



Select Committee on Communications and Digital

Corrected oral evidence: Freedom of expression online

Tuesday 2 March 2021

3 pm

[Watch the meeting](#)

Members present: Lord Gilbert of Panteg (The Chair); Baroness Bull; Baroness Buscombe; Viscount Colville of Culross; Baroness Featherstone; Baroness Grender; Lord Griffiths of Burry Port; Lord Lipsey; Lord McInnes of Kilwinning; Baroness Rebuck; Lord Stevenson of Balmacara; Lord Vaizey of Didcot; The Lord Bishop of Worcester.

Evidence Session No. 13

Virtual Proceeding

Questions 114 - 123

Witnesses

I: Alan Rusbridger, Member, Facebook Oversight Board; Kate Klonick, Assistant Professor of Law, St John's University, New York.

USE OF THE TRANSCRIPT

This is a corrected transcript of evidence taken in public and webcast on

www.parliamentlive.tv.

Examination of witnesses

Alan Rusbridger and Kate Klonick.

Q114 **The Chair:** Welcome to our witnesses for this week's evidence session in our inquiry into freedom of expression online. Our witnesses are Alan Rusbridger and Kate Klonick. Alan is a former editor of the *Guardian* newspaper and a member of the Facebook Oversight Board. Kate Klonick is assistant professor of law at St John's University. Thank you both very much for joining us and for giving us your time. It is very generous of you.

I will ask you to briefly introduce yourselves and tell us a bit about your respective roles and your thoughts on the issue we are investigating, which is freedom of expression online, from the perspective of your brief. When we have done that, we will go around members of the committee, inviting them to ask further questions. We will try to accomplish all that in about an hour. Alan Rusbridger, would you like to go first?

Alan Rusbridger: Thanks for having me along. I edited the *Guardian* for 20 years, I chair the Reuters Institute for the Study of Journalism in Oxford, I sit on the Committee to Protect Journalists and, for the last year, I have been on the Facebook Oversight Board. I joined that because, although we could spend the next hour being very critical of almost everything that social media companies do, I still have a view that they also do things that are for good. Sometimes they are the only platforms for people who would otherwise be voiceless or who live in repressive states. It is very easy to see this argument through a western prism. It is very easy to reach for regulation, but that begs the question of what regulation would look like in Turkey or Pakistan as opposed to America or the UK.

I witnessed these companies, which are built by very talented engineers, floundering around with issues of ethics, free speech and the kind of decisions that editors make every day of the week, and thought that, if I could try to help them think in a more sophisticated way about content, that would be a valuable thing to do in a world in which none of us really wants Mark Zuckerberg to be making those decisions. Those of us who have studied 300 years of free speech are quite nervous about Governments making decisions about content. That is why I joined it and, so far, I am finding it highly absorbing.

Kate Klonick: I am an assistant professor at St John's School of Law in New York. I have spent the last five years doing a type of ethnographic or journalistic research on what these companies—Twitter, YouTube and, most recently, especially Facebook—are doing to govern online speech. I call it governance, because really what I found is that five years ago the community standards that Facebook posts now, which we are all very aware of, were not even publicly posted; they were not posted until 2018. Part of what I was working on is to uncover what those rules were, how they were enforced and how they had a recursive nature of changing policy through enforcement—in a common-law-type approach—of facts to law.

In doing that, one of the things I found out was that, yes, this is a type of governance system for freedom of expression online. It is transnational. It operates within and between all these countries. This was five years ago. I saw it pointing towards a type of power that was not within any one country or nation state to really turn off. The march towards that was certainly a little frightening, but I did not necessarily think it was the fault of the companies or of non-regulation; it was just something that needed to be studied.

If we are going to do it right and preserve, as Alan said, all the wonderful things that technology brings us and how it empowers people in incredible ways, especially in non-western more authoritarian states, while also making sure that we can answer for the harms it is creating, those will be really difficult trade-offs and a fascinating set of questions. To that end, I asked for and received access, and was independently funded to study and follow, inside Facebook, the build-out of the Facebook Oversight Board.

My findings were written up in the *Yale Law Journal* and, most recently, the *New Yorker*, detailing exactly how Facebook founded and created the oversight board that Alan sits on and the formative couple of months or year that it first started operating. I did that, because there is no one lever that will solve the problem of online speech; it will be a series of levers, one of which will be building various types of accountability mechanisms into the tech platforms themselves, of which the oversight board may be one of the first and most ambitious projects.

The Chair: Let us start looking at the work of the oversight board and the challenges that you face, and then we will move on to some wider issues.

Q115 **Baroness Rebuck:** Thank you both for being with us today. I realise that we are just at the beginning and in the early days of the oversight board, so my first questions will mostly be for information.

Alan, you have answered this in terms of your own membership of the board, but what were the general criteria that were selected for putting this board together in the first place? Now that it is going to double in size, will you, as a board member, be involved in recruiting further board members, or will the co-chairs?

Alan Rusbridger: They wanted a diversity of candidates who represented legal backgrounds, human rights, free speech, editorial and government. That is who we come from, really. They wanted us to come from all over the world. Between us, we speak 29 languages and have lived in 27 countries, so it is certainly the most diverse board that I have ever been on. It was very important that it was not perceived to be dominated by Americans.

For the next 20, we were all invited to nominate people. We need more people who understand technology, to be honest. There are some bits of the world that are uncovered. There is a nominations committee that is crunching through all our recommendations. They were quite specifically

targeted to the kinds of skills gaps and regional gaps that we have at the moment.

Baroness Rebuck: Would you say that the cases that you select from are, to the extent that you know, representative of Facebook users' concerns, and possibly even broader societal concerns, about Facebook? I believe you select from a wider group of cases supplied to you from Facebook.

Alan Rusbridger: I think so. We also have a case selection committee, which I have not yet sat on; we are all going to rotate on these committees. They can see all of them. There is a secretariat that helps to winnow them down. We are trying to choose cases that seem representative of the most significant problems that Facebook is experiencing, or the most important issues. We are making quite narrow decisions on the specific context and facts of individual cases—a bit like a supreme court, although we shy away from the exact analogy—but you would expect this to have a wider ripple effect from each case.

Kate Klönick: In the first stage of picking the initial board members, there were four co-chairs whom the governance team, which was internal to Facebook, selected from a huge pool of nominations and recommendations from stakeholders. The four co-chairs then worked with Facebook to select the other board members, such as Alan.

Everything Alan said about the diversity of the board and what they were hearing is absolutely correct. There was a lot of discussion among stakeholders at all the meetings I was at about the unavoidable chicken-and-egg conflict at Facebook of building out a board that was initially staffed by Facebook and the taint that that would create. The idea of getting started with only 20 members instead of the 40 that had initially been planned was part of the solution to that problem.

Baroness Rebuck: Alan, apart from what I have read about community standards, what other guidance did you get as a board member, apart from having been selected because of your years of experience, to help in your decisions? Are the principles easily communicated and articulated among the board members?

I am also interested in how the five-member judging panel is selected in each case. Is it the board itself that selects them? Is it the chairs? Is there much conflict in these early days in arriving at decisions? I suspect that the board's deliberations on Donald Trump will be quite complex and high-profile, but, in a case like that, might that be a whole-board decision, or will it still be a five-member decision?

Alan Rusbridger: I am told that the panels are chosen completely by lot, although there is a preference to have one person from the region, so that, if you want regional expertise, you have that. The Donald Trump case will begin with an all-board meeting to take our views, and then it will be considered by a five-member panel before coming back to the full board.

In the cases that I have been involved in so far—we have been practising for a long time, even if we have done only a limited number of cases—we

look at the community standards and Facebook values, both of which are published documents. Then we apply a human rights filter, which is a more complicated thing than I had anticipated. As an editor, you are used, classically, to balancing Article 8 versus Article 10—privacy versus freedom of speech—but I was unfamiliar with a huge raft of human rights protocols and norms from around the world.

Your question about whether this is easily communicated is a really good one that we are wrestling with a bit. It is important that the decisions we come to are understandable by human moderators and, ideally, also by machines. There is a tension there, because sometimes you have looked at the facts of the case and decided in a particular way with reference to those three standards, but in the knowledge that it will be quite a tall order for a machine to understand the nuance between that and another case. These are early days.

Q116 **Baroness Rebuck:** We will get to questions about machine moderation a bit later. Building on that question, Kate, thank you for your *New Yorker* article, which I found very informative. At the moment, the board is focused on take-downs, but you suggest that, in the middle of 2021, it will look at keep-ups. Then you suggest various other things that it could look at, such as a virality lens, which is quite interesting. I do not know how easy that would be to put into play.

To what extent should the oversight board, if it can be characterised as the ethical supreme court of Facebook, have a voice in what is currently outside its remit, according to Helle Thorning-Schmidt, namely the take-down of a newsfeed that happened temporarily to a nation state? I am interested in starting with you on the slightly broader question of which direction this board should be going in, and I will then ask Alan the same question.

Kate Klönick: I do not have as much of a stake in what they should be doing from a strategic position on the oversight board, so I will answer just from an outside perspective of what is best for freedom of expression online.

Those are really complex decisions, which expand outside of saying, “Facebook, we’re going to recommend that you stop running your newsfeed”. Without anything more—any direct harm that has necessarily been proven or shown—it would be a confusing signal for the board to send if it were to step outside its role in that way. There is more opportunity for it to build slowly over time and to start listening to individual users on some of these instances of censorship and take-down of their speech. That has not been heard for a long time.

We hear so much about Donald Trump and high-profile actors who are censored by Facebook, but there are millions of people who are censored and their speech taken down at really crucial moments right now, and the board is their last source of any way to get their account back. We are in the middle of a pandemic, and one of the things that strikes me as really meaningful is how much we have all come to rely on these platforms to meet, to see each other, and to conduct business and

government. When you decide to take down someone's entire account or to ban their speech, you deprive them, very truly, of their right to association and their right to live their lives.

Those are not small questions, so they need to be the focus. The board has plenty to do in that category before it gets into the business of taking down or attacking certain products that Facebook puts out where there has not yet been any type of evidence presented about where the harms necessarily are.

Alan Rusbridger: In general, I was very struck by all the people working on new technology who are fascinated by Gutenberg. My colleague Jeff Jarvis wrote a book about Gutenberg. Say that was 1450; it feels to me that we are at about 1457 at the moment; centuries of turmoil and forces will be unleashed on society, and we are really in the foothills. None of this will be solved quickly.

I agree with Kate that the board will want to expand in its scope. We are already a bit frustrated by just saying, "Take it down" or "Leave it up". What happens if you want to make something less viral? What happens if you want to put in an interstitial? What happens if, without commenting on any current high-profile cases, you did not want to ban somebody for life but you wanted to have a sin bin, so that, if they misbehaved, you could chuck them back in again, like a yellow card?

All these things are things which the board may ask Facebook for in time, but we have to get our feet under the table first and prove that we can do what we want. At some point, I feel sure that we will ask to see the algorithm, whatever that means. Whether we can understand it when we see it is a different matter.

Baroness Rebuck: Will you understand it? Indeed, yes.

Alan Rusbridger: These are all things that are down the road.

Q117 **Lord Griffiths of Burry Port:** Thank you both for being here. I must commend Kate for the extraordinary *New Yorker* piece that she has written, which was thrill-a-minute stuff and helped me to understand things in a very pithy way.

You have both mentioned the slow speed at which we start, even harking back to centuries ago, with Gutenberg and so on. The fact is that things are now expanding so exponentially quickly that we do not have centuries, and the rules are being made by the platforms, which Governments are trying to catch up with subsequently.

It is really cultural difference that I am interested in here. How on earth do you deal with it? We have such little evidence to go on at the moment, so forgive me if I am pointing to things that may be atypical. Of the cases that are highlighted for us to look at, one deals with the treatment of Uighur Muslims in China compared with the killings in response to cartoon depictions of the Prophet in France, and the other deals with Armenians and Azerbaijanis. Those are two cases where there are such fierce oppositional forces in play and where any judgment that comes out would be perceived by some to be at the expense of the

other.

How do you hold all those things in tension? How do you come to a decision? I notice that one has been overturned and one upheld. I do not want you to go into the particulars of these cases, but they are illustrative of the complexities that you are dealing with as you move forward.

Kate Klonick: I could not agree more. One reason why I was drawn to this project five years ago and writing about this was seeing these incredibly vital conversations about trade-offs relating to historically oppressed groups or to censorship in an authoritarian state being made by individuals who were working as contract workers. The policy that underscored how they were enforcing this was being created by a bunch of people who did not necessarily have a background in free expression. They were businesspeople sitting in offices. Some of them were just wonderful, purely by accident, and were quite good at their jobs, but it just felt like the stakes were so high that we needed to create something other than that small group of people being the ones who decide these trade-offs of speech.

I loved the two cases that you highlighted, particularly for their difficulty. Knowing a bit more about some of those cases and how hard they were for the board to decide, I have to say that there is no way you can make decisions like that at scale. I do not know how you decide to make those kinds of trade-offs. Those would have been cases that took three years for the record to develop and be heard in a US court if they had been challenges to free speech.

What is so remarkable is that the whole world is coming alive to how difficult these questions are, and to the fact that these decisions are being made without them knowing about it and that the people doing it have been saying, "We don't need to point fingers. We just need to create an expert panel, which is what we've been doing". Having inquiries like this about how those decisions are being made is exactly the type of process and transparency that we need in order to get rid of the ghost in the system. Your points are excellent, and I completely agree.

Alan Rusbridger: The question about scale is very interesting. That is why it is not going to be done by next Tuesday. One of the most interesting cases that we have had so far was the French doctor who was very keen on hydroxychloroquine. Facebook banned that, because it said, "That's a crank solution and we don't believe in misinformation". If you looked at it, it was addressed to the French regulation agency and was trying to influence them to be more receptive to this.

I am very worried about banning that kind of speech, because where does it go then? It will go underground or into encrypted channels. I am afraid that I am an old John Stuart Millite; the way to tackle bad speech is to produce good speech, and you are not really solving anything by pushing it underground. I have some sympathies with Facebook, because a blanket ban on Covid misinformation would be an easier thing for moderators and machines to understand.

Lord Griffiths of Burry Port: I am also interested in the users and whether you can establish positions on some of these complex issues that you feel are faithful to the way, in general, a global audience might have a view, if it has a view. In her article, Kate points to these workshops that were held and that kidded themselves for a moment that they could crowdsource and get all kinds of people in to help to shape a consensus or something, but it fell flat on its face. It is *festina lente*: you are moving forward slowly in order to make real progress. How do users fit into the picture that you are shaping?

Alan Rusbridger: We are getting a pretty good response from people on these cases. They all know what is coming up, and people are feeding in, an awful lot about Donald Trump but not just about him. Again, the secretariat is boiling those down. Some quite prominent people are going on Twitter and saying, "Here is my submission and how I think the oversight board should solve this case". That is a useful thing.

It brings in a question that might be coming down the slipway about how you reconcile the cultural sensitivities of different parts of the world. Take something like nudity: different countries see nudity differently, so how do we settle a question like that? We had a nudity test case—we seem to have spent quite a lot of time thinking about nudity—that was interesting because of the inability all of us had to determine what we are dealing with.

It is not an old-fashioned media company. Nobody thinks any longer that they are just platforms. You can call them publishers if you want, but they are not publishers in the sense that the *Guardian* or the *Financial Times* are. It is plainly different, so we have started calling them the public square. That works quite well, but one of our board members, half way into one of these discussions on nudity, said, "Where I come from, if you start taking your clothes off in the public square, that kind of thing is frowned on", so we all said, "Okay, it is not a public square". We are so early into this discussion, and regulators and politicians are struggling with this too. We do not even know what to call them.

Kate Klonick: That is true. None of the analogies fits perfectly. To the question about users, what really struck me in talking to members of the board about some of the decisions that were made in the first round and the resounding statement back to Facebook to restore the majority of the speech that had been removed is that, in a much smaller sense than a supreme court or something like that, this is a dispute resolution mechanism for users. I heard from board members, but more generally, that some of the statements that have been made and released are sympathetic plights: you can be sympathetic to why they want their speech back up.

The other thing that you will hear, which has always been the case, is absolute frustration at not knowing specifically what rule was broken, how to avoid breaking the rule again, what they did to get there, or to be able to tell their side of the story. What you are seeing in the board's decision is, first and foremost, an attempt to build some of that back in. That is the signal that they are sending back to Facebook, which is pretty

low-hanging fruit, to be honest: “Let people know the exact rule. Give them a fact-to-fact type of analysis or an application of the rule to the facts. Give them that kind of read-in on what they are seeing”. People will be happier with what is going on, or at least it will just seem like there is more of a process and it is not just this black box that is censoring them.

Q118 Lord Lipsey: Thank you for this very helpful session. This question is aimed particularly at Alan. I want to just drill down a bit into the independence of this board. How long are you appointed for? Can you be reappointed, and can you be fired?

Alan Rusbridger: You are appointed for three years, and the appointment is renewable twice, so you can do nine years in all—three terms of three. Can we be fired? We are reappointed by the trust that has been set up over the board. I do not know what it would take to be fired, but there is the power to fire us.

Lord Lipsey: This could be thought of as a set of terms that are rather compromising of your independence, because you are paid a very decent amount of money to be on this board. I have no objection to you being paid a decent amount of money, but there will always be the question whether, as you get closer to the time of reappointment, you might start to look at the judgments you are making from the point of view of what is in Facebook’s interests or in the interests of those who are appointing you, rather than taking a purely objective view.

Alan Rusbridger: I honestly do not think it works like that, because we are not there to please Facebook. If you had wanted to please Facebook, you would have chosen a different group of people. My experience of my colleagues so far is that they are quite bolshie. They do not want to have anything to do with Facebook. They have turfed Facebook out of our meetings when we have realised that there are some people sitting there. We do not feel that we work for Facebook at all, so there is no obligation to be either nice or horrible to them.

By the way, one of the things that is nice about this board is that occasionally people will say, “But if we did that, that would scupper Facebook’s economic model in such and such a country”, to which we answer, “That’s not our problem”, which is a very liberating thing.

Lord Lipsey: We were distributed some of the judgments you had made, and there were far more of them where you said that Facebook should not have taken something down than those that said Facebook was right to take something down. You described yourself as a John Stuart Mill liberal earlier on this evidence session. There are two sides to this issue: people who think, maybe as you do, that we should have the absolute maximum of freedom and that freedom of expression trumps all other virtues; and people—I would put myself in this camp—who are much more nuanced about this and would not worry too much about stamping on an opinion that just might be legitimate but is mostly damaging. Does your view reflect that of your colleagues as a whole in being on the more liberal side of the spectrum? Why do you find yourself

on the liberal side of the spectrum?

Alan Rusbridger: I would not use the word “trump” in the sense of a trump card that beats everything else. It is a balancing act. Nevertheless, I believe that freedom of speech is a hugely important right; it is not more important than the right to life, but it is hugely important. In most judgments, I begin by thinking, “Why would we restrict freedom of speech in this particular case?”

That gets you into interesting questions. The right not to be offended has been engaged by one of the cases, as opposed to the borderline between being offended and being harmed. That issue has been argued about by political philosophers for a long time and it will never be settled absolutely. If you went along with establishing a right not to be offended, in the end that would have huge implications for the ability to discuss almost anything, yet there have been one or two cases where Facebook, in taking something down, has essentially invoked something like that.

Lord Lipsey: I was reflecting last night on something very interesting from an American fact checker—I think it is called “News Quest”. It had taken apart a very sophisticated anti-vaxxer called Kennedy, who puts all sorts of arguments that purport to be scientific but that fall down at the first bit of the searchlight. In a case like that, surely it is right for a well-informed platform like Facebook to say, “We’ll take you down, although you’re saying something plausible, and indeed something that you may well believe, because by doing so you’re going to kill people”. People will die as a result of this man.

Alan Rusbridger: Harm, as opposed to offence, is clearly something you would treat differently. We are in the fortunate position of being able to hire in experts and say, “Please advise us on the harm here”. To take the Covid case, I thought there was a difference between a medically qualified doctor arguing for a drug before a regulatory agency, and somebody saying, “Inject bleach”. Injecting bleach would be tremendously harmful and dangerous, and is very actionable, whereas the argument that a drug should be considered by an agency is not. We have not had a vaccine case yet, but instinctively I rather agree with you that that trespasses into harm. Kate may feel differently.

Kate Klonick: I suspect that you are using the terminology “harm” and “offence” in a way that I am slightly unfamiliar with as an American, but maybe not. When you have certain types of line-drawing, it often, in these hypotheticals, seems very straightforward. I would just point out in particular the coronavirus/hydroxychloroquine case that Alan referenced earlier, which on the one hand could be disassociated from the page that it was on and maybe seen not as a policy argument but in fact as some type of call for something, but it was much closer to a policy argument and a pushback on a certain type of speech. The harm seems slightly less than a viral messaging system that is going around telling people to inject bleach, take drugs that might hurt them or not solve their problem, or not be vaccinated.

I emphasise again that these are all really hard questions. Should Facebook make a decision on them, the only thing that we can ask for, in

terms of freedom of expression in a harm situation, is to have some type of transparency and appellate process in place to re-review these. Facebook's bottom line is served by over-censorship in that particular scenario. They would be best served by not creating a huge PR rumble by keeping up a lot of potentially very dangerous speech. They would be better served by taking it down and then letting the board make those types of determinations for them.

Just to be clear, this loops back to something that you said earlier about independence, which is that the independence of the board is, in some ways, very much in Facebook's best interests, because it makes a lot of these hard decisions that Facebook does not want to be on the hook for making. That is a very important thing to be aware of, in that, as the board grows and takes more cases, the nature of the symbiotic relationship between these two institutions is not clear.

Q119 Baroness Bull: Just to push a little bit on that relationship between Facebook, the board and the trust, I see from oversightboard.com that the oversight board comprises what it calls three interlocking elements—it recognises that the board, the trust and the administration interlock and that each has roles in ensuring accountability. One always has to ask, "Who governs the governors?" to see where power really sits, and we all know that \$130 million has been given by Facebook to establish this independent trust. Kate will correct me, I am sure, if I am wrong there.

Alan, if you were still editor-in-chief of the *Guardian*, what questions would your paper be asking to get to the bottom of the relationship between the three? What would you want to be picking away at?

Alan Rusbridger: I guess I would hope to do what Kate did, which was to spend a lot of time having complete freedom to speak to everybody and to tease out any anxieties that people have had. Kate did a very good job in that *New Yorker* piece. The question has arisen already as to whether we feel in any way obliged or in debt to them, and I honestly do not. As I said, I have not met Mark Zuckerberg since this happened. We have not met any Facebook executives since the early days. I feel highly independent.

Baroness Bull: You talked about them attempting to be in the room but being kicked out. Can you explain what you mean by that?

Alan Rusbridger: Sadly, we have never met as a board because of Covid, so the whole thing took much longer to set up than it would have done. We had an immense number of training sessions in human rights, systems, technology and how moderation happened. Sometimes we were being trained by Facebook. Kate writes about this in her piece. At some point, we got to the point where the training had ended and, on a Zoom screen much like this, you saw there were still one or two Facebook people with their names there. One or two of us said, "Hold on a minute. What are they still doing here?" Once the training and

induction had ended, we kicked them out and we have not seen them since.

Kate Klonick: I want to quickly say how it was told to me, which is exactly as Alan said, but I also want to emphasise that I thought that was an interesting detail to put in. It was a slightly esoteric point. When you are building a watchdog institution, it is incredibly hard to break those bonds and set up new people with, frankly, a huge set of problems with new technology and a new back-end in a content management system. Facebook had to be there to some extent, and this was exactly the type of moment that, having watched this, I knew had to happen; there had to be some type of formal break. It was told to me in a way that seemed like this was just a natural moment. They had done their training and this was going to be a moment of pushback and breaking away from the nest.

This reminds me of one small point that I want to make about the 20 board members who were initially chosen and some of their qualities. I will add one detail. One of the things frequently mentioned to me about how they chose their first 20 members was that not only did the members have to be diverse and from various places, but there was also a very real concern about having people who had history with institution-building and who would work well together, because that was seen as essential in the initial phases of this.

It is not a huge, salacious detail or anything, but it is an interesting piece. Alan talks about building out other types of areas in the coming year and having 20 new members. Some sacrifices were made, such as having people with a history of institution-building and leadership on the one hand but not with super-deep content moderation or technological sophistication on the other. That was where the balance was struck, and that will be corrected or readjusted going forward.

Q120 **Viscount Colville of Culross:** Good afternoon. Thanks very much for such interesting evidence. Inevitably, the global scale of the platforms means that they are becoming ever more reliant on machine learning to moderate content. However, the companies' use of AI for moderating platforms has increasingly come under scrutiny, because machine-learning tools have difficulty in accounting for context, sarcasm and cultural meaning.

Alan, you already talked about the difficulty even for humans to look at the different cultural meanings in nudity. Are you concerned that, as a result of the increased use of machine learning for moderation, it poses a threat to free speech?

Alan Rusbridger: There is a tremendous penalty in bad moderation, whether by human beings or machines. That is just very bad for free speech. Again, without repeating the Gutenberg analogy, we are only just at the foothills of learning, so I have no reason to think that the machines will not get better and that the information going into the machines will get more sophisticated.

As I said to begin with, we need more technological people on board, or people who can give us independent advice from Facebook, because it will be very difficult to understand exactly how this artificial intelligence works. I have heard, as I am sure you have, this sense of the machines taking over at some point. They break free and they start making their own independent decisions. I have no idea if that is true or not, but I feel as though we need to educate ourselves in that. I am sure Kate understands this better than I do.

Viscount Colville of Culross: That is one of the great concerns of all of us: how to understand algorithms and what they are doing for users and for the effectiveness of the platforms. Is it ever possible to get proper transparency in the use of algorithms, particularly when it comes to the moderation of content?

Alan Rusbridger: It is essential. Again, I hear the same things that you hear. People say to me, "You're on that board, but it is well known that the algorithms reward emotional content that polarises communities, because that makes it more addictive". I do not know if that is true, and as a board we are going to have to get to grips with that. Even if that takes many sessions with coders speaking very slowly so that we can understand what they are saying, our responsibility will be to understand the metrics of the machines that are going in, rather than the machines that are moderating.

Viscount Colville of Culross: Kate, are you concerned that having machine-learning moderation of content poses a threat to free speech?

Kate Klonick: Yes, absolutely, for the same reasons why notice and take-down regimes, copyright and content ID are incredibly damning. One of the things that has not been mentioned enough is that one of the reasons why we are talking about content moderation right now is that it has started to happen to powerful people. It started to happen with the "Terror of War" photo and with the events in Norway in 2017. That involved a powerful person and another powerful person, political and non-political, exercising their power to get their content reinstated. The posting of the "Terror of War" photo had been a mechanism for censoring individuals for a long time and, in some cases, getting their accounts suspended.

One of the things that is super-important to underscore here is that, when we use these high-speed algorithms at scale and have false positives that do not hurt the company that much from an economic perspective, you end up with mass censorship and no recourse. That is fine if you can go to other platforms, but, as we know, there are network effects and all types of other reasons why these platforms are not created equally.

What I am most afraid of right now is that you put these non-nuanced "Keep it up or take it down" type of solutions in place when there is no need to continue with that binary. We have technology, as Alan mentioned, that can be implemented on interstitials, or on de-ranking things or other types of mechanisms that are more thoughtful, like

labelling, and not have the censorial power rested in the hands of these tech giants and their non-transparent algorithms.

Viscount Colville of Culross: Could the algorithms ever make subtle enough decisions to look at de-ranking and making it more difficult to get access to a post? I am sure that they could; it is about what you feed into them.

Kate Klonick: I will give you an example. Algorithms are not one type of thing. Also, they are written and made by humans. They are informed by data that is generated by humans. It is such an easy "point at the robot" type of thing, and so easy to say "that", but it is not a "that"; it is an "us". I just want to make that very clear.

That said, let me talk for one second about how algorithms factor into some of this stuff. It is not just newsfeeds and other types of things. One of the pieces that I wrote for the *New Yorker* before this was after the Christchurch shooting. The livestreamed video of the shooter going into the mosque and committing this horrible atrocity was posted on Facebook. Facebook immediately tried to take it down and created what had previously been a very rigorous way to take down videos, which was through hash technology.

Then all these people on 8chan, 4chan and other sites circulated alternative versions of the video that clipped out the hash or flipped the video backwards, so that it would play in reverse image and escape the take down, or changed the soundtrack so that it would escape other types of detection. That went on for 48 hours, with a team of people around the world chasing down this horrible video. Do you think that Pepsi or General Electric wants their ad up next to Christchurch? Facebook has no incentive to keep that up, but it stayed up because bad actors decided to keep it up and keep it posted.

I am not trying to make Facebook a hero in this story. There is always more that can be done. An algorithm is like a rule: as soon as there is a rule and you know what the rule is, there will always be ways to break, manipulate or twist it to your means. These companies are dealing with those types of hard choices all the time. Some of them are not striking the right balance when it comes to where they are putting their money and effort, but algorithms as a concept are an oversimplified way of thinking about this.

Viscount Colville of Culross: It is very good to have that clear in our heads. Thanks so much. That is great.

- Q121 **The Lord Bishop of Worcester:** Thank you for really helpful evidence this afternoon, both of you. I want to look a bit more closely at a particular aspect of moderation. As you are probably well aware, the UK Government are due to publish their Online Safety Bill later this year. It has been a long wait. They have already published their initial and full responses to the White Paper. Can you tell us your view of the proposals as we understand them so far?

Alan Rusbridger: From what I have seen, the idea of individual nations protecting especially children but online safety in general is a completely reasonable and necessary thing to do, so what I have seen of it so far looks perfectly sensible. You will need a form of independent regulation as well, but the broad framework of this approach to harm seems to me a sensible one.

Kate Klonick: I agree. It is always important not to underestimate the bad actors in these scenarios and to make sure that we are targeting them and helping the companies to target them. Generally speaking, the categories created for the different tech companies seemed slightly soft. I do not know how that will be applied or whether it will be just thrown out and we will see where things land.

I do not know where Microsoft, for example, would land in this scenario. It does not make consumer-facing products like the ones you are talking about, but it does operate the cloud, which is a source of CSAM and other types of harmful material with children. There is no clear delineation, in my mind, between the types of companies that are targeted in different places in the stack, the stack being the internet stack of your Fios cable and your basic pipe all the way up to Amazon Web Services and domain-name providers and to platforms. Microsoft is somewhere in the middle. Facebook is also somewhere in the middle and also at the top. That part of this Bill strikes me as still underdeveloped.

The Lord Bishop of Worcester: There is a lot more that we could talk about there, but we are pressed for time so I will just highlight one particular issue. What sort of sanctions should be applied for non-compliance? The White Paper does not go into details, although it states that Ofcom may have the power to block non-compliant services, fine them up to 10% of global revenue, or bring criminal or civil proceedings. It has been suggested that fines for platforms should be linked to how long illegal content has been on their website, and that directors should be criminally liable. Do you have views as to what the maximum penalties might or should be?

Kate Klonick: I had thought that the criminal parts of this had been taken out. I thought that it was only fines. Is that not the case? I thought that there would no longer be criminal liability, but maybe I misread it.

The Lord Bishop of Worcester: That is not my understanding, but tell me what you think.

Kate Klonick: Piercing the corporate veil here and going after directors may be a little disassociated from the problem. Maybe it will provide a better incentive, but it seems a very blunt tool for dealing with this. Generally speaking, the fines are very incentivising and appropriate.

Alan Rusbridger: I am not sure that I know enough about this to argue very sensibly, but the one thing that makes me slightly anxious about enormous fines is that Facebook can pay almost limitless fines. Google can too. If we are trying to encourage innovation and new players, and to not have these gigantic, monolithic companies, the prospect of huge

fines, say, for failures of moderation makes me nervous, because that will deter people from coming in and challenging them.

The Chair: My reading of where the Government are on criminal liability is that they have hinted that it is not what they are minded to do, but I do not think they have confirmed it.

Q122 **Lord Stevenson of Balmacara:** Kate, in your article in the *New Yorker*, which we have all read and think is very helpful and useful, you say, "Facebook had modelled itself as a haven of free expression on the internet" but has been overtaken by "conspiracy theories, hate speech and disinformation".

We have heard a lot of evidence in our earlier sessions that the business model under which Facebook and other social media companies are established needs the hate speech, the conspiracy theorists and the disinformation spreaders to be on their site for as long as possible in order for their numbers to stack up and the overall income that they are going to generate, because they are not there just as people and eyeballs looking at advertising; they are there because they can harvest the data. If the business model is saying, "We don't want you to start taking stuff down. We want more of this awful stuff", does that not strike at the heart of what Alan is trying to do in relation to the board?

Kate Klonick: Alan mentioned this. Interestingly, the oversight board is saying that a lot of the stuff that is problematic should be put back up. There will be this interesting trade-off. To counter your point, when people see unpleasant things for a long time on Facebook, they stop going on Facebook. A site with photos of puppies and cats would be a good reason for people to go on Facebook all the time, especially these days. It goes both ways. It really depends on the person's internal view of what they are seeing in the news, but generally speaking I agree with what Alan said.

Alan Rusbridger: I cannot see why Facebook thinks it is in its long-term interests to have that kind of content. There is a sense of exhaustion there. They are engineers. Yes, of course they like making lots of money, but the grief that they are getting at the moment, from all sides, because they do not have this under control must be so debilitating. They want it to go away and they want it to be a better, safer place. Whether the machines need to be significantly tweaked in order to achieve that, I do not know, but I would not be too cynical about what they want there.

Q123 **Baroness Grender:** Thank you both for a fascinating session. Alan, you mentioned the need for the oversight board to have more understanding of tech, and the need, in turn, for the machine to understand more. A hotly contested issue in our last inquiry, which was into the future of journalism, was the curation of content and the application of algorithms. I would love to hear your views on that. There is some suggestion that, for instance, Google gives higher prominence to one media outlet versus another. Is that something that an oversight board like the one at Facebook would ever have an engagement in?

Alan Rusbridger: I hope we do. This question of metrics and incentives is so important in both the old world and the new world. Since I stopped editing, which is only five and a half years ago, the ability of every editor to know exactly who is reading what, for how long and what they do afterwards is just extraordinary. You see lots of newspaper companies now saying, "Our aim in life is to drive subscriptions, so we're going to prioritise the kind of content that drives subscriptions".

That sounds reasonable, but what happens if people do not like reading about climate change? There is some evidence that people are frightened of reading about climate change. Climate-change articles do not drive subscriptions, so are we going to downrank climate-change pieces because that is not what our metric says? We have to be in charge of the metrics, rather than let the metrics be in charge of us.

Kate Klonick: I completely agree. The one-to-one between good content and money is a lot less understood than people suggest. You hear both sides: that the problematic and salacious content is what drives people to click and to visit, and the exact opposite—that people do now want to see the really problematic content. They want Facebook to take it down. There is something to be said for the fact that people want speech that they want, and do not want speech that they do not want. They want Facebook and some of these tech companies to read their minds. That will be an impossible standard, but it is also not a new standard. I used to get *Esquire* magazine, which would regularly have a bunch of articles in it that I did not like reading, but I did not call up *Esquire* magazine to cancel my subscription. That is something to consider.

The Chair: Alan Rusbridger and Kate Klonick, thank you both very much indeed. It has been a very interesting session. Kate, you have driven a pile of new readers to the *New Yorker* magazine, and we found your evidence absolutely fascinating and useful. The inquiry will go on for some time, so if either of you has further thoughts or evidence that you would like to put in front of us or send us for our background reading, we would welcome and appreciate it. Thank you again for joining us today. Thank you very much indeed for the evidence and the time that you have given in preparing for the meeting.