, and perhaps some useful transparency to users, could be a positive development.

Christoph Schott: Some of the most sophisticated disinformation actors have more possibilities to understand the algorithm than the normal user. They can test many types of content and understand how it works. Yes, we believe we should not put out millions of lines of code for everyone to see, but the idea would be to have trusted, independent experts with clear oversight and understanding. I am not from Britain, so I cannot speak specifically about the role of Ofcom. Having the platforms submit quarterly reports on how many fake accounts they have, and what they do about disinformation, will give more transparency. It will also give you and us more information and understanding of the best policy choices to make.

Alaphia Zoyab: As a user, at the moment there is no explainability as to why I have been recommended something over something else. If I am being sent down a rabbit hole of disinformation, there is no explainability in the process as to why that is happening. A degree of transparency is important, both in auditing the algorithms using certain metrics and in explainability for users.

Lord Black of Brentwood: How might we produce this level of transparency and oversight by regulators? The commercial success of these companies is built on the secrecy of the algorithms, and they are going to fight tooth and claw to protect it. Presumably, you see no other way than legislation, which would be quite complicated, to force them to produce greater levels of transparency. They are not going to do it as a result of any voluntary code. Perhaps I am being overly pessimistic.

Dr Jennifer Cobbe: The black-box nature of these systems is a choice on the part of the companies, for the reasons you say. There are intellectual property and commercial secrecy reasons for keeping these things as secret as possible. Any company should be willing to open up its algorithms to regulators for this kind of investigation, if it can be confident that that is not going to be passed to competitors. That seems like a reasonable thing to require of them.

You are right that it needs some kind of legal duty, because they are not going to do this voluntarily. If they were going to, they probably would have done it already. We need some kind of statutory or legal duty on these platforms. That could be tricky to do, but not insurmountable, and it is necessary.

Baroness Kidron: On that point, I am interested in how much commercially sensitive transparency you need to the regulator. Could you look at the path or the impact? Could you look all around the black box and not necessarily challenge the IP?

Dr Jennifer Cobbe: Generally, when it comes to transparency of algorithmic systems, the model is incidental to a large extent. We need to know what is coming in and out. That is something that we can absolutely legislate for.

Alaphia Zoyab: Regulators routinely see commercially sensitive information. We do not think social media companies should somehow be exempt from that degree of scrutiny.

Christoph Schott: There is a specific piece that you can see coming out. Our work is focused on the harm that comes out of the algorithm. But it is a glass of water and you do not see what percentage of the accounts is fake. Only if you could see inside it would you see the mixture and content of the algorithm. At the moment, only the platforms can do that. Not even they can; otherwise they would act more strongly. Some factors are impossible to see from the outside. I am not an expert on transparency regulation, but there could be a way to ensure transparency on specific issues, related to democracy, that might be causing harm and that we need to understand more deeply.

Baroness Kidron: I speak for myself, but the Committee would probably be very interested in what that list is, from your point of view, if you would care to give it to us after this session.

The Chair: On the issue of transparency, the written evidence from Avaaz says that 77 per cent of citizens across the United Kingdom "think social media platforms should be regulated to protect our societies from manipulation, fake news and data misuse". First, where does that statistic come from? It chimes quite interestingly with another piece of written evidence we got last week. YouGov suggests that 87 per cent of the population believe there should be a legal requirement on political advertising.

Alaphia Zoyab: The 77 per cent statistic is from a YouGov poll that Avaaz commissioned last year. Ahead of coming today, we asked YouGov to run the poll again in order to see the difference in sentiment between last year and this year. We have seen a significant jump and we will share those numbers with the Committee. There is increasing evidence that the public absolutely want regulation. They want to see corrections in their news feed. They believe that their trust in social media would increase significantly if there was fact-checked, verified information and they were notified. We have seen the numbers move from the high 70s

to the 80s. The public mood is shifting towards a greater demand for transparency from social media platforms and requirements for regulation.

The Chair: This is quite hard-core evidence for us. It is really important.

Lord Lucas: Would you be prepared to share the full YouGov stuff with us? Sometimes the methods of respondent selection are pretty dodgy.

Alaphia Zoyab: Yes, we will share all the questions with you.

Lord Lucas: It is the methods, so we know whether it is a percentage of the population or just of the people who could be bothered to reply to YouGov.

Christoph Schott: You just touched on the numbers. Considering it is a widely run poll, 81 per cent support regulation on social media to ensure prevention of harm; 84 per cent support the idea of correcting records or issuing corrections if they have seen false information; 87 per cent think that platforms should have responsibility and must be held accountable if they recommend false news to millions of people.

Lord Lucas: These are really important figures, but 87 per cent of whom? Is it 87 per cent of the real population or some subset that is self-defined in some way?

Alaphia Zoyab: When we commissioned the study with YouGov, we said specially that it had to be a nationally representative sample.

Lord Lucas: Sharing with us the details of how that selection was done becomes very important.

Alaphia Zoyab: Yes, absolutely. It is broken up by Labour or Conservative voters. We will share the full details as opposed to just the top-line numbers.

The Chair: The point Lord Lucas makes is very important. If we are to convince government that this shift is taking place, we need to be very clear that our own evidence base is adequate.

Q167 **Baroness Morris of Yardley:** There seems to be evidence that some groups, maybe by gender or ethnicity, are biased against in the algorithms. I know that is the way they are built up, but what can be done to reduce this amount of bias? Where does that responsibility lie and what actions might be taken?

Alaphia Zoyab: I will speak a little narrowly on the subject; maybe you could take the broader question of bias. Avaaz did a study on hate speech in the state of Assam in India. Publicly, the Government in India were saying that they were doing a citizenship count. We looked on Facebook to see what government officials were saying. We found an absolute tsunami of hate speech against the minority community that was deemed to be illegal in the state of Assam.

Algorithmically, we found that the platforms were failing that particular group of people, because they had not built algorithms and fed datasets to those algorithms in the languages in which it operates. For instance, one major reason for the failure of Facebook to track hate speech in Myanmar is that it did not have enough staff who spoke the language. Crucially, the algorithm had not been trained with the datasets to understand what hate speech looks like in Burmese.

We found that to be the case in Assamese as well. The algorithm had a blind spot. It just was not picking up hate speech in Assamese. There was a double whammy for the community that was targeted by hate speech. They did not have the resources to flag it online. When you do not flag it online, the algorithm does not learn, so it does not flag it either. A whole community was left as sitting ducks and targets of hate speech online. That was the clear blind spot that we saw, in terms of algorithmic bias, when it comes to language.

Dr Jennifer Cobbe: It is important to understand that machine-learning systems almost inherently have some kind of bias in them. It is exceptionally difficult, if not impossible, to fully eradicate bias from these systems. Computationally, it is exceptionally difficult. Mathematically, there are different formulations of bias and fairness that are contradictory. If you take one particular way of trying to deal with this, you can make another worse or not really address it. It is very difficult.

Fundamentally, these issues come from a number of causes that are outside the algorithm. At the heart of all these systems is a model that has been trained on a large training dataset. It could be millions or hundreds of millions of records. If that dataset is not broad enough to be representative of all the conditions the system is likely to encounter in future, you will have gaps in that. You will prioritise the things it has been trained on and deprioritise the things it has not been trained on, so you have gaps there.

You are looking at historical datasets. All algorithms are inherently trained on data about the past. Even data you collect in real time is data about the past, by the time it gets into the system. Structural issues in society that may not even exist anymore, but may exist, can become encoded into the model because the dataset reflects those underlying issues. There are other ways that this could be done. If the system designer simply has not tested or audited the system properly or widely enough to look at all the potential biases and issues, you can get this stuff coming through.

The risk of bias in these systems is very high. It is impossible to remove it entirely, but it can be minimised. That is something that can be addressed and we need to be looking at.

Baroness Morris of Yardley: When there is evidence that that is the case, because it does seem to have happened, what is the incentive for the people controlling the algorithm to change it? If they just leave it as it is, their financial model will still run and they will still get the income.

They might get a bit of flack. What needs to be put into the system to incentivise them to change?

Dr Jennifer Cobbe: We need legal, regulatory incentives around this. If all they care about is profit — and it is all they care about — there is no reason for them to change this.

Baroness Morris of Yardley: There has to be a heavy hand.

Dr Jennifer Cobbe: There has to be something else that forces them.

Lord Harris of Haringey: You described how the bias occurs, but is the bias occurring wilfully or by accident?

Dr Jennifer Cobbe: It could be either, or both. Ultimately, the fact that it is there is a problem that needs to be thought about. If it is there by accident, so by omission rather than deliberate, that is still because they have not gone in and actively audited or tested their systems properly.

Lord Harris of Haringey: Any control mechanism on that would have to be essentially saying, “The legal obligation on you is to audit what happens with your algorithm”.

Dr Jennifer Cobbe: Yes. Auditing and testing as broadly as possible, making sure that your datasets are as representative as possible, are all things you can do to help eradicate this. Even where it is not deliberate, it could just be overlooking things, but if you have overlooked those things you have done something wrong.

The Chair: So I understand this, a passive unconscious bias can be ignited by the very things we are talking about here. You have passivity turning into activity, as it were.

Dr Jennifer Cobbe: Yes.

Baroness Kidron: To pick up on your very good example of where a whole language is absent, I want to ask you briefly about third-party flagging. We have huge global companies making absolute fortunes; we have a community of flaggers, who are volunteers a lot of the time; and sometimes we have fact-checkers. I want you to speak to whether we have a problem, because they will be from parts of the world and particular communities that already have strength, so we are going to see a bias in what is checked. Is that correct?

Alaphia Zoyab: Yes, that is correct. Right now, there is no legal obligation for example on Facebook when it enters a new market, operating in a new language, to have requisite systems in place, where it can check for straight-up illegal content. A lot of what we detected was straight-up hate speech that should be taken down, by Facebook’s own community standards. Yet over a couple of months we found ourselves writing letters to them over one or two posts. There is no requirement on them, right now, to make sure their algorithms are unbiased. There are no audits.

The algorithms are being taught bias by an extraordinary extent of fake activity. If you have fake accounts boosting a certain piece of content, it is teaching the algorithm to be biased in a certain way. The only people who know the extent of that bias, examining the extent of fake activity, are the platforms. Right now, we are in a situation where they are judge, jury, witness, defence and prosecutor, and nobody knows what the problem really is.

Baroness Kidron: I am sorry to give you so much homework, but I would be really interested if you would point us to what you think the audit should include, because that is another very practical thing that is interesting for us. Thank you.

Lord Knight of Weymouth: To be really clear, is it realistic to ask platforms to weed out fake accounts?

Christoph Schott: Is it possible for them to do so? We believe it is, in many cases. Facebook has said it has taken down, I think, 3.3 billion fake accounts in the last quarter. That is more than half the world population. It was mainly new accounts that were trying to sign on, so they catch most of them when they sign on. The problem is that they have 120 million active on the platform already, which have likely been there for years, so they get more sophisticated.

In a new study, for example, about 25 per cent of the climate denial activity on Twitter is from likely bots. There are lots of ways to track what is a likely bot or fake account, to the extent that you can be 95 per cent to 99 per cent sure. We believe that the platforms could do a similar job and then, for that small percentage, find ways to make sure that this is a real person. Right now, they have no obligation. This is really bad for advertisers. You go on Facebook and say, "I want to target 1,000 people in east London". There is a good chance that you target only 900 people and 100 of them are fake. There are studies showing that you can target more people in a specific area of the US than there are people living there.

Lord Knight of Weymouth: The advertiser's content could be made by a bot as well, so it is just bots advertising to bots.

Christoph Schott: Yes, but money is being spent by someone real.

Q168 **Lord Holmes of Richmond:** It has been an excellent session. It seems somewhat trite to try to reduce this to a grain of sand, but I will do it anyway. If technology platforms could do one thing to improve their recommendation algorithms, what should it be? If the Government could do one thing to regulate it, what should they do?

Christoph Schott: We have a hard time with one thing.

Lord Holmes of Richmond: Treat yourself to two, as there are so many birthdays in the room.

Christoph Schott: There are three key principles that we have talked about. One is to detox the recommendation algorithm, so strip out the disinformation from them. Do not delete it, but do your best to strip it out. The second is to correct the record. Issue fact checks from fact-checkers to people who have seen or interacted with disinformation. The third is to ban fake accounts and make an obligation for platforms to do that. We are pretty sure they will find better ways to do it than they have now. The last point is to have stronger transparency and oversight through independent experts, to understand more broadly what is going on.

Alaphia Zoyab: This is one thing governments can do. A concern for us in the UK is that disinformation is being treated as something that you can deal with through media literacy. The tactics we have researched suggest it is very hard for the average person to understand when something is disinformation and when they have been targeted by a disinformation campaign. We would urge the Committee and the Government to make sure that there is a statutory code governing the distribution of disinformation, so we do not have to get into ministry of truth stuff. It is just the distribution of disinformation that must be regulated.

The Chairman: That is very helpful.

Dr Jennifer Cobbe: One of the most useful things platforms can do is to actively address this problem and try to down-rank this kind of stuff. Some platforms have begun to do that, but not all. From the point of view of Government, the most useful and achievable intervention would be to make the right to use open recommending conditional on showing that you are actively making progress towards reducing the dissemination of this kind of stuff. Open recommending is not the only way to personalise services for users. It is not the only way to disseminate user-generated content. There are other ways of doing this. If platforms cannot show that they can do it responsibly, they should not be allowed to do it and should be made to use one of the other ways.

The Chair: Thank you very much indeed to the three of you. It has been a very good session.