

# Select Committee on Democracy and Digital Technologies

## Corrected oral evidence: Democracy and Digital Technologies

Tuesday 17 March 2020

2.35 pm

Watch the meeting

Members present: Lord Puttnam (The Chair); Lord German; Lord Harris of Haringey; Lord Holmes of Richmond; Baroness Kidron; Lord Lipsey; Lord Lucas; Baroness McGregor-Smith; Baroness Morris of Yardley.

Evidence Session No. 25

Heard in Public

Questions 311 - 325

### Witnesses

[I](#): Katy Minshall, Head of UK Government, Public Policy and Philanthropy, Twitter.

## Examination of Witness

Katy Minshall.

Q311 **The Chair:** I am going to read the police warning; then we are off to the first question. As you know, this session is open to the public. A webcast of the session 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 an opportunity to make minor corrections for the purposes of clarification or accuracy. I wonder if you might introduce yourself for the record.

**Katy Minshall:** I am the head of UK public policy, government and philanthropy at Twitter.

Q312 **Baroness Morris of Yardley:** The first question is about authenticity. What would Twitter understand by a "fake account"?

**Katy Minshall:** Can I start by thanking the Committee for inviting Twitter to participate today? I wonder if you would allow me a minute or two to give a scene-setter on Twitter. It might be helpful for people who are not regular users.

**The Chair:** We did not have any formal evidence to look at, so that would be very helpful.

**Katy Minshall:** We are here today to talk about political discussion on the service, but it is important to acknowledge that people log on every day for all sorts of reasons. The most popular conversation on the service last year was actually about K-pop. We saw over 6 billion tweets about K-pop, a genre of pop music originating in South Korea.

When it comes to political discussions, we are proud that Twitter has given a public platform to many voices that, 10 or 15 years ago, lacked that opportunity, particularly the young and marginalised groups. In the 2019 general election here in the UK, we saw an increase in the conversation by 66 per cent compared with the 2017 general election. That is more people coming online and talking about politics and policies that are important to them.

The key distinguishing feature of Twitter for the purposes of this discussion is that we are the only major service to make our public conversations available for the purposes of research. Over the past decade, at least tens of thousands of researchers have accessed our public API and we are really keen to double down on that.

That is not to say that we are not acutely aware of our responsibilities here. We welcome opportunities such as this to work with experts and policymakers, and to think about what more we can do. I have two examples, if you will allow me, and then I will come to your question. The first is our announcement last year to ban political advertising globally, which I am sure we will get into. We took that decision based on the principle that political message reach should be earned and not bought.

The second is our investment in technology, which is starting to pay real dividends in the context of safety. One in two tweets that we take down for abuse we have detected ourselves, reducing the burden on victims to report these issues to us.

Let me come to your question. For fake accounts, I will talk about how we define them and then how we go about enforcing those rules. When you sign up to Twitter, we ask you for your full name and your email address or phone number. When you get on to the service, we ask you to abide by a series of rules, including that you cannot mislead others on Twitter through the use of fake accounts. The sorts of signs we take into account are the use of stolen or stock profile photos; your location — if you are saying that you are somewhere you are not; or the use of a copied bio. That rule is part of our wider policy on platform manipulation, including that you cannot mislead others on Twitter through bulk, aggressive, deceitful activity. That includes activities such as operating multiple accounts on the service to try to disrupt conversations or make something trend artificially.

The question then is how we enforce that. On the one hand, it is important that any user feels able to report their concerns on these issues to us. If they see an account on Twitter that they think is fake, they can, with a couple of clicks in-app, make us aware. However, to really get at this at scale, we have to use technology. We have been able to automate a number of processes when it comes to malicious behaviour detection. For example, an account tweeting at an extraordinarily high volume on a specific hashtag is something that a machine can detect. It might be a straightforward case of sending that account a password reset to see what is on the other end of that account. More complex cases are escalated to our site integrity team, which is charged with identifying and removing bad actors from the service.

To give you a sense of the scale, we release a transparency report twice a year where we share aggregate data on actions we have taken against our rules. Our most recent report came out towards the end of last year and we shared that we had challenged 97 million accounts in this context over the first six months of 2019.

**Baroness Morris of Yardley:** I shall resist responding to that, because a lot of it will be covered in follow-up questions elsewhere, but thank you for that introduction. Let us go back to the name of the account. What is the advantage of allowing people to use names that are clearly not their own or that cannot be recognised by someone else? Why do you do that? What is the advantage of doing that and what does it add to the process?

**Katy Minshall:** You can be pseudonymous but you cannot have a fake account. I could not say that I am Baroness Morris of Yardley.

**Baroness Morris of Yardley:** You cannot have a fake name.

**Katy Minshall:** I could not say that I am you. That would be against our rules. You can be pseudonymous. The reason is that we have seen all

sorts of really powerful use cases in countries around the world where pseudonymity has enabled people to speak out, even in the UK. Last summer we saw an account that got a lot of coverage, @thegayfootballer. Now, this purported to be a high-profile footballer who was gay and who was struggling to come out. The existence of that account sparked a wider conversation about homophobia in football.

Shortly after I joined Twitter, lots of people were engaging in a hashtag, #WhyIDidntReport. It had examples of individuals talking about why they had not reported sexual harassment and sexual assault. Being able to speak using a pseudonym enabled more voices to join that conversation and share their perspective on that issue.

**Baroness Morris of Yardley:** They are not all well motivated, are they? You have taken some examples, but I would still argue that it is better to do it in your name or find another way, because I am not sure what it adds to the process. A lot of abuse is done not in users' own names. If you weigh it up, without knowing the answer to the question, I suspect that far more abuse comes through pseudonyms than honourable causes such as those you have just mentioned. That cannot be the reason you allow it to be anonymous: that someone who is not quite ready to come out as gay will talk about homosexuality in football and that is the only way they can have that conversation. That cannot be the reason you have this policy. What is the advantage of it across the board?

**Katy Minshall:** There are a couple of questions there. If you put to one side the powerful use cases of people being able to use pseudonyms, there is the question of how you verify people's identities and whether it will work in solving the problem you have outlined.

Let us take the first question. You will be in the space of asking what ID it is appropriate to check. In the UK, you might think of a driver's licence and passport. Large swathes of the population have neither and you would not want to restrict those communities' abilities to use a service such as Twitter. Twitter operates a data minimisation approach. At this time, do people really want to be sharing more private information and data with companies such as Twitter?

Put that to one side. Will this work? There are not many use cases to draw upon here but there are a few. South Korea brought in a law in the 2000s trying to get at this exact issue where, on websites with over 100,000 users, you had to share your real identity. That law was overturned about a year later, because it was found not only to have had a negligible effect on malicious communications but to have created a database that was vulnerable to constant attempts to attack it and take that data. From my day-to-day experience at Twitter where cases of abuse come across my desk, plenty of individuals are willing to abuse not only using their full name but sharing all sorts of personal information about where they live and work.

**Baroness Morris of Yardley:** Why, when people register with you, do you ask people for their real name and address, if it is pointless?

**Katy Minshall:** There are a few things. The most important is cybersecurity. We strongly encourage two-factor authentication. That means, if I log on to Twitter from a new device, my password entry will be coupled with an email or a text from Twitter, with a code allowing me to access Twitter.

**Baroness Morris of Yardley:** You cannot have it both ways. You cannot say that it is really strong for that purpose but really weak in stopping people using a name that they claim is their own.

**Katy Minshall:** Sorry, do you mind repeating that?

**Baroness Morris of Yardley:** I knew I had said that badly.

**The Chair:** Try again.

**Baroness Morris of Yardley:** I will start again. In your follow-up, when I had asked why people do these abusive things in their own name, you said that it would not work, people could give false names and we could not rely on that. Then you gave various examples. You seem to be saying that, if you register in your own name with your own address and date of birth, even that is not effective. I could register as Toby Harris, pinch his ID from his pocket, and that would be that. You are saying that it is a frail system. Then I asked why you use it at all, and your second answer was that it is really robust and you ask for lots of ID. The point I was making was that you cannot have it both ways. It is either robust, in which case you can identify people online, or it is not robust. It is not a big question; it was just a clarification of what I had felt were contradictory statements.

**Katy Minshall:** To be clear, asking for email or phone in the context of two-factor authentication is a really strong intervention to provide cybersecurity on Twitter and services like it. On your wider question about identity online, our CEO has said that we should be sceptical of most things we see online. Even websites that purport to have a real-name policy will face the exact same challenges that I have outlined of how to identify people in the digital age.

**Baroness Morris of Yardley:** It is partly about the message you give out to people. If you allow them to use names that have no resemblance to their own, you are giving the message that it is okay to hide behind a pseudonym. You have made a case, which I do not agree with, for allowing it. I was not of the political era where we got as much abuse as there is now, but I would much sooner have abuse from somebody who put their name and address on the letter—it was letters in my day—than somebody who sent it with no address and with a false name on the bottom. It was just different. You might not agree with the abuse they were giving you, but you felt they were honest enough to actually state that and you could write back in whatever tone you wanted.

Are we really a better society if we make it easier for people to abuse others without saying who they are? Does that really make us a better society than saying to people, "To be honest, you could be brave and be

accountable for what you say”?

**Katy Minshall:** Any account that breaks our rules, whether it is a verified user or someone using a pseudonym, by abusing, harassing or engaging in hateful conduct, will be subjected to enforcement, regardless of who does it.

On your question about the experience of politicians, I have testified on this issue before and, as I said then, I do not think anyone could see the experience of politicians today and not be alarmed. For over a year now, we have had a really close partnership with the security team here and we have found that invaluable. They report abuse of Members on Twitter through a direct portal to a specific team on our side. That is coupled with a weekly phone meeting between me and the security team here. As well as unblocking issues with that process, they can give a heads-up on any upcoming debates that might be contentious, so I can tell the safety team that we could expect a high volume of reports at that time.

More recently, that has been coupled with as much training as we can offer. I was here with counterparts from Google and Facebook a couple of weeks ago running drop-in sessions for Members to talk through our approach, answer any questions and work through any specific issues they were having.

**Baroness Kidron:** On that last point, I think both Houses really appreciate that concentration on their particular problem. How do you deal with the broader problem of, say, women in public life, who get an incredible amount of abuse and are not necessarily Members of this House?

**Katy Minshall:** First, we should be working directly, as closely as possible, with expert organisations looking at these issues. The Committee on Standards in Public Life, for instance, produced a report on intimidation in public life. It made a series of recommendations, all six of which I believe we have now moved on in quite a substantial way, and we stay in touch on a regular basis to track progress. That is based on its expertise, talking about the issues of people in public life and what it wants services such as Twitter to do.

Secondly, we cannot solve problems of moderation using humans alone. We have to double down on our investment in technology. We see about 500 million tweets every day on the service and, operating at that scale, we have had to ask machines to do as much of the work as they can. That is starting to show real results. In the past year, we have increased the number of accounts we take action on by 105 per cent and that is using interventions such as machine learning.

Q313 **Lord Lucas:** Can you tell us about Twitter’s experiments with labelling misinformation? Has Twitter opted for a community-led approach rather than one based on expert fact-checkers and, if so, why?

**Katy Minshall:** I believe you are referring to a design that was covered on news websites a month or two ago that showed, as you say, a

community-led approach to misinformation. We are exploring many approaches to misinformation. This particular one is not currently staffed and is not comprehensive of the range of approaches we are looking at. We want to work through several more iterations before considering whether to test it further.

That is not to say that we do not have a community-led approach in other ways. The first is with researchers. As I said, tens of thousands of researchers have accessed data using our public API over the past decade. About a year and a half ago, we went a step further and curated a dataset of information operation we had seen on the service over a number of years by the Russian IRA. It was about a terabyte of data that we released in late 2018 representing thousands of accounts, 10 million tweets, two million photos, GIFs and images. Since then, we have shared 22 further datasets of state-backed information operations on the service. They are available on our website right now and have been accessed all over the world by researchers thousands of times.

Why did we do that? We wanted to increase our collective understanding of these issues. In terms of the community-led approach, there are experts around the world who can bring to bear on our datasets their knowledge and their background, and ensure that we all, as a society, understand these issues better.

**Lord Lucas:** Are there any other approaches that you are looking at more positively?

**Katy Minshall:** Another key thing is how we make rules, which again is community-led. Earlier this month, we launched a new rule on synthetic and manipulated media, essentially deep fakes and shallow fakes. That rule states that you cannot deceptively share synthetic and manipulated media on Twitter with the intention of causing harm.

Where did that rule come from? It came from a public consultation. We engaged with experts and academic researchers as well as doing a public consultation. Even on Twitter, you could contribute using #TwitterPolicyFeedback and we got 6,500 responses. On another public consultation for a different rule, we had 8,000 responses. Involving our community as much as possible will mean that our policies and rules are really robust.

**The Chair:** Katy, does the way that, for example, President Trump uses Twitter equate to your understanding or our understanding of accountability?

**Katy Minshall:** Could you elaborate on your question?

**The Chair:** Does the way in which President Trump chooses to use Twitter conform to what we understand by democratic accountability?

**Katy Minshall:** I do not want to speak for your own view but there are a couple of things here. Politicians all over the world use Twitter every day to engage with their constituents and engage in policy debates.

**The Chair:** I was using him only as an example. I am really questioning whether very short bursts of opinion are, in any way, an acceptable alternative to being properly questioned in an open forum on the views you have. We have a danger of politicians, in particular, retreating behind Twitter tirades instead of being accountable in the way that we, certainly in this building, understand accountability to operate.

**Katy Minshall:** We do not see ourselves as a replacement for that. We see ourselves as complementary to those processes. In the election, we worked closely with news partners so they could live-stream things such as the debates. We had nine different live streams on election night that had over three million views. We see very direct uses of Twitter, particularly for diplomacy, where leaders across the world share their statements and reply to each other. We even saw, in the US, two politicians from different sides of the aisle finding an issue they could agree on and doing so through Twitter, but we see that discourse working alongside all the other existing mediums through which politicians engage with their constituents.

**The Chair:** I am struggling with the idea of Twitter being discourse but I will leave that for the moment.

Q314 **Baroness McGregor-Smith:** Could you tell us what role you believe human moderation plays in improving online experiences? How could you make the processes more transparent and consistent? I would be interested to know whether you have, or have considered, a public database of decisions you make.

**Katy Minshall:** Transparency is the right question. My concerns with a public database, although it is an interesting idea, are twofold. We already provide aggregate-level data about the number of reports we receive on accounts against different rules and the number of accounts we take action on. The more information we provide, the more there is a risk of inadvertently sharing private information. The more insurmountable challenge with the public database is that it could enable adversaries to build solutions designed to game our enforcement efforts.

As I said, transparency is the right question and I can talk through how Twitter thinks about transparency to our users and the general public more broadly.

**Baroness McGregor-Smith:** Surely, though, the right way to be transparent and to explain your values and ethics is to publish the way you moderate decisions. I watch Twitter and it is interesting how you comment on Twitter. I see it used as a vehicle for hate pretty much every day. The commentary on Twitter is not of the level that you would necessarily get in written media, because they are governed by different rules. But we do not know what your rules are. Your rules are not transparent and, as a user, I would not even know where to look at how you decide to moderate. I do not actually think you do any moderation because I cannot see it. Every time I look at Twitter, it is pretty abusive in many areas.



**Katy Minshall:** Our rules are on our website. We have a shortened version designed to make them as easily accessible as possible and, for each one, we have a longer version with more detail about the criteria content moderators will consider, and example tweets.

The issue you pick out is the right one. We are certainly keen to do more and think more innovatively about how we can communicate our rules and our approach. To give you an example, last month for Safer Internet Day, we made a series of videos with influencers on Twitter who had experienced abuse. They talked about their experiences and the tools they used, such as muting words in conversations and blocking accounts, to manage that every day. That had close to two million impressions with a view-through rate of almost 60 per cent. Rather than someone scrolling past, that is someone meaningfully watching and engaging. Those partnerships may be a more compelling way to reach people on Twitter compared with just expecting our users to go on our website and read our rules.

**Baroness McGregor-Smith:** If you put together a list of the things you moderated, and explained how humans moderated for you and the decisions against which you moderated, surely that would give new users to Twitter a lot more comfort about what they are opening themselves up to in this online world. Every time someone I know goes on Twitter, they are potentially subject to abuse. Although you know the rules, they do not know how you make decisions about those rules. It is a bit like walking down a dark street; you make a decision about whether you want to walk down that dark street. With Twitter, you do not know what you open yourself up to, because it is not easy to understand. Although you may have things on your website and you may argue that there is transparency, when I look at the 140 characters, including the example of the President of the United States, people are not that kind. Everyone judges kindness and humanity in a different way but, if you put this together, people would probably really admire you guys for it. Admiration is not something that I think a lot of people have at the moment.

**Katy Minshall:** Can I just check I understand? It would be something on our website to say what our rules are and give examples.

**Baroness McGregor-Smith:** It is about how you enforce them. "These are the things we have taken off. These are the things we have moderated. These are examples and this is how we enforce, humanly, our rules as an organisation".

**Katy Minshall:** I totally agree. We have on our website now what the rule is and the criteria that the moderators take into account.

**Baroness McGregor-Smith:** Do you have examples?

**Katy Minshall:** We have examples of hypothetical tweets. They are not real-world tweets but hypothetical ones we have come up with. We will share, where relevant, the breakdown of where we have proactively used

technology to find that content and make those decisions versus where we have relied on reports from users.

**Baroness McGregor-Smith:** Why not do it with real tweets and say, "These ones are unacceptable so stop doing it"? Human beings listen to truth; they do not listen to vaguely moderated hypothetical examples. They think, "Right, that is the thing we cannot say". Would you at least take that away and think about how you can do it?

**Katy Minshall:** I am happy to take that away. I would want to check our privacy policy, but I am very happy to come back to you on that.

Q315 **Lord Harris of Haringey:** Can we go back to the question of consistency? To be quite clear, I have not checked back on the various instances where I have seen this, but I have seen people complaining about inconsistency between how their tweets and other people's tweets have been treated. What is the process of making sure you apply those criteria and your rules consistently? Would it not be easier for people to believe that you were applying them consistently if you published more information about it?

**Katy Minshall:** The fact that we are a public, open platform, and researchers publish information and findings on Twitter data all the time, helps get at some of these issues. The challenges on Twitter are well documented. We want to double down on research to support that and work very closely in collaboration with external experts as we find solutions.

To come back to your question on consistency, when content moderators start at Twitter, they will undertake certification exams with strict grading criteria. That will really be tested for quality and accuracy of review. Should they pass it, as their career at Twitter continues, they receive training on an ongoing basis and are spot-checked to test for quality and accuracy.

There is also a question on the product side. How do we consider safety by design in the product life cycle? On the one hand, any big product change is inherently cross-functional at Twitter. Product, engineering and design work closely with our trust and safety teams, and potentially with my team and the Twitter service team, which enforces our rules. The most important team of all is our product trust team. Product trust is a team within Twitter charged with preserving user trust and doing what it can to keep the service free of abuse and spam.

They will do that in all sorts of ways. To give you a sense, if there is a big product change, or for big products on Twitter on an ongoing basis, they have a dedicated product trust point of contact who gives timely, actionable feedback on any proposed changes. They suggest mitigations. They create new policies and systems if we are rolling out a new product. Before we even get to content moderation, we are really proud of what you might describe as a safety by design approach, where we are thinking about this from the very beginning.

**Lord Harris of Haringey:** I am not sure that quite answers my point about the consistency. You seem to be saying that consistency stems from your rigorous training programme and people cross-checking. How long does somebody spend on the training programme?

**Katy Minshall:** I do not have that information to hand but I am happy to take that away.

Q316 **Baroness Kidron:** In this context specifically, how is Twitter finding the current misinformation about the virus? How are you doing on that and what are you learning from it?

**Katy Minshall:** Right now, we are not seeing significant co-ordinated attempts to spread disinformation about this topic at scale on Twitter. I spoke to the Government and DCMS last week. They also said that they are not seeing disinformation on this topic targeting the UK at this time. We will absolutely remain vigilant on that.

**Baroness Kidron:** Can I ask more broadly than the UK on this particular issue? It is a global virus and a global issue.

**Katy Minshall:** The big drive for us, at this time, is what we can do to make sure that anyone who comes to Twitter is met with credible, timely, authoritative information about the virus. We have a prompt in place so that, if you search for coronavirus, COVID-19 or a range of other trigger terms, you will be met, if you are on mobile, with a half-screen prompt directing you to the NHS website, where you will find the latest information on coronavirus. That has been in place since January. We worked with DHSC and the NHS to get that up and running. It is now in place in 64 countries across the world.

Then there is a question of what you do for users who are not looking for information on coronavirus and COVID-19 but are just logging on to Twitter. Right now, every time a user in the UK goes on their home timeline on Twitter, they will see at the top what we call a "moment". You will normally find moments in the "explore" tab—the magnifying glass on mobile. A moment is a selection of tweets that a team within Twitter has curated to depict a summary of a news story or a conversation on the service. Right now, if you click on the coronavirus one at the top of all UK users' feeds, you see the latest information, tweets from the WHO, from experts and from Governments.

**Baroness Kidron:** Katy, I hope I speak for everyone when I say that we all really appreciate those efforts and would love to see them mirrored in other areas. But I actually asked you whether you are finding a level of misinformation globally, co-ordinated or not, whether you are learning from it and whether you are worried by it in this particular context. I formally recognise all the good things that are going on.

**Katy Minshall:** As it stands, we are not seeing disinformation at scale about the coronavirus. We are absolutely remaining vigilant and working with peer companies and Governments around the world to very quickly respond to any issues that any of us see. For instance, it was in the news

last week that we had worked with the NHS to take down an account purporting to be an NHS trust.

**The Chair:** Katy, you are a smart woman and I do not mean that in any sense patronisingly. We are in a very particular moment in time. What do you feel that the company, Twitter, and Jack Dorsey are learning about the situation we are in that would make you a better service and a more valuable public resource?

**Katy Minshall:** Do you mean the situation we are in with coronavirus?

**The Chair:** Yes.

**Katy Minshall:** That is a good question. I do not mean to sound like a broken record, but it is about doubling down on working with researchers. We are keen to make our data available to researchers who are looking at Twitter data in the context of contagion. Just before this committee hearing, Yale tweeted out that its researchers were using Twitter to find and source information as they think about responses to coronavirus. It has confirmed what we suspected: the importance of those partnerships.

**The Chair:** These are important questions for us because we are looking for good news from the digital world. Do not think that we are being negative at all. Our report will be meaningless if it is only a tirade of bad news and destruction, so I am very grateful.

Q317 **Lord Holmes of Richmond:** Good afternoon. Other platforms have stressed how their algorithms attempt to de-emphasise sensationalist or borderline content. Is this a focus for Twitter? What is Twitter doing to promote trustworthy content on its service: #thoughts?

**Katy Minshall:** For context, Twitter as an open, public service is in a different space here. We are not a closed, private platform where you may think more in the context of echo chambers. My colleague has a saying that the hashtag pierces the filter bubble. What do we mean by that? If you looked on Twitter in December and clicked on the general election 2019 hashtag, you would have seen a wide range of viewpoints on that hashtag. If a political leader tweets something, you see a wide range of views in the replies below. There is something there about the nature of Twitter.

Thinking about what the research says, we are proud that that is being confirmed by external experts as well. The Oxford Internet Institute did a study on the UK general election looking at the prevalence of junk news on Twitter. It found that, of the sources shared 10 times or more on the service in relation to the election, only two per cent were considered junk news and the overwhelming majority were professionally made journalistic content.

**Lord Holmes of Richmond:** On the general election 2019 point, it is probably worth getting that on the record. How does the service work to curate, prioritise and align what somebody would have seen if they had put in that hashtag?

**Katy Minshall:** In the three days before and on election day itself, we injected into the top of all UK users' home timelines a prompt that we made in partnership with Democracy Club, a London-based organisation that you may be aware of, where you could insert your postcode to find out where your local polling station was and subsequently who you could vote for. That was clicked on 170,000 times. That shows, to your point on good news, the opportunity of digital technology as well as the challenges. That was at the top of people's home timelines.

In addition, over the course of the campaign, the curation team in Twitter that brings together the moments on the explore tab, which people have gone on to search for, curated over 120 moments of the most relevant and informative tweets covering a certain topic. Eight of those moments were focused specifically on myth-busting. Our team, which works with news partners, put together a series of live streams on election night so news partners could make the most of the service to get their content out to their audience. When you are on Twitter and you are looking for information about the general election, there was plenty of high-quality, informative, credible content.

**Lord Holmes of Richmond:** In that process, what is the determinant and who determines relevance?

**Katy Minshall:** I can explain how our algorithm works in general, which would apply to the election as well. On the Twitter home timeline, you may have the ranking algorithm turned on. Historically, there was no algorithm on the Twitter home timeline. You would just see, in reverse chronological order, tweets from accounts you followed. That was the case for most of Twitter's existence. In 2016, we brought in an algorithm, after extensive user feedback that seeing tweets in that way was not helpful. People did not want to have to log on every five minutes to keep on top of Twitter. They wanted to be free to log on however often they wanted and see the tweets most relevant to them.

The key thing is that we give all users the option to just turn the algorithm off. If you are on mobile, in the top-right corner there is a sparkle icon where you can turn the algorithm off and once again see tweets in reverse chronological order from accounts that you follow.

**Lord Holmes of Richmond:** That is very helpful. Thank you.

Q318 **Lord Lucas:** I have done that because I found that your algorithm was hiding a lot of my right-wing friends. There is a very wide spectrum of people I follow on Twitter and it had become very politically emphasised. Is there more you could do on Twitter to make it a friendlier place for debate? There are a lot of sentiment analysers people use on Twitter. Could you use that yourself so that you are emphasising kinder conversations? Could you do it yourself so that someone who has started a thread can tag a reply lower down, say, "This is good; this is authoritative" and bring it back towards the top of the thread?

**Katy Minshall:** These are exactly the kinds of questions we think about every day. Our aim is for the public conversation to be as healthy as possible. What we mean by that is healthy debate, conversation and critical thinking. The converse of that is abuse, spam and platform manipulation. In terms of how we get there, yes, we absolutely think about what we could do to bring to the surface the most engaging, interesting responses.

In the way our algorithm works, you do not necessarily see replies in chronological order. If you tweet something out and you respond as the tweet author to one of the replies, that is at the top. If someone you follow replies to your tweet, again, that is brought to the top because, as you point out, the chances are that that will be more relevant and interesting to people who want to keep track of the conversation you are having with your audience.

Q319 **Lord Lipsey:** Facebook put out a white paper in February that focused on procedures and processes as a way of dealing with the various problems we have raised. Your response seemed to hint that you took a slightly wider view of things. Is that right?

**Katy Minshall:** Our CEO has said that we absolutely welcome smart regulation. The key thing is that we are not waiting for regulation. I work closely with colleagues in DCMS and other government departments all the time as they think really carefully about the Online Harms White Paper. I think this was first proposed in 2017 and we are really keen to move on the concerns people have as regulation has developed concurrently.

To your question on systems and processes, Facebook's white paper made a compelling point. If you just focus on the outcomes, such as the number of reports you get about X or the percentage you act on, there is a very real risk of creating perverse incentives. It might encourage a company, for example, to narrowly define its rules. It might encourage a company to make it harder to make reports because it is not prepared to deal with the big volume. It might encourage a company to focus overwhelmingly on responding to reports at the expense of proactive work and investing in technology that could help get at the problem at scale.

It is worth regulators considering a more holistic picture, looking at systems and processes, which seems to be the view of the online harms White Paper here anyway.

**Lord Lipsey:** As a company, would you welcome greater regulation from regulatory bodies on the grounds that it would create a level playing field or are you fearful of the impact on free speech for which Twitter is such an ally?

**Katy Minshall:** We welcome smart regulation. In our view, freedom of expression means that all users feel safe expressing their unique point of

view. If collaborations with safety experts, with governments and with whoever help us get there, that is something to welcome.

**The Chair:** We took evidence on this, as you know, from the ICO and from Ofcom. We discussed the process of regulation and, indeed, we discussed penalties. I could not make head or tail of where Facebook sits on appropriate penalties. Where do you sit? By that I mean, with anonymous use of Twitter, you know who the people are. If somebody has been continuously abusive towards an MP or somebody else, what form of penalty would you regard as acceptable and who would impose the penalty?

**Katy Minshall:** I saw some of the testimony this morning. On Twitter, we can suspend someone from the service. We have to work very closely with lawmakers and law enforcement if they want to take next steps. With malicious communications, I believe only three per cent of cases result in a charge. There is a need for all sectors to work together and think about what appropriate penalties are in place for behaviours that are illegal both online and offline.

**The Chair:** The ICO made it clear that, now she has the power to charge up to four per cent of revenues, her position vis-à-vis the people she is regulating is dramatically different from when she had a ceiling of £500,000. Are you in favour of higher penalties to get better regulation and, indeed, more compliance with regulation? Do you feel that freedom of speech then gets squeezed out and there is a problem at that end of the spectrum?

**Katy Minshall:** Our global data protection officer has met with the ICO every six months for a long time now, since before this change occurred. The detail of regulation really matters before you even think about enforcement action. Clarity would be really welcome when a company such as Twitter adapts to a new framework. Do you have a good point of contact with law enforcement? Do you make your rules publicly available? Do you do proactive work to convey those rules to your users? Those sorts of systemic process requirements are more straightforward for companies to know what they need to do. When you get into the detail of trying to define something as complicated as misinformation, it puts a company in a more challenging place in understanding what it would need to do to comply with that regulation and avoid the risk of fines or equivalent measures.

**The Chair:** At the risk of being a bit boring, I am fascinated by the issue of regulation and particularly the history of the media. I was quite surprised that one reason freedom of the press was established in the middle of the 18th century was that a lot of anonymous pamphlets were published but the person who took responsibility was the printer, and the printer's name had to be on the pamphlet. A lot of printers were prosecuted but the courts continually found the printer not guilty. It was the courts deciding in favour of freedom of expression that created the environment we have in the media today. I just wonder what the parallel would be, frankly, in the online world.

**Katy Minshall:** This is an interesting question. I would not want to say anything with certainty without reading more into the scenario you outline. I guess you are getting at this question. Could there be an opportunity with regulation for people to have more trust in services such as Twitter, because a regulator is saying, "You are doing what you said"? There is value to that but, as we are trying to be transparent with Ofcom, which the Government are minded to appoint as the regulator, that does not solve the problem of transparency with users directly. Academic research and proactivity in your communication to users are important.

Q320 **Baroness Kidron:** In a way, do we look at this with a blunt instrument: regulation or no regulation? Is there not a sense in which the culture of Twitter should be really transparent and upheld, so what you say you do in your terms and your community rules is done, and the job of the regulator is to come in and ensure you have done what you have said you have done? There is then a two-step process. In this conversation, I hear the argument about freedom of expression, but you are a private company. You have your rules, standards, corporate values and so on. Even your algorithm seems to be rather to the left, we find out. That is also fine, so long as we know what you are and you are what you say you are.

In a way, the job of regulator is slightly different from the one it is always cast in, which is saying, "Okay, you have said what you are and you are doing what you said, so you get a big tick". There may be things at the edges about illegal content and so on, but in my experience, both as a company and through the Internet Association, techUK and various other forums, there is always a pushback about regulation and not a huge amount of transparency. I appreciate all the positive things you have said but, on this point, is that a healthier way for us to look at it as a committee: we hold you to account for what you say and give the regulator teeth and possibly a fist to make sure you do what you say?

**Katy Minshall:** This is the opportunity with the approach you outline. If we had fixed in law what we thought of as priorities for the internet space in 2017, the chances are that there is a whole host of work that that would not have captured. The risk of the regulator, Ofcom, or the Bill itself trying to fix specific codes or requirements on a whole multitude of issues is that it will not be as future-proofed as it could be. If you adopt the approach you outline, trying to get at whether companies are doing what they say they are doing, companies might be better able to respond to emerging issues more rapidly.

**Baroness Kidron:** For clarity, there always has to be a floor if not a ceiling.

**Katy Minshall:** Yes.

Q321 **Lord German:** Can I turn to what you do allow the users of Twitter to do, which is to hide the replies to their tweets? Can you tell us how many people, either globally or in the United Kingdom, have taken up that opportunity and are using it?



**Katy Minshall:** To give some context, we want everyone on Twitter to feel safe expressing their viewpoint. To get there, we took the view that we had to change how conversations were working. The issue, as we have talked about, was that repliers were detracting from the tone or topic that a tweet author wanted to have with their audience. Last year we took the decision to trial the ability to turn off replies in Japan, Canada and then the US. Just a few months ago, we rolled that out globally to all users. I do not have the results of the global rollout yet, but I have the results of the trial if you would like me to read them out.

**Lord German:** Yes, please.

**Katy Minshall:** People mostly hide replies that they think are irrelevant, off-topic or annoying. The option is a new way to shut out noise. Of the people who hide replies, 85 per cent are not using block or mute. People were curious to see how public figures like those in politics and journalism would use this update. So far, they are not hiding replies very often. In Canada, 27 per cent of people who had their tweets hidden said they would reconsider how they interact with others in the future. They also thought it was a helpful way to manage what they saw, similar to muted keywords. We learned that you may want to take further action after you hide a reply so we now check to see if you also want to block the replier. Some people mentioned that they did not want to hide replies due to fear of retaliation as the icon remains visible. We will continue to get feedback on this. These are early findings and we look forward to continued learning as the feature is used by more people.

**Lord German:** I did not quite get from that the percentage of users in Canada and Japan who took up that option.

**Katy Minshall:** I do not have those figures for you today but, after watching the hearing this morning, I have contacted colleagues in the US to see if there is further data we can provide for you.

**Lord German:** Thank you very much for that. The other issue related to this is how easy it is. I was interested that you said politicians in Canada used this more often than anybody else. Of course, politicians are quite robust people, although we do not like being trolled and abused. The people who have an opinion, lay it out and then get abused find it most difficult to sort this out. You have a choice, you say, to block or hide; it can be reawakened. Is that all user-friendly? If the people using it are mostly the politicians and the sort of people who are expecting to get that response, clearly it is not good enough for the people who feel most hurt by it.

**Katy Minshall:** To clarify, we saw that public figures, such as those in politics and journalism, were not hiding replies very often. To your question about what next and what more we do with this, we announced in January that, partly as a result of this, we would now look to experiment with the ability to turn off replies or to specify who could reply to your tweet. At the moment, if you tweet, anyone can reply to that. We are going to trial conversation controls. You can keep it as it is;

you can have it so that only people you follow can reply to that tweet; you can, for example, invite the Committee here to reply to your tweet; or you can turn off replies completely. The thinking is that putting those controls in place or making them available might encourage people to feel more confident about tweeting, because they are less concerned about the barrage of replies they may get below.

**Lord German:** Can I go back to a question you spoke about right at the beginning? I understand that hiding a response can automatically happen, or there is a tool for making it happen, where you do not have an email address, phone number or profile picture. From the conversation at the beginning, I thought that you had to have their email address or phone number, to know that they were a real person behind whatever was said. This is not really a user device, because the only one that knows whether you have a registered email address or phone number, or are not changing a profile picture, is Twitter. Is Twitter doing this on request or do you do it automatically?

**Katy Minshall:** This is the ability to hide replies.

**Lord German:** Yes. It is also about users restricting the notifications they receive so they are not informed when they are mentioned by someone who has not had their account authenticated. I thought you had said that everybody who had an account had to have it authenticated.

**Katy Minshall:** You are absolutely right. You can control what notifications you get from Twitter. This is particularly useful to politicians and others in public life who get high volumes of notifications. As you say, you have the option to get notifications only from people who have verified an email address. You can also turn off notifications for everyone except those you follow.

By the way, tying back to your earlier comments, we want to do as much as we can to make people aware of the tools. A good thing, which may be a challenge as well, is that on Twitter you can mute, block and hide replies, limit notifications and, in months to come, turn off conversations. That is a lot for users to take in. We are going to look for really good opportunities to clearly articulate how you can use these safety tools and in what context.

It is important to mention that we are looking at whether external developers can make use of the "hide replies" API so you can automatically control what replies are filtered. For example, a safety organisation might create a plug-in where it would automatically hide replies to your tweets that included certain words. That is a really blunt example, but it gives people more control and opportunity to shape their experience on Twitter.

**Lord German:** I just want to get at the very important point that I was repeating from an answer you gave earlier. You said that all accounts had to be authenticated. Here is this device with which users can switch off any unauthenticated account. Why would they have an account if you had

not authenticated it?

**Katy Minshall:** Not all bots are bad; that is one way of putting it. We will ask for an email address or a phone number, but you can operate multiple accounts. You can have a bot on Twitter without necessarily a specific human on the other end. To give you an example, there is a bot right now that reminds people to wash their hands and goes off every so often. To give you another example, last year there were lots of petitions related to Brexit. The petitions website crashed a few times because so many people were going on it to see how many people had signed up. There was a bot that would, every 15 minutes, give the updated number of signatures, trying to prevent that website from crashing by giving people the information they needed.

Yes, we ask for an email address and a phone number, not least for the purposes of cybersecurity, but historically people have created accounts on Twitter that are bots, with no human on the other end, for good reasons.

**Lord German:** You allow bots.

**Katy Minshall:** Yes. We do not have a rule against bots. We have a rule against automated accounts and multiple accounts trying to spam a conversation or engage in a malicious way on Twitter.

**Lord German:** Someone will have to explain to me the difference between a bot and an automated account, but there we are.

**Lord Lucas:** Looking at how we should keep tabs on all these changes going on and their effects, would you be content if Ofcom had a duty to comment on the way in which you ran your site and whether it was supportive of or antithetical to democratic engagement of citizens? In other words, we give Ofcom a general duty to keep an eye on what you are doing and say, "Hang on; the way you do that is causing problems. Why not do this, because it would make things better?"

**Katy Minshall:** That is the right question to ask. We should all be talking about how we are engaging with democratic discussions. It is an interesting idea. I did not watch the Ofcom testimony. If they had some thoughts, it would be interesting to read them. If you try to specifically define misinformation, which therefore translates to what services such as Twitter do or do not take down, and what platforms should or should not do, I can imagine some issues in the context of democracy. It might be difficult, but your question is the right one. Yes, every service should be thinking about how it reflects political discussions and engages in democratic processes.

**The Chair:** I think you might find yourself quoted in the report.

Q322 **Baroness Kidron:** Can I ask you about the new mechanics of conversation hiding and so on? Are you not concerned about what we have seen on other platforms where conversations get private, people try to monetise that and powerful people in the space ask for money to get

into the private conversation? That is one version. Another version is even less savoury about grooming and so on. I am interested, and I may have misunderstood what you said, but you seem to suggest that there are a whole number of things that could mean you go out of sight in some way. The answer may be no, but I am interested to know whether those considerations have been in mind.

**Katy Minshall:** Much like our political advertising ban, we very much stand behind freedom of expression and freedom of speech, but freedom of reach is different. To be totally clear, what I have talked about is in no way a change to Twitter's public, open nature. Even if you were to hide my reply, that does not prevent me tweeting that you have hidden my reply. It does not prevent me sharing that viewpoint.

**Baroness Kidron:** Nothing you have described is actually to the side of the public platform ultimately.

**Katy Minshall:** Correct, yes.

Q323 **Lord Harris of Haringey:** You reminded us earlier that political message reach should be earned, not bought, so I am going to ask about that. You have banned political advertising. I understand Mike Bloomberg, in his brief and ineffective presidential campaign, paid influencers to promote his campaign and paid regular members of the public to tweet positively about him. Are you going to stop that and, if so, how?

**Katy Minshall:** The political ban was informed by three principles. As you have just said, political message reach should be earned and not bought. In addition, advertising should not be used to drive political outcomes. The interaction between civic conversations and microtargeting is not yet sufficiently understood and poses new risks, so we should adopt a cautious approach. We already ask that any user posting an advertisement as an organic piece of content discloses its commercial nature to users. We have best practice on our website, such as saying "#ad", and we will continually evolve our rules and processes as the nature of content changes on Twitter.

The key thing here is that, if a political party makes an arrangement with a high-profile individual or celebrity to post something on Twitter, it still has to earn that reach. It cannot buy the reach on our service.

**Lord Harris of Haringey:** If it pays the celebrity, who has a high reach because they tweet about football or whatever else, is that not the same?

**Katy Minshall:** The question is the right one. The challenge is that we, Twitter, would not have visibility of an arrangement made between a political party and an influencer or high-profile individual. That political party in the UK should be disclosing its digital spend to the Electoral Commission. This Committee is talking to Twitter, Facebook and YouTube but in the 2017 general election, not just the most recent one, we saw political parties creating their own apps. There are now over two million apps on the Android app store. The way people engage with democratic life through technology is not fixed. Platform by platform, service by

service, may be the wrong way of looking at it, when there is already an Electoral Commission that can make these decisions and enforce them in that way.

**Lord Harris of Haringey:** What I am trying to get at is whether there are ways to circumvent your rules. We have seen the phenomenon of Twitter storms to get particular political messages across. Those are not paid for but may use a network of activists or whatever else. Would you see those things falling into the category of political advertising?

**Katy Minshall:** Sorry, can I ask what you mean by Twitter storms?

**Lord Harris of Haringey:** I was rather hoping you would know because, when I hear there is going to be a Twitter storm about a particular issue, I try to keep away from it. As I understand it, hundreds or thousands, depending on the organisation concerned, deliberately tweet on the same topic at a specified time, so they get a trend and the trending sign appears. This is then seen as an effective way of campaigning. That can be orchestrated either by the sheer enthusiasm of a political party's supporters or by paying people to do it for you.

**Katy Minshall:** It is already against our rules to artificially amplify a conversation. Political party or not, you cannot create a load of accounts and ask them to all tweet the same thing at the same time, and like each other's tweets, to try to get something to trend artificially.

**Lord Harris of Haringey:** How do you stop it?

**Katy Minshall:** To go back to my earlier response to Baroness Morris, it is a mixture of people reporting it to us and our proactive technology. A computer can detect extremely high-volume tweeting on the same hashtag and intervene.

**Lord Harris of Haringey:** Hang on. COVID-19 is no doubt a very popular hashtag at the moment and it has suddenly peaked. You do not pick that up but you would pick up if somebody was tweeting something about universal credit.

**Katy Minshall:** Whether it is about COVID-19, universal credit or any topic, if someone or a group of people creates a series of fake accounts to try to orchestrate or amplify a conversation, that is against our rules.

**Lord Harris of Haringey:** Does it get taken down and just disappear?

**Katy Minshall:** There are all sorts of ways we could intervene. We could reduce the visibility of the trend. We could find and remove the accounts, and suspend them from the service. Yes, we would intervene in multiple ways.

**Lord Harris of Haringey:** How quickly could you do that?

**Katy Minshall:** It varies. We want to be as quick as possible and stop things before they reach the point of trending. In other cases, we would ask people to report that to us.

Q324 **Baroness Kidron:** I am not sure you have answered the first half of my question, but I want to give you the opportunity to say anything you feel you have not said, which is really about how you come to design and roll out new products and the forces that might make you choose to do so.

**Katy Minshall:** Our aim is for the conversation to be useful, interesting, timely and informative. The products we roll out, the products we change and the products we deprecate are trying to serve that purpose. As well as the work of the product trust team, which I outlined, it is worth talking about our prototype app. This is something we launched last year and invited ordinary Twitter users to join, where we will test new designs or products directly with users, to get feedback from them and to see how a particular change might roll out in reality before we consider rolling it out on the service writ large. As well as the internal safety by design approach, we have an external test bed to look at these new innovations.

**Baroness Kidron:** I want to ask you about safety by design. I understand that methodology, but so many problems in this area have been dealt with retrospectively. Could we look at the development of Twitter and say, "In August 2017, it changed from looking over its shoulder to looking forward in terms of its product"? Has there been a change of attitude and what happened to cause that?

**Katy Minshall:** I have a timeline of safety changes we have made over the past few years, which I will provide to you after this hearing. No one at Twitter says that we have everything perfect—far from it. There is a reason why, if you go on our blog, the title right now talks about our safety and it says, "Progress and more to do". Because we are a public open platform and there is so much research, the challenges are well documented on Twitter. I am sure that, in the way that we have already changed a lot over the past few years, particularly using technology more and more, we will continue to change and adapt again. We are starting to see really significant results that give us confidence that our approach is paying dividends.

A few months ago, we announced that we had seen a 27 per cent drop in bystander reports. If I was abusing Baroness Morris and Baroness Kidron reported it, that would be a bystander report versus Baroness Morris reporting it herself, which would be first-person. That drop in reports suggests that people are seeing less abuse on Twitter than they used to. That gives us confidence that the investments we have made and the technology we are putting in place is having an impact. Like I said, no one at Twitter would say we are perfect. No one would say we do not have work to do.

**Baroness Kidron:** I would be very interested in seeing your timeline. If you were willing, I would be interested in your view about what preceded the change, whether it was regulatory, whether it was reputational,

whether there was a particular incident and so on. That would be very interesting and you may indeed star in our “good news” chapter.

I want to ask about the business model. I do not want to ask just about Twitter. If you think about organising all the world’s information, connecting all the world’s people or creating a fantastic public conversation, these are all things that everyone in this room would like to see happen. But it has turned into a bit of a war of attention, growth and user numbers in order to advertise. I wonder whether you have anything to say about the culture wars, if you like, between your mission on paper and the need for growth because of the business model. Do you consider that there is another business model available for Twitter?

**Katy Minshall:** Principally, our business model relies on people coming on to Twitter every day. That is how services such as Twitter succeed. People are not going to come on if they do not feel safe. They are not going to come on if they are not seeing useful, relevant, high-quality information and content. Our business model and our priorities as a company of safety and health are absolutely aligned.

**The Chair:** You are suggesting that people do not come on to Twitter in order to shout, abuse and insult but that they come on to listen.

**Katy Minshall:** People come on to Twitter for all sorts of reasons. We define 40 per cent of users as listeners, in that they do not tweet, but they are there, interested in the news, sports and entertainment. I talked about K-pop. We published research last year looking at the UK specifically and the communities that have formed. There are the big ones that you might expect, such as gaming, sports and football, but communities have also formed around cheese, wine, gardening and bird-watching. We are here today to talk about political Twitter, but there are so many different forms of Twitter that go far beyond talking about politics or Brexit every single day.

**The Chair:** Is that not the reason why we should, in theory, be helpful to you? We are talking about the potential reputational damage that then disrupts or upsets all the good things you could be doing. Because of exactly what is happening on social media, the reputational damage argument dominates in exactly the same way that, unfortunately, on too much social media, the negative platform manipulation dominates. It may sound strange but we are endeavouring to be helpful to you, because we believe that, as a sector, you have a reputational issue. You at Twitter probably have it less than anybody else.

**Lord Harris of Haringey:** While Baroness Kidron was asking her question, I typed in the search term on Twitter “#TwitterStorm”. What immediately came up was a tweet from @Artists4Assange yesterday saying, “Why should #JulianAssange be urgently released from British custody? Please answer in your own words & retweet! If we get 100 retweets, @Artists4Assange will launch another #TwitterStorm calling for Assange’s immediate release”. Now, is that a manipulation and, if it is a manipulation, is it something you pick up? Obviously you did not pick it

up and, as far as I can see, it got 96 retweets, so it did not meet its own criteria. The point is that, if you have a community of people, you would use something like that to try to amplify it. Are you saying that you would or you could stop it?

**Katy Minshall:** I do not want to give you any incorrect information, so let me take that example away.

**Lord Harris of Haringey:** I suspect that that example is irrelevant. It is clearly a tactic that people are aware of. Similarly, the next one is about finding a lost dog, #FindBertieLakeland. There is a picture of a very fetching dog. All I am saying is that there seem to be people on Twitter who see this as a tactic to maximise their message. Finding the dog may be a very beneficial one. With Julian Assange, it depends on what you think about him. It seems to be a manipulation that people are well aware of. Can you track and deal with this, if you think it is unhelpful, abusive or anything else?

**Katy Minshall:** Yes. To your point, there is a difference between someone retweeting a missing pet or a missing person, encouraging people to share to get as much reach and exposure as possible, and me, with my friends, saying, "Let us create a load of fake accounts and get this particular hashtag to trend", which would be completely artificial in nature. There is a difference between that and saying, "I am watching a film. Retweet and tell us what your favourite film is". That would be an authentic engagement versus disrupting a conversation on particular terms in a very non-genuine and artificial way.

I am sorry that that is not a very black and white answer, but that speaks to the fact that it is not a black and white tool. Bad actors game it whereas others use it every day for good.

**Baroness Kidron:** We often get the cast of bad actors. I absolutely agree that they are out there and it is troublesome to deal with them. The Committee struggles with the more systemic issues. Where is the system itself creating a problem or unfairness? I want to quote from the developer who invented the retweet. He said it was like handing a loaded weapon to a four year-old, due to the way the feature prioritises sharing content at speed rather than encouraging users to read or think about the content they are sharing. He is making the point there, from an internal perspective—and, as was pointed out, the more traffic there is, the more Twitter is winning—that there is something intrinsic in the conversation that is destabilising to productive conversation. That is slightly different from whether people who like cheese are tweeting about gorgeous cheese.

**Katy Minshall:** Our CEO has been pretty upfront about this. He has said he is interested in rethinking the fundamentals of the service. The fact that we are considering rolling out the ability to just turn off replies to tweets speaks to that. Yes, we absolutely think about the fundamental parts of Twitter. Again, this is where we have to work with researchers. We have a partnership with UC Berkeley right now, looking at the



interaction between machine learning and social systems. What impact does it have on people when you have a certain algorithm in place? Does it cause them to change their behaviour?

Q325 **The Chair:** That is why I challenged you on the use of the word "discourse". It is not discourse at the moment. As long as we know it is not discourse, we are not likely to be quite as confused as we sometimes are. Tweeting is much closer to a football crowd chanting, but that is a personal view.

In what ways are you working to improve academic access to Twitter data?

**Katy Minshall:** I have already talked about our API and the information operations datasets we now have on our website. The step change we are really keen to work on now, rather than just making data available, is to work directly and closely in research partnerships. We have a dedicated team at Twitter working with academic researchers. In January, we launched a new hub for academic researchers so they could quite quickly find out what information they could get from Twitter and how to get there.

Last week, as a result of feedback from our engagement with the academic community, we made some changes to our developer policy. They said that it would be helpful if they could share an unlimited number of tweet IDs or account IDs with their peers for the purposes of peer review, and our policy was getting in the way of that so we have added that clarification.

As a good example of the direction of travel for us, which is on our website right now, last week we announced a new partnership with the New Zealand National Centre for Peace and Conflict Studies. It did some work looking at the aftermath of the Christchurch massacre and the way Twitter was used, overwhelmingly to share grief and outrage at the atrocity. Its research with us is going to focus on how social media can be used to promote inclusion and tolerance, and fight against exclusion and intolerance.

We are keen to double down on our partnerships with researchers. Thinking back to our discussion about regulation and the White Paper, what would be great, before the next iteration of the online harms Bill or White Paper, is detail on how companies could be incentivised to work more closely with the research community.

**The Chair:** Your company should be very proud of you. Thank you very much.

**Katy Minshall:** Thank you.