



HOUSE OF LORDS

Select Committee on Democracy and Digital Technologies

Corrected oral evidence: Democracy and Digital Technologies

Tuesday 17 March 2020

10.30 am

Watch the meeting

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

Evidence Session No. 24

Heard in Public

Questions 297 - 310

Witness

[I](#): Karim Palant, UK Public Policy Manager, Facebook.

Examination of Witness

Karim Palant.

Q297 **The Chair:** Welcome. As you know, this session is open to the public. The webcast of the session goes out live and is subsequently accessible via the parliamentary website. A verbatim transcript of the evidence will be taken and put on the parliamentary website. You will have the opportunity to make minor corrections for the purposes of clarification and accuracy. Please introduce yourself for the record and then we will move to the first question from Lord Black.

Karim Palant: Thank you. I work for Facebook on public policy for the UK and have done so for four and a half years. I have worked through a number of elections at the company, including the 2019 general election. My background is in politics: I worked for the Labour Party for many years, including in the elections unit during the 2015 election, so I have a bit of experience on the other side of things as well.

Q298 **Lord Black of Brentwood:** One of the biggest issues that anybody looking at this area faces — I am talking principally about independent researchers, including people attached to this Committee — is Facebook's reluctance to share data for study. The reason that is often given is an interpretation of GDPR and the obligations under GDPR and the Data Protection Act. Could you say a little more about that and explain why you worry about incurring liability for sharing data with researchers drilling into this area?

Karim Palant: Sure. Over the course of this Committee's investigations and inquiries, things have moved on a little. The good news is that we have a collaboration with an organisation called Social Science One where we have released a very large number — I do not want to get it wrong, but I think it is more than one quintillion megabytes — concerning a set of URLs, the engagement with those URLs that were shared more than 100 times, how they were engaged with and the levels of reaction to them. It is a database over two years. That dataset was first discussed with Social Science One 18 months or two years ago and we announced that we would be releasing that data.

As we worked with them it became clear that the safeguards that needed to be put in place needed to be much greater than those that were initially put in. Part of that, it is fair to say, is that Social Science One made clear that they recognised that we have been subject to a great deal of regulatory investigation around the Cambridge Analytica issue. That involved providing data to an academic for research purposes. The initial issue that arose was that that was then misused, although I have not been that close to those proceedings, so I would not want to delve too deeply into that. The thing to say is that in those circumstances, ensuring that this was done in a privacy-protective way was highly important.

It has taken more like 20 months and then two months to get to that point. It is, of course, a source of great frustration to a great many

academics that we have done that. However, a lot of the technology and the technical fixes to ensure that that data can be transferred in a privacy-protected way has been done and was done for these purposes. Much of the thinking, weighing up the different risks, the legal analysis and the designing of how to share this database has been done.

In my experience — again, I was not directly involved in that project — once these things have been done once and you have built the systems and the tools, it is a lot easier to learn the lessons from it and do it much more quickly in future. I cannot guarantee that those discussions are happening right now, because I am not close to it, but I know that it was announced in the last two or three weeks that we had finally got to the point where that was now handed over.

It was more challenging: Social Science One has said that it disagrees with our analysis of the legal risks and so on, but as it said itself, it was not the one that had a \$5 billion fine from the FTC and it respects the fact that we may well have a different analysis. That is where it has ended up. It has taken longer than we would have liked, but we are a lot further along than we were when your investigation started.

Lord Black of Brentwood: I appreciate that progress has been made with Social Science One, and that is good, but these things take time. Did you share with them your, presumably written, legal opinion about the application of GDPR in this area?

Karim Palant: I was not involved in these discussions directly but, as I understand it, there was a constant process of iteration over that time, a sharing of potential technical fixes and challenges and a dialogue to try to understand where what they saw as our conservative legal view differed from theirs and how we might technically get over the line on that. I was not directly involved in those conversations, but I know we had regular, ongoing dialogue where we explained the barriers that we faced and how we were seeking to overcome them. The fix that has come about was not something we could have done overnight, and in doing it we would have had to make sure that it was going to be of value to them before we did it. We certainly shared that conversation along the way and were in dialogue with them throughout.

The mechanism they have described in their release around this data was something called “differential privacy” which has been used, for example, in the US census. Effectively, when you query the data, the closer you get to identifiable information, the more it slightly obfuscates what you are looking at, so you cannot triangulate and identify an individual from it. That is a complex way of constructing a database, and it was done in collaboration with Social Science One, not separately.

Lord Black of Brentwood: Given how important that whole issue is, would Facebook be prepared to publish legal analysis in that area, so that parliamentarians, among others, could engage with it?

Karim Palant: That is an interesting question, and I would be very happy to come back to you on it. As with all these things, we all would benefit from greater clarity and comfort from regulators on where they see the boundaries. For example, the Information Commissioner, who I know was here yesterday, has made it clear to UK-jurisdiction companies and agencies that she recognises that in a crisis such as we face with COVID-19 a balance needs to be struck between safety and security and data protection. Therefore, some of the restrictions that we might place on analysis or sharing of data have to be weighed against those safety concerns. That is an interesting model. It is certainly something that we have experienced in other areas. Greater comfort and clarity from regulators, and greater dialogue with them, would be welcome and we would wholeheartedly endorse it.

Lord Black of Brentwood: Thank you. That is very helpful. If you could come back to us on that point about publication, we would be grateful.

Karim Palant: Absolutely.

The Chair: Lord Black is suggesting that other research organisations could look at that legal advice and say, “Well, we could comply with that” or “There are elements here which no self-respecting university could comply with”.

Karim Palant: Absolutely. I accept that. Social Science One being an independent agency helps in this regard, because it essentially holds the data and then allocates access to it to universities. It is therefore able to give clarity about the legal basis on which people have access. That can help, but I agree that the more comfort and transparency everyone involved in this process can have the better. It is clear that we have all learned lessons from the past.

Baroness Kidron: I want to pick up your point about the ICO guidance and the FTC. I do not notice Facebook being shy in engaging with regulators around the world on other subjects. There has been such a move on the part of the FTC and the ICO to have research-app data. Did you take the problem you had with GDPR and privacy to the regulator and say, “Look, we want to open our house to these researchers. Our legal people are saying this. Could you give us a clear marker that you won’t come after us?” Did you have those conversations during the 18-month process?

Karim Palant: Again, I was not directly involved in that. The FTC and the Irish Data Protection Commissioner would have been the people to have that conversation with. I know for certain that there are a great many areas where we have that kind of conversation and say, “Look, we would need comfort around being able to use some of our classifiers for certain purposes because we are worried about the interrelationship here, but there is a safety consideration” and so on. I know that we do that. I do not know 100% when or whether that conversation took place, but, again, I can come back to you on that.

Q299 **Lord Scriven:** On the differential privacy that you described, some commentators, particularly in the social science world, describe it as extra technical noise over what would normally be expected for privacy. Why has Facebook decided to do that when there was no need to do it to comply with data protection rules? It stops certain aggregation by social scientists.

Karim Palant: I think Social Science One has described it as effectively bringing you the accuracy and granularity of a very large sample.

Lord Scriven: May I stop you there? I am not interested in what they say; I am talking about people who can potentially use it. They have described differential privacy as extra technical noise over and above that which they would normally expect when interrogating a database. Facebook insisted on that; no one else did. That stops certain data investigations taking place. Why did Facebook insist that such extra technical noise be included?

Karim Palant: The short answer is the privacy of our users.

Lord Scriven: They did not have to do it to be compliant with the law. It is over and above what was compliant. Facebook insisted on it. Why?

Karim Palant: As I described before, some people have the opinion that it is not necessary to comply with the law. Our legal advice was that it is.

Lord Scriven: Are you prepared to put that legal advice on record? That is the crux of whether the technical noise stops social scientists getting to data that they would otherwise be able to get to.

Karim Palant: As I have said, I will look into that and write back to the Committee. The value of this data as described by Social Science One is that it would be like having a very large sample size and studying an issue on that basis. It is therefore hugely valuable. It is an unprecedented data source that has not been provided by any company before. It is clearly a new and emerging world that we are working hard with researchers and others to learn about and get better at. It is an incredibly fraught area that has been of huge regulatory and public interest from the other direction over the past two years.

Lord Scriven: Let us be clear: other companies — not necessarily in your space — have allowed people access to anonymised data, including digital data. You have put extra technical noise in; others have done it without. All I need to understand is what makes Facebook so special that it needs to put in this extra technical noise which stops validation of data which would be possible without it. It is key, because it appears to the lay person that you are trying to restrict analysis of the data by putting in this extra technical noise.

Karim Palant: We are, I think rightly, under huge, increased scrutiny from both regulators and the public on these issues. That explains why we are particularly careful in this space. We are also dealing with huge quantities of very sensitive data around people's political views and their

engagement with political information. That would also explain why we might be in a slightly different position.

We are also talking about providing a solution that goes an awful long way to allowing researchers to do the kind of research that they have been keen to do over a long period. This is a very recent conclusion to the process; I think that a lot of researchers are still to interrogate the dataset that has been provided. I am certain that a whole range of interesting insights will come from it and a whole range of interesting questions as to why there is not more data in the following areas. That is to be expected; I would not expect it to be the end of the conversation by any stretch of the imagination.

We have provided URLs as part of this, which apparently is unprecedented. That is something else that we have worked incredibly hard to go further on.

Lord Scriven: It comes down to this extra digital noise. I want to understand why that was put in.

Karim Palant: Specifically, if you have several datasets and data points about individuals or individual pages, you can, if you are highly technically adept, triangulate and identify personal information.

Lord Scriven: But you can do that with any dataset. The issue is why Facebook went beyond the norm by adding this. You are going to write to us; it is really key. The Committee needs to understand why you have put in this extra technical noise which stops normal interrogation. Despite the point about triangulation in other areas, I want to understand why Facebook's legal opinion was that you needed to put it in, because it stops certain analysis of data.

Karim Palant: Absolutely.

Q300 **Lord Lucas:** If your newsfeed is found to include misinformation and hate speech — perhaps you have some opinions on who should be the judge of that — how should you be held to account for it? Should you be fined, or is there a better way of dealing with things?

Karim Palant: Any statutory regulatory system with no penalties would come under a lot of criticism, so we fully expect there to be penalties, whether fines or anything else, as part of any regulatory regime. In a sense, the question of how big the fine is, what kind of punishment there will be and so on is easy. If you are being punished because you did not do something, the really complicated question is: what is that thing? What things do we need to do? What does the regulator want us to do? Who is the regulator or judge of these things? What is proportionate? What proportionate expectations might a regulator have?

To be frank, my hope, and my full expectation, is that, in any regulatory environment that might be created in the UK, we would be an awfully long way from penalties because we would engage directly and proactively with the regulator and be in full compliance from the

beginning. That would be the goal. That would include iterative conversations with them about their expectations and what is proportionate to prevent the spread of hate speech or anything else.

Initially, the priority would be to have a sense of the regulator's expectations and, secondly, an information-sharing process through which we would provide it with information about how we were doing in that regard. It would be able to interrogate that and interrogate us. Then, if there was room for improvement, we would look for a conversation with them in which they could tell us where they thought there was room for improvement and get undertakings from us to do that.

At that point, if there was disagreement further down the line, you would of course expect there to be penalties in any regulatory system. However, an awful lot of water has to go under the bridge at this stage in the UK about that first, big process. That is much more of a dilemma. But yes, absolutely, we fully expect the regulatory process to have penalties, including fines, which have been talked about, and so on.

Lord Lucas: Would it make things easier from your point of view if the first requested response would be that you either linked whatever was being complained about to a fact-checker or gave some rebuttal of what they were saying, particularly when there is a question of whether something is misinformation? If you can argue about the side of a red bus with a straight face, then defining misinformation is quite difficult. Putting up a counter-argument may be an easier way of dealing with that, but would that not work from your point of view?

Karim Palant: There are two things to unpack there. First, there are questions to do with which harms the regulator will focus on first, how it will address them, how it will talk to companies and how companies currently address those harms. I would break misinformation out from, say, hate speech, nudity and those kinds of things, which are not allowed on platforms such as ours. I break those out because we are very clear about them; we have had policies in place for over a decade that set out what is allowed on our platform and have built and built our enforcement tools to catch and remove that stuff.

Misinformation is considerably more complex. Even if this was within the scope of what the regulator was supposed to look at—I know Ofcom was here yesterday and said that that was still unclear—and it wished to engage on it, there is a whole range of ways in which you can tackle it. We break it down with a policy to remove, reduce and inform, but a regulator may choose to break it down differently.

There are certain types of misinformation that we simply remove, for example, if it might lead to real-world harm or if there is incitement to violence. In the case of the current COVID-19 crisis, for example, misinformation could lead people who should seek treatment to not do so, or lead people to use treatments that might cause harm or prevent them getting treatment. There could also be misinformation about the provision of healthcare which might lead to other real-world

consequences. We would remove that kind of misinformation. There are also the actors and individuals engaged in misinformation, such as fake accounts. We remove over 2 billion fake accounts every quarter. That kind of removal is one aspect of what you might do.

The “reduce” policy is the kind of action you talked about. You might identify something that is at the margin. We know that, wherever we draw the line on our policies, the stuff that would get attention is the stuff that goes right up against that line, wherever it is; it feels sensationalist and shocking, so people engage with it more. We try to identify at scale content which might be close to those lines and downrank it relatively compared to other content. Also, as you said, when content is flagged to fact-checkers by users, the media and others, fact-checkers review it and it is marked to show that a fact-checker has marked it as false. It gets a screen over the top, as during the general election, and it gets downranked in the newsfeed and so on. That is one aspect.

The third aspect is to inform, for example by ensuring that the correct information can get to people and that legitimate news sources are able to thrive on our platform. I imagine that there will be considerable conversations about this in the UK in the coming months and years, whether through the online harms process or the outcomes of the Cairncross review into the sustainability of journalism. There will be conversations about how we can all do better at all three aspects, but I do not think the regulator will mandate that this is the silver bullet and that you must do this one thing. When the regulator is confronted with the evidence, I do not think that that is how it will respond.

The Chair: Lord Lipsey has a supplementary question but, first, I should say that I do not want you to truncate any of your answers, because they are important to us. This session will undoubtedly run over, so I hope that is not inconvenient. If we lose one or two members of the Committee, so be it.

Karim Palant: Quite a few things have been cancelled recently, so —

The Chair: Excellent answer.

Q301 **Lord Lipsey:** I understand what you are saying about dialogue with the regulators, but one of the things some of them have said is that they would like statutory codes of guidance. Would Facebook be happy if such a code of guidance was applied in this area?

Karim Palant: Do you mean a statutory code that would then apply to what, say, Ofcom would do as the online harms regulator?

Lord Lipsey: That is correct.

Karim Palant: There is a spectrum across a shopping list of things which Parliament might wish to demand that a regulator pursues that is essentially done ad hoc, which might distort what a regulator does and stop it being flexible in the face of emerging problems. For example,

health misinformation might not have been on the list, but it is suddenly a hugely important thing, so that kind of list would be overly restrictive.

The flipside is that, if there is no framework through which a regulator can judge how Parliament is asking it to prioritise harms, and proportionality from Parliament on how to balance freedom of expression, innovation, safety and privacy, that will not work for the regulator. Ofcom is currently in an unenviable position, because it can see that it has potentially quite big responsibilities coming down the track but, at the moment, Parliament has not spoken about what it would like those responsibilities to look like.

Overall, my sense is that we need a way to establish what a harm is and a framework that gives Ofcom the tools to make that assessment. I think Ofcom has talked about the need for it to be quite transparent about how it assesses the level of harm and risk and how it balances those risks in advance of any potential crisis or question that it might get—being transparent from the get-go. That will be the right way to do it. I do not know whether giving Ofcom that framework would involve statutory codes or whether it can be done on the face of a Bill; I am not the expert on that, but I would say that it is definitely up for discussion. It is certainly not something that we would be afraid of.

Lord Lipsey: It does not have to be on the face of a Bill. It could be done through a statutory instrument.

Karim Palant: The Bill itself may give enough of a framework, but it may be that Ofcom wants to draft a code and have it signed off.

Lord Lipsey: You were complaining that there might be a lack of flexibility. If health came on the agenda, it would be much easier to change something done by statutory instrument than something on the face of a Bill.

Karim Palant: Yes.

The Chair: May I ask you something simply in order for you to possibly correct the record? Yesterday, in a question to Ofcom, I used two examples of harms — you said “the thing” — which struck me as extraordinary. “Any Facebook user can pretend to be a relative or friend, and to declare any other Facebook user deceased”. That was from Dave Eggers’ lecture to PEN 18 months ago. Is that true?

Karim Palant: We have a process in place — this has happened with relatives of mine — whereby if somebody is deceased who was a Facebook user, you have a choice as a family member: you might want the account deleted, or you might want it memorialised — it stays up, you can still access the photos and so on, and you can list the person as being deceased so that for example birthday reminders are turned off. There is an extensive blog about this explaining how all that works.

In order for that to work, somebody needs to tell us that the person is deceased, which is obviously incredibly difficult to manage, especially at a

time when inevitably people are grieving and so on. So we try to make that as simple as possible. As I understand it, there are two ways in which this can be done. One is that you as a Facebook user can set who you would like the person to be who, if such a thing were to happen, can tell us and make your intentions clear. You can set your intentions, so if you want to do that, you can.

There are other ways in which we will ask people to prove their connection to the person, such as a death certificate. I am very happy to send you the full breakdown of exactly how that works. We try to make the system as low-lift as possible for people who are grieving, so it is perfectly possible that people could do that maliciously, and so on. That would obviously be hugely unfortunate, and I am confident that we have looked at those systems quite recently and tried to work out—

The Chair: On the point about the malicious opportunity, where would the penalty go in that instance? Is it you because your compliance processes have not checked, or is it the person, with your help, checking back on the person who posted the false death notice?

Karim Palant: This is a problem with a lot of what goes on in the online space. The only punishment we could have for somebody who was trying to abuse our systems would be their removal or blocking them from using those tools again. We cannot pursue people through the legal system, for example, except in extreme circumstances, so people will try to push the envelope of what they can do using our tools. In that circumstance, thankfully the person involved is not deceased, so at least the harm is less than in the reverse circumstance.

The Chair: You know that is very arguable. The person who is deceased is probably the one suffering the least harm.

Karim Palant: No, but I can imagine why family members would be very upset if someone were to do the reverse, which in many cases would be worse. It is clearly something that we should be designing to prevent. I know that we require certification, for example, because I have had to provide it myself in such circumstances.

I cannot guarantee that those systems are fool proof if somebody really wants to do this, but it is an example of where our systems have to be robust when people are seeking to silence other people. If somebody wants to silence somebody, that might be a way in which they try to do it. Our systems have to be robust to stop people using them not just to share bad content but to try to silence others. Where someone tries to do that, we have to make sure that our systems are robust for that.

The Chair: Just one other point before Baroness Kidron asks her question. The issue of false information was raised in our session yesterday and it is important that you get a chance to say something about it. UNICEF has had a serious problem with false information about COVID-19 being planted in its name and attempts to use its logo. What do you do about that, and what do you do if it has not been taken down?

Karim Palant: As I understand it, this was over the weekend in Myanmar.

The Chair: It was last week. I will give you the whole file.

Karim Palant: I ran an internal investigation when I heard you mention it yesterday, and I believe that quite a lot of content has been taken down; it might not be all of it. I am not sure whether there is content that you are aware of and we are not. I contacted colleagues internally and my understanding is that we have removed a significant amount of content in that space.

Clearly if people are pretending to be a health agency and sharing misinformation about healthcare in this space, especially if the agency has been in contact with us about it — as UNICEF has — we seek to act and investigate that very quickly. I know that we have done that in this case. As you say, there may be content that we have not been able to remove at this stage, and I will investigate that further if you can provide that information.

Q302 **Baroness Kidron:** I will just ask a quick question on a point of information and then ask a broader question. You just said that Facebook does not really have any way of enforcing anything or chasing after someone who does wrong, except to prevent them using your tools. What is Facebook's policy on people who misuse your tools? Do you block them for a minute, a month, a lifetime? What is your gradation of punishment?

Karim Palant: As you would expect, it varies based on the type of violation of the policy. Some violations are very high-end harms, such as terrorism content, so even if someone has been engaged in terrorist activity nowhere near our platform and has simply done it offline—

Baroness Kidron: I will ask you the question at the level that Lord Puttnam just outlined. If I said that someone was dead who was not or posted something in UNICEF's name, what would you do to me as a user? The terrorism thing we all understand, and Facebook has a half-way decent record on it.

Karim Palant: If you are talking about something like misleading information which, as in the UNICEF case, might spread real-world harm, there are circumstances in which someone might do that innocently — they might have seen a piece of content which they found shocking and shared it. We give some latitude where individuals may have done something unintentionally. But if somebody is found to do so repeatedly in a space, their account will be removed and blocked entirely.

Baroness Kidron: For how long?

Karim Palant: In some cases, completely, forever. In some cases, you will be prevented from posting for a month. In some cases, you might not be able to post for longer. We might suspend your account. In many cases — for example, where people are spreading misinformation — one thing we can do is look at the account to judge how authentic we think it

is, and we can remove it if we think it is inauthentic, or we can ask that person to prove their identity and they cannot post until they do. We might give them, say, 28 days to prove their identity, and we will remove them if they do not — that kind of thing. There is a whole gradation, depending on the type of harm.

Baroness Kidron: Is that public and transparent so that people understand misdemeanour and enforcement; there is a sense in which you know what the game is?

Karim Palant: How it works is absolutely public and transparent. There are really good reasons why we might not publish the exact numbers of violations you are allowed to have for each type of offence. Should that, and could that, be scrutinised in a conversation with a regulator, for example? I think that would be much easier.

Baroness Kidron: To go to my other point, something disturbed me when you talked earlier about the online harms and the regulator somehow having to articulate what the harms are. You used the phrase — I do not want to misquote you — "before we are in crisis". Do you not feel — I want a very personal view, not a Facebook view — that there is some sort of crisis already: the crisis of misinformation, the crisis of trust, the crisis of child sexual abuse imagery on your platform? That is not mentioned in your written evidence.

Do you not feel personally that we are in this crisis and that the first step is admitting that you are part of this crisis?

Karim Palant: Yes. I would zero in a bit more. I am talking specifically about a short-term thing like the current COVID-19 crisis. In that scenario, if there were a major news cycle about or parliamentary scrutiny of a particular piece of content, it is important that the regulator, more than us, can say, "Here is the judgment call we have made over time", recognising absolutely that there are very serious issues that need to be focused on, even if the media or Parliament are not focusing on it right now. They are huge issues, we need to focus on them, and that is the day-to-day work that you would have with the regulator.

Were there in a crisis to be a moment of great scrutiny for that regulator, which is what I was referring to, the regulator needs to be able to point to something and say, "Here is a framework within which we prioritise working with these platforms on these five issues and not this sixth", for example. That, I think, would give comfort to a regulator in that scenario. For example, it might be — this is only theoretical — that a regulator would have sat down with us at the beginning of 2019 and said, "Over 2019, I want you to deal with child sexual abuse material. I want to talk to you about that and about—" and list maybe 10 issues that might not include health misinformation. If, in the current climate—

Baroness Kidron: But would you not accept that fundamentally this is about veracity, spread, business model and so on, and that all these are manifestations of similar sets of systemic issues? A regulator that was

completely on top of one issue would probably be able to respond rather quickly to the others. The fact that you can list all these things is part of the problem we are tackling here.

Karim Palant: Absolutely. I totally agree with you. I am suggesting that if a regulator cannot show a paper trail of having interrogated us on an issue that becomes news, at the very least they need to say, "Here is the framework by which we are operating. Here is our prioritisation framework. Here is how we think about these issues". That was the only comment I was making. Whether the regulator drafts that or Parliament does so is really important, but I was making quite a narrow point.

Q303 **Lord Lucas:** What can the regulator, or politics and the Government more generally, do to help you make these sorts of decisions? It is noticeable that some very hard things are being asked of the likes of you and Twitter, such as deciding when a political argument goes over the border somewhat. Some things seem to be allowed to get extremely hard-headed, and then people get turfed off Twitter for saying that a woman is an adult human female. These seem to me to be really hard things to ask of you. What can we do to make such questions easier for you?

Karim Palant: I do not think there will ever be a point when it is easy for platforms, Governments or a regulator. It will always be complicated, but there are some quite straightforward things that could move forward quite quickly and tackle some stuff quite effectively. If we look at the current issue with COVID-19, for example, there is a globally recognised centre, the World Health Organization, for accurate information that we can all rely on, and that makes the dilemma a lot more straightforward. When anybody searches, anywhere in the world, for "coronavirus" on Facebook, they get sent to the World Health Organization. We can make sure that happens; that is straightforward. With something more complicated, where truth may be more disputed, it is obviously going to be much more challenging.

There are loads of areas where the Government could require or mandate more clarity where they currently do not. A really good example—we referenced it in the submission and we have certainly said it before—is transparency around political activity and political advertising. If the Government were to make that a legal requirement online, it would not just be us building the tools for transparency and enforcing them ourselves; there would be real-world sanctions for people who broke those policies. It would make a massive difference in terms of the deterrent effect of those rules. As I said before, we can remove people from our platform, but if they are not breaking any law, then that is it.

Another example of something that is very hard at the moment is around court proceedings. If somebody contacts us and says, "This is subject to court proceedings. This person has posted something that is in contempt of court", it is incredibly difficult to find out whether that is the case or not. There is no central place you can go and say, "We have been told there are court proceedings here and this might be in contempt of court".

At the moment, that is very difficult. We have a good relationship with the MoJ, which will bring cases to us and give us full chapter and verse, but that can never operate to scale; it can only ever be small numbers of cases.

Those kinds of things can be repeated across the board — for example, having conversations with the Food Standards Agency so that we know when somebody is registered with it and we can easily check that. That is something we are talking with it about. Across the board, government agencies can work more quickly and better with us to give us clear and unambiguous information that allows us to enforce policies that will include, in many cases, a requirement to abide by local laws and regulations. Therefore, we will be able, even under our terms of service, without any kind of legal proceedings, to say, “This agency tells us that this is in breach of local law, local regulations”, and we can remove it under our policies.

Q304 **Baroness McGregor-Smith:** Moving now to how we can make social media platforms a bit more understandable to the public, I am interested in how your newsfeeds and your privacy settings work. I am a member of Facebook and I look back on how privacy settings have changed over the years. How do you make that more understandable? Also, what tests have you done to make sure that the “Why am I seeing this?” button becomes more understandable to users? I have to say that I am still not sure that I quite get it.

Karim Palant: Those are very fair questions. One of the real challenges here is that there are lots of different ways in which people want to control their experience, and therefore you have to give lots of different settings. That presents a challenge in itself, because most of the time when people are online, they are perfectly happy with their settings, they are just going about their business, trying to post things and so on, and they do not want the user experience to be really unpleasant. We have all had the frustration of going to a new website, especially on our mobile, and having a cookie notification pop up so we cannot read what we want to read. That is a real dilemma, and it is a design dilemma as much as a regulatory dilemma as a dilemma for users.

Baroness McGregor-Smith: Thinking about that, although it is a big dilemma, when you have some quite young audiences on Facebook and you know the harm that can be done if their privacy settings are wrong, surely there are some very simple things that can be done. These are the real basics: “Look, whoever you are, this is what you need to know about your privacy settings. This is how we keep you safe”. When I look at it I go, “Yes, it is complicated”, but if somebody told me, very simply, what could harm me, I would be more inclined to change my privacy settings.

Karim Palant: Sure. There are a few things there. One is about under-18s. If an under-18 is about to accept a friend request, they will get a notification. When they set up an account first of all, their settings are “Friends only” and they have to opt to make that broader. If you are under 18 and you accept a friend request, it pops up and says, “If you

accept this friend request, this person will be able to see your stuff". You can do those kinds of things as you go about the process. The age-appropriate design code, for example, is a good trigger for a lot of companies, ourselves included, to have that conversation again and think about what more we can do in that space.

A lot of thought goes into that. For example, on Data Privacy Day this year — I think it was 6 February, but I cannot remember — we showed a privacy check-up, a simplified version of the privacy settings that requires everybody to go in and say, "Right, what are the top privacy settings that people use most, and how can I look at that?" We put that at the top of 2 billion newsfeeds and made sure that everybody saw it.

The "Why am I seeing this?" message in the newsfeed serves a number of purposes. One is to satisfy curiosity: people want to know why they have been shown a piece of content and want to understand their experience. It is also about transparency for interrogation by regulators, journalists and others who want to understand our systems better. It is a balance between giving all the information we possibly could and giving comprehensible information in a quantity that people will want to absorb at that moment.

For example, when you click "Why am I seeing this?", it will tell you, "You tend to like video; you tend to like content from your friends and family, so here's why we have given you it". You might say, "Well, I want more granular data that allows me to query that". That is also important, but if you are 15 years old and do not understand how a newsfeed works, is that the right moment to provide you with that huge subset of data and information? It may be or it may not be. Those are real dilemmas.

Baroness McGregor-Smith: I would challenge that slightly. If you want it, you should be able to get it, whoever you are and at whatever level.

Karim Palant: We provide that on News Feed FYI, where we set out in detail how newsfeed works and talk about the algorithm changes that we make, why we make them and what their impact might be for certain types of page. We do that separately at a macro level. In the "Download Your Information" tool, we will also provide you with all the information that we have about you that has gone into informing that decision. That is a huge amount of information if you have been on the platform for a long time and it might not be what you want to see at that moment you ask that question.

That is always the trade-off in whether we can give someone comprehensible, valuable information. We know that 7 million people in the UK have used the "Why am I seeing this?" tool, so there is clearly interest in it.

Baroness McGregor-Smith: Should you not be promoting that and explaining to everybody that they ought to do it, so they understand it? Every three months someone goes on Facebook, should you not say to them, "We are Facebook. Please look at your privacy settings. This really

matters to us”?

Karim Palant: As I say, last month we promoted it to everybody. We do that at regular intervals. We are always working with partners, NGOs and so on to think about what more we can do.

Another big area here is conversations off platform with school groups and NGO partners, where we can be as open with them as we can be and support their work. We have a big programme brewing at the moment — whose launch I do not want to pre-empt — with a number of parent groups, school groups and teen groups in the UK. We have a long-standing partnership with the Diana Award, Childnet and other NGOs that has reached thousands of schools and hundreds of thousands of school kids with information about what they can do and how they control their experience. A lot of the time, they do not want to hear from parents, from teachers or from us; they want to hear from their peers. If we can make sure that what they hear from their peers is good advice about controlling their experience, that is important.

Baroness McGregor-Smith: That is really good, but my concern is that, when you are on Facebook, you really need to know. Yes, you can train and teach parents and everybody else, and that is incredibly important, but it is about training people to understand what they do with the tool when they are on it. I am still not sure that enough people get just how easy it is for those privacy settings to change. That is going to be an ongoing battle.

Karim Palant: You want them to be easy to change so that people can control them, but you also do not want people to do something accidentally.

Baroness McGregor-Smith: Precisely.

Karim Palant: That is always a tension. It is something we work incredibly hard on, and on which we always take feedback from NGOs and others. We continue to iterate on that.

Q305 **Lord Scriven:** How much control do I have as an individual user in your algorithm to determine what I see?

Karim Palant: You can choose what type of newsfeed you want. You could leave it up to us, for example, and say, “Well, I’ve got 300 friends and I’m a member of up to 50 groups. I could therefore see over a thousand pieces of content in a day. I’m never going to be able to see all that. Based on my prior engagement with the platform and what I’ve been interested in, I’m going to let your algorithm choose in which order I see that content”. Just to be clear: you have chosen all that content; the question is the order in which it goes. Apart from with paid advertising, where it is really clear that that has happened, we do not show you anything that you have not chosen. If your friends have shared something into your feed, it is because they are your friends. Other than that, you have chosen everything.

Even with ordering, you can say, "No, I want to see these 10 friends or these five pages first". You can click on any individual profile or page and say, "I want to see that first". You can also change the feed so that it is ordered only in a certain way.

Lord Scriven: What percentage of your users understand how your platform works and how do you check it? Do you do an audit? What systems are in place?

Karim Palant: We are constantly testing our tools to make sure that people are using them. It is systematic to how we build the product.

Lord Scriven: I was very specific: I can use a tool, but I might not understand it. How does Facebook check that people understand the systems it puts in place to keep them safe and to determine the order in which information appears? Do you have any systems in place to check that people understand the options and the implications, or is it just, "These are the tools that you can use"?

Karim Palant: When we roll out a new tool such as "See First", we will test it in a few countries and see whether it is something that people use.

Lord Scriven: Can I be clear? I am not asking whether people use it; I am asking whether people understand the implications of use. What systems does Facebook have to test that people are not just using a tool but understand the implications of the choices they make on your platform for what they see and how they see it?

Karim Palant: Obviously, we have extensive conversations with users and consumers all the time. We are also in constant contact with a wide range of NGO partners around the world who work with young people and parents and give us constant feedback on the products. Ultimately, the biggest test of whether people understand the tools and feel in control of their experience on Facebook is whether they are using them.

Lord Scriven: So are there no management tools in place at Facebook level to check systematically that what you are giving to the user is understood? It is quite an important issue. Somebody said to us that you can only be literate with something that is readable. How do you check that the system is readable and understandable so that people make informed choices on your platform that mean that they do not get information that they do not seek, and in an order that they do not seek? Is there a management system in place which can be used by anybody in Facebook to make changes on the platform related not just to use but the implications of use?

Karim Palant: Some of what you are talking about is divining the intent of users and how they feel when they are using these products. That is best done through research off platform with users of the platform to understand how they go about it. Did they feel in control of that experience? Did they understand the information they were getting and how it came about? Of course, we do a huge amount of that. That is part

of being a business that wants to understand its users and to provide products that they use. We also seek out a huge amount of feedback on those products from NGOs and groups that work with user groups, particularly underrepresented groups or younger groups.

That is an important part of what we do. When it comes to the extent to which people understand literally what this button or that button does, you know that partly by whether they are using it, but that does not get to the point you are talking about, which is the extent to which—

Lord Scriven: I think the answer is no, from what you have said. I have another question. You talk about me having control over the algorithm. In May 2018 Facebook did a survey, four or five questions about who I engage with and what is my favourite way of looking at things. That fed your algorithm. At that point, in May 2018, it was a revenue-driven thing — that is what your algorithm is about, to be quite honest. Do I as an individual know, when I am filling in that survey, that it is about changing your algorithm, about how I will potentially see things if I have not changed things on my timeline? Are users given that information? The elephant in the room here is that the algorithm is driven by revenue, not by user content. That is the whole reason Facebook had the algorithm, and user content is back, so when you are asking me to fill in surveys to change the algorithm, what knowledge do you give me as a user that that is why you are seeking my information? Or is it not given?

Karim Palant: I am not familiar with the survey you are talking about—

Lord Scriven: You do a number of surveys. I used the May 2018 one because it was a big survey particularly around the algorithm.

Karim Palant: What we did in May 2018, I believe, is ask many questions of our users about how they wanted their data to be used on our platform, as part of the rollout of GDPR. I think most companies did that. I am not familiar with another survey at that time, but that will have had an influence on what data we were able to use to prioritise content within your newsfeed. That was about giving users control over that information. I am not familiar with the survey you described, but clearly, if we asked for information from our users, we would do so in a way that was clear and transparent and conformed with GDPR and any relevant regulations. Therefore, we would have to be very clear with you about what that information was being used for.

On the question of why we have an algorithm, why we prioritise things on your newsfeed, and what we tell people about it, the first thing is that you, as a Facebook user, could potentially be served well over 1,000 pieces of content on a given day because you have, say, 300 Facebook friends, all of whom might have posted, you have pages you like and groups that you are a member of. That content has to be prioritised in one way or another. It is always a choice: there is no non-choice.

Lord Scriven: You have just admitted to the Committee that Facebook has no way of showing whether people are making informed choices

about what they are seeing, because there is no management tool that gives you information to show that people understand what is happening on the back of this algorithm and the individual choices they are making.

Karim Palant: To be clear, nobody is seeing any content on the platform that they are not choosing to see. If you like a page, you see content from that page. If you are friends with people, you see the content from those friends. If you are a member of a group, you see the content from that group. That is how you see content on Facebook. I think that is pretty clear to the vast majority of users.

Lord Scriven: The bottom is that the algorithm is the main way in which Facebook is able to determine revenue on the back of what is posted where and what is advertised where: is that correct? It is a revenue-driven algorithm, rather than predominantly a user-driven algorithm?

Karim Palant: I disagree. For example, we looked at all the published evidence that existed at the time, about 18 months ago, about the wellbeing of people who use social media. We published an analysis, a literature review of that, in which we looked at the extent to which engagement on social media drove broader well-being. In that scenario, we looked at something called passive scrolling versus something called meaningful social interaction — interaction with your friends and family on issues of shared interest.

We found very clearly that that kind of interaction was positively correlated with well-being, whereas passive scrolling, looking at public video content, was correlated with more negative well-being. That drove a change in the algorithm which was hugely controversial with many of our commercial partners. It was hugely controversial because it drove a decline of 50 million hours a day of Facebook engagement and activity and we did it driven by wellbeing. So, I disagree slightly with the revenue-driven aspect.

We have to prioritise somehow. We could have it as purely chronological — see the latest thing — but that might mean you might miss something from one of your close relatives that you are particularly interested in, versus something from someone who you knew 20 or 30 years ago who is posting about something that is not directly relevant to you. That is a choice we have made about how we build the product: it tends to be pretty much what users want.

Lord Scriven: Finally, in a democracy, why does Facebook not allow me to see things that I might not like? Why is the algorithm driven by likes? In a democracy, during an election, would it not be good to have an algorithm that occasionally showed me things that I don't like so that I can get a different side of the argument? Why does Facebook not do that? Why would that not be a good model to integrate into your algorithm?

Karim Palant: About four per cent of what is in the newsfeed is news, generally speaking. For news content, we will often show other related

stories, which will often be from a wide range of different viewpoints. If you have a friend who might share into your newsfeed content that you disagree with, we might also share content below that that you might agree with. As I think the Reuters Institute told a different Lords committee this week, the evidence is that with social media, people tend to see a much broader range of news content than people who do not use social media.

The Chair: What must be clear to you, Karim, is that there is deep disquiet, not just in this Committee, about the decision-making process of algorithms, in that they are not wholly transparent and that they are capable of causing confusion, to put it kindly. I am assuming that you, Nick Clegg and others are the good guys. Is it possible, particularly in the present circumstances, that a whole series of decisions have to be made to clean things up, improve and increase transparency and at least try to get rid of a lot of the disquiet that Lord Scriven and Lady McGregor-Smith have just been describing? There is something wrong, is there not? That is the problem. You can feel it in this room. There is something wrong.

Karim Palant: I have been in this space for about four and a half years. I think there was a time when it did not have a very high profile in the public debate. I have seen that escalate very sharply in the last few years. It is clearly the case that there is a good deal of disquiet about how the ways in which we communicate digitally have fundamentally changed the way we all communicate in a range of ways. Many of those have been hugely positive, but some have caused a level of disquiet and confusion. There are clearly very real-world harms that can come from that. I completely agree with you on that.

The Chair: Lord Harris has been unbelievably patient.

Lord Harris of Haringey: Ultimately, of course, it is reputational: if your reputation tanks because of how this operates, that hits you commercially, whatever the original motivation of the algorithm. I just pick up one point and I may have misheard you; I think you said that 7 million people in the UK had used the "Why am I seeing this?" feature. Just remind me: what is the number of Facebook users in the UK?

Karim Palant: I believe that 32 million use it every day.

Lord Harris of Haringey: Right, so less than a quarter. How many of the 7 million people who have used the "Why am I seeing this?" feature have used it more than once?

Karim Palant: I do not know. I will find out.

Q306 **Baroness Kidron:** In the interests of time, could I ask you to give short answers on this? Some of it is just for the record. First, would you tell the Committee what the role of human moderation is in improving the online experience?

Karim Palant: Absolutely. Human moderators look at user reports that have been flagged to us. Increasingly, they also review content surfaced

to them by algorithms that scan content for violations of our policies, in the large part. For example, when we say that 80 per cent of the hate speech content that we removed from our platform in the first quarter of this year — that is 3.5 million pieces of content — was removed proactively by our tools and processes, what that largely means is that our algorithm and AI tools spotted that something may have been hate speech and surfaced it to a human reviewer who made the final decision. That is a large part of what they do. The third part of what they do is sampling content so that we can publish our transparency report about the prevalence of certain types of violations. That is the role that they largely play. We have 35,000 people working on safety and security at Facebook; about half of those are content moderators.

Baroness Kidron: You will not be unaware that there are some concerns about their consistency, the transparency of that process and whether there is sensitivity in different markets. What do you think the steps could be to make that transparency not pretend but absolutely real and trusted by Facebook users and perhaps the regulators that look at these things?

Karim Palant: I joined Facebook four and a half years ago and the system then was an awful lot less transparent. I take no credit for the change, by the way; I have not worked very closely on those things, so I am making it sound as though I have had more of an impact there than I have. Four and a half years ago, it was very opaque externally, for a whole range of reasons. There was a genuine concern that, had we talked about the use of AI and algorithms to spot and surface content, people might jump to saying, "Why don't you use them for everything?". These were very nascent tools that were being developed at that time and there was a fear that it might look like there was a magic button that could solve the problem. There were lots of concerns to do with that.

There are a few areas in which we are now a lot more transparent than we were and a few areas that we continue to work on. The rules by which people are judged when a piece of content is sent to our reviewers used to be very high-level: "You must not post hate speech", that kind of thing. About 18 months or two years ago, we published the detailed guidance that we give to the reviewers. We basically said, "Here is the detail of the policy. Here is how we have written it and why we have written it in this way. Here is the rationale by which we came to it". It gave a much more detailed breakdown of what we consider to be hate speech and why we came up with it in that way. That was step one.

Step two was being much more transparent about the number of people working on this, where they are based and so on. The third is the outputs of that process. A couple of years ago, we published the first community standards enforcement report, which was our first attempt to publish information on the scale at which we are removing content, the amount that is removed because we proactively found it using our tools and the amount that was flagged in user reports. We do something called a prevalence score, which we think is really important. It is the best

measure of the scale of the problem, the backstop measure of something such as nudity, spam or child nudity.

To give a very quick explanation, the score essentially gives a percentage based on a sample of what people would see if they opened Facebook right now, to two to three decimal points. It asks: if 10,000 people opened Facebook right now, what would they see? How much of that would be bad content of particular types? It is based on a very large sample size, so it allows us to say with very high accuracy that 0.014 per cent of the content you might see would be hate speech. In fact, I do not think we have a metric for hate speech yet because it is very hard to calculate, but the score offers a level of transparency. We publish that every six months and are broadening the types of content that might be included. For example, last time, we added Instagram content to that measure and broke that out from the main measure. We also added self-harm content to those measures. Every time, we are building on what we can do there and doing more and more.

There is that level of transparency. There is also an assurance that the people who are doing that content moderation are looked after. You may have seen the announcement made overnight that we are asking all the offices to close if they can and people to work from home during the current crisis, but we will continue paying all those contractors and will make sure that we build systems to ensure that we can continue to do as much of the moderation as possible during the crisis. We have also arranged for a large number of journalists, policymakers and so on to visit the sites where this reviewing takes place.

Baroness Kidron: One idea that has been presented is that of putting forward a database of archetypes and decisions and making that available to the public. I suppose that would be a positive double whammy in the sense that it would make people understand what you are looking for and what they are not to do. It would also give a little more clarity about the sort of problems you face and the decisions you have made about them. What do you think about that idea?

Karim Palant: It is important to be clear about who the audience for some of this is. As you know, there are already databases of known images of child sexual abuse, for example. The audience for that, for want of a better word, is other companies, law enforcement agencies and so on. They need to share access to that so that we can scan all images that are uploaded against the database and remove those child-abuse images. A similar system is in place for terrorism material and companies are working together quite closely on COVID-19 misinformation, so that is shared across industry.

When this is external and about educating users, I wonder whether a heavy-lift database is the right way for people to engage with that content. Could we be better at explaining and publicising our policies and finding ways to make them real for people? Yes, that is certainly true. We are always trying to work with NGO partners, such as hate speech partners, in areas in which we think this could be most effective, to

educate both the people who might be victims of that kind of hate speech on their rights and how they might contact us to raise these issues, and people who might be perpetrators, so that they understand what they can and cannot do on our platform.

One of the areas we are thinking harder about now is how, when somebody has breached one of our policies once, we can give them more information to explain how they breached our policies and how they can avoid doing so again. That is certainly an important thing we can do.

Baroness Kidron: Bearing in mind that this is on the record, how confident are you about the penalties, that you are consistent in the way you behave to the perpetrators and that you use what you have available?

Karim Palant: First, I am personally not that close to some of this. Secondly, it is not something that we would necessarily disclose in a public forum because we want to chase down people who might wish to abuse our platforms.

One big shift that is under way at the moment is an additional shift towards looking at the behaviour of individual actors and profiles, which might or might not be authentic—we have done a huge amount of work on identifying authenticity. We are looking at authentic people who are routinely misusing to get a better understanding of how that kind of behaviour, as well as the associated content, can be identified. We have a system that effectively says, “If you have done this X number of times, your profile comes down”, but this behaviour could be quite granular and linear. We are thinking about internal signals that might tell us a bit about an individual and whether they are systematically behaving in a way that might not breach our thresholds but is causing harm, and whether we can be better at identifying the people who are doing that.

The straight answer to your question is, yes, I think we could improve. We could improve on all this, by the way — let us be clear.

Baroness Kidron: Why did you decide that the decisions of the oversight board should not create binding precedents? That seems like a useful higher authority.

Karim Palant: I think that it is a process that will grow over time. In the first instance, an example might be where some content has been removed and people can say, “Actually, I want to go and escalate this to this oversight board”. A system for exactly how that should be done is being released shortly. We will explain how one might do that. Who is on the board is being established. That content will go to it, and it will be able to say whether it was the wrong decision. For example, it could be that interpretation of the policies was incorrect in that instance. It could be that the policy has been drafted poorly and we could get some advice from the oversight board to change it. If we have been told that a decision was incorrect and needs to be reversed, the expectation is absolutely that we would act on that.

There are potentially technical barriers to operationalising the policy at that level, so I do not know exactly where the line has been drawn between the oversight board effectively making a rule that has to be applied to all similar content ever produced and it giving strong advice that is binding on us in terms of recrafting our policies and having to go away and work out what is technically feasible.

Baroness Kidron: Do you accept that the optics of having an oversight board to which you have not given the ultimate authority makes one feel that it is less oversight and more doing the bidding of the platform?

Karim Palant: It has ultimate authority over that piece of content.

Baroness Kidron: Yes, but, as Facebook always says, we are talking about a system with billions of pieces of content, so systemic solutions are obviously the ones that we are all reaching for.

The Chair: Let me interrupt for a second. There is really something crazy here. I have some 50 years' experience of working with the British Board of Film Classification. There is no question of needing to reinvent the wheel here. You have a scale issue, but you definitely do not have to reinvent the wheel. You defer to a body that is preferably thoroughly independent and then you abide by its decisions. As Facebook, you do not override them any more than I can make movies and say, "I'm very sorry, BBFC. We're putting it out anyway". This is quite normal stuff.

Here is the danger you have. It is the same danger that auditors have when a client is vastly bigger than them. The temptation to keep the client, as opposed to losing the business, is colossal. You are the employer of moderation centres; you have decided to employ separate companies. They are reliant on your income, so they are not going to upset you and offend you. You have the human problem of desensitisation. If you have watched 24 child rapes, you would not react to the 24th in the way you did to the first. These are things that you will have to grasp. There are rules. There is a way in which the world and human beings operate. What is coming across to me is that you are kind of light-touch. You want to do the minimum you can do to run your business in the way you run it, to not interfere too much but to be seen to be okay. You are not coming across as a business of really good guys who want to make this a better world.

Karim Palant: I slightly contest that.

The Chair: I would hope so.

Karim Palant: I am not sure where the suggestion of not having precedents has come from. It is not the case that we have said, "Oh, by the way, this will not be binding". It will be binding; it is about what it is binding on and to what extent it has the reach.

To use your film analogy, if you took a film to the BBFC and it said, "Cut that out", you would cut it out, but if you produced another film that had a similar scene in it, you would expect the BBFC to watch it again.

Therefore, it is not binding 100 per cent of the time; you would expect the BBFC to make the judgment call on the next film just as much as it did on the first.

Baroness Kidron: Just to be clear on this point, I believe there was a workshop of the oversight committee about how it was going to set out. It came to a decision that it would have authority over a single piece of information, but that it would not necessarily be binding throughout. It is just a question of optics. Ultimately, if you have billions of parts and you are not building up a culture of what is allowed, and the oversight body is not given some authority, that is problematic.

Perhaps I may finish by making a small suggestion and an urgent request. We spoke earlier about access to research, independence of research and blocking of research. That has been important in all the conversations that we have since had. In Lord Scriven's interrogation, in the question about moderation and in the question about what people understand, there has been an urging that we have some independent oversight that one can absolutely believe is independent — that is neither paid for nor monitored nor whose edge is created by Facebook. With great power comes great responsibility, so you need such oversight. In addition to the things that you offered to do earlier, please could you come back to us on the iterations with the regulators about research access and what they are doing to get in the way of that being a fruitful and transparent action.

Q307 **Lord Holmes of Richmond:** It is great still to be able to say, "good morning"; I thought that I would have to start my question with "good afternoon", but there we go.

I am going to start with a supplementary from an earlier conversation and then come to the main question. When you were discussing with Lord Puttnam the UNICEF issue and Myanmar, you said that you had launched an investigation and thought that most of the content had been taken down, but the Chair may know of content that Facebook does not. Does that not concern you?

Karim Palant: It is clearly a serious issue. When anything of that nature is flagged to us, we will undertake a deep investigation of the connections. For example, if it is activity that seeks to spread co-ordinated misinformation, we will undertake an investigation to look at that whole network and what we can remove. It is clearly a matter of significant concern.

There are two levels of investigation. When I heard what Lord Puttnam said to the Committee yesterday, I undertook an investigation to find out what we knew about this and what had been done. Secondly, there is the actual investigation into the content. My investigation was on a deadline of coming today to answer Lord Puttnam, but the internal investigation is not concluded and is ongoing.

We will never catch all content of a negative nature if people deliberately obfuscate or seek to get that content on our platform. I am confident in

this case that we have found and removed a significant amount of content. I am not at this stage able to say that we have removed everything that UNICEF raised with us, because I do not know all the reports that it might have brought to us. I know that we have removed the content that it flagged to us in our initial conversation, but I do not know whether we have removed everything, because I was not able to establish that in the time that I had yesterday evening.

Lord Holmes of Richmond: So, in spite of all the spend, all the human checkers, all the auto-checkers and all the resource that you put into it, Lord Puttnam is still better than all the systems you have.

Karim Palant: I do not know the answer to that question. Lord Puttnam says that the content is still online and that he has printouts of the content with him. I do not know whether it is still online or where we are up to.

Lord Holmes of Richmond: Okay. Why does Facebook think that content moderation decisions should be made with partial knowledge of the context and speaker? Does this not purposely hobble the moderation process? To what extent and level can moderators choose between removal of posts from the site and removal from the recommendation system of the newsfeed?

Karim Palant: There are a few key reasons why we restrict what content people might see when they are reviewing. First, this is personal information about users and, as I said, there are thousands of reviewers who have access to those reports. Therefore, it is important that we are privacy-focused in what we share with them. There have been instances in the past when bad content was not removed because reviewers were not able to see the appropriate content and we keep that under regular review. We will share more or less information depending on the kind of content and the reason the report or AI tool has surfaced it, to make sure that we are giving the reviewers the appropriate, privacy-safe context.

The other important thing to say is that our rules are designed and written to be enforced at scale based on millions of reports. If you want to avoid bias and make your decisions as consistent as possible, you need to look at the content itself and whether it violates a policy. That is clearly appropriate only at scale, for large-scale content violations. There will be instances in which that does not satisfy, so for particular group and issues, there is a different system — I am talking here about elections, for example.

In the UK, for example, we know that Members of Parliament feel very vulnerable and that they are subjected to abuse both on and offline, on a wide range of platforms, over the phone and by email. There may be context about a particular instance or report on our platform that we do not know about and which we could not show to our reviewers. In those cases, the way we get around the problem is by having direct points of contact with either the vulnerable groups themselves, as we do with MPs, or NGOs and organisations which work closely with those groups,

understand the issues better and can help us with additional context when we receive a high-touch review.

A Member of Parliament who reports some abuse on the platform may feel deeply upset about it, and that may be partly because they know that that user has been engaged in a wide range of off-platform activity, which is context that we do not know about. We may not act on that report using the traditional reporting tools, so we have made sure that every MP has direct email contact with us so that they can contact us and give us that additional context. Those are the kinds of things we try to do to get around this very real problem, but this is very hard to operationalise at scale. In any case, even if we could, we would not have all the information we would need for every report.

Lord Holmes of Richmond: And on my second question?

Karim Palant: Sorry, do you mind repeating it?

Lord Holmes of Richmond: How much are the moderators empowered to choose between removal of posts from the site or removal from the recommendation system of the newsfeed?

Karim Palant: When content moderators are reviewing content, they are deciding whether it violates our policies or not. If it violates our policies, it comes down completely or, in some cases, such as if it were graphic content, they might mark it as disturbing or for over-18s only. Those options are available to reviewers. At the moment, they cannot downrank content in the newsfeed.

If there are a significant number of reviews saying that content is misinformation or fake, it might get sent to fact-checkers who are independent of Facebook. They would have absolute discretion on whether to fact-check that content and, if they do, to declare it false. It would then be downranked in the newsfeed.

That is the case for content that does not absolutely violate our policies, but which has been marked as being false by a fact-checker. It can stay on the platform, but it will receive considerably less reach as a result and will be marked as false. However, if something violates our policies, it simply comes down; it does not get removed from the recommendation system.

The Chair: That is a good segue to the next question.

Q308 **Lord Harris of Haringey:** I want to move on to third-party fact-checking. Can you tell us what evidence you have of its effect, as it were?

Karim Palant: We know — some research has been done by independent academics — that if you put a fact-check marking on something and say that it is false, it can do a number of things. For example, it could make people who see stuff that does not have a false marking trust that content more, which might not be correct. We know, for example, that if you put a false marking on something, people might

be more inclined to be angry and disbelieve it. We know that, on its own, marking content does not do the job. We downrank very substantially, in terms of how much reach it gets, anything that has been marked as false, partly for that reason. If a fact-checker says something is false, it does not get circulated widely. We have estimated that this means an 80 per cent reduction in reach; it reaches about 80 per cent fewer people than it otherwise would.

Lord Harris of Haringey: Okay. Have you looked at this scenario? I have looked at some content quite early on, it has then been fact-checked and turns out to be rubbish, so it is marked as such. Have you looked at a mechanism for going back to the people who looked at that content at an early stage and saying, "You looked at this. It is now labelled as misinformation".?

Karim Palant: Yes, we have. There is lots of work going on internally about when and how that might be appropriate. One of the things we did during the UK general election was that, when somebody saw something that might be considered to be voter suppression information — telling people the wrong date of the election, for example, or telling people not to bother voting because it is all fixed anyway, or whatever — we went back to them with a notice giving the correct election information, telling them the date of the election and so on. We did that during the 2019 general election in the UK. It is something that has real potential and it could be an interesting way to move in the future, so we are working really hard on ways to do that.

Lord Harris of Haringey: Have you actually judged the effectiveness of that? Or are you simply saying that you tried it and it looked as though it was good?

Karim Palant: We tried it out. We obviously do not know how persuasive the initial misinformation was. Something that was shared a lot during the UK election campaign was a jokey meme suggesting that we should just tell the Tories that the date of the election was 13 December — the day after the actual date. Having worked in political campaigns, I know that that is a joke that you quite often hear in political committee rooms, and people were sharing it online. We removed that at scale, even though people in political party committee rooms might be a bit like, "Come on, that was just a joke". We removed it and people who had seen that content got a notification telling them the correct date of the election. Without knowing how effective the initial content was, it is hard to say whether that drove changes in behaviour or whether people were then more likely to know the correct date of the election. It would be very difficult to unpick, but it is certainly something that we look at closely.

Lord Harris of Haringey: You have told us that one of the possibilities is that, if you label something as misinformation, people assume that anything that is not labelled as misinformation is therefore good content. That then calls into question the process by which the algorithm refers something for fact-checking. What auditing of that algorithm have you done, and to what extent has that auditing been done independently?

Karim Palant: Part of what we hope Social Science One will do is audit that kind of thing independently. The dataset that is shared includes URL links to precisely the kind of news content that you could study. We have obviously done a significant amount of auditing on the quality and nature of the news that is shared in the newsfeed. We think that there has been a dramatic reduction in the amount of poor-quality, clickbait-type, misleading news that is shared. I saw the figure of a 50 per cent or so reduction since 2016, but I could not tell you precisely the metrics on which that is being judged.

One thing that is worth looking at is that the total amount of news in newsfeed fell, as a result of the meaningful social interaction changes I talked about earlier, from about five per cent to about four per cent of newsfeed; so there was a falling back of the total amount of news in newsfeed. However, the amount of traffic that went to broadly trusted or prominent publishers did not fall in the same way over the same period, so we can see that the share of the pie that is going to established news publishers has grown as a result of that kind of activity. That includes, for example, removing spam and clickbait — click farms and that sort of thing. We have developed a whole range of policies and tests to remove that.

Some of this is in the eye of the beholder; some of it will require independent analysis rather than analysis by us. I know the Reuters Institute gave some evidence to a different Lords committee a couple of weeks ago and talked a bit about the effectiveness of some of the actions we have taken and what it has seen, which I think is very instructive.

Lord Harris of Haringey: You are saying that the data you have handed over is sufficient for the effectiveness of your algorithm to be externally audited?

Karim Palant: It should be for that kind of external audit. It may be that that is not the only one; there may be a lot more to do. This is a very complicated area that is in the eye of the beholder. Essentially, someone has to judge whether something is a legitimate news source or not before you can measure anything. Is this fake? Is this accurate? You have to make that initial judgment, for a start, and then you can measure it. Of course, we are going to make one judgment, and someone else will make a different judgment: a wide range of different people will be looking at that content and making a judgment call as to how effective all this has been. The purpose of the transparency and the handover of data is to give that external assurance. As I say, we have seen some really encouraging numbers, but it will be us marking our own homework, which is not good enough.

Lord Harris of Haringey: Okay. Quite properly, you are saying that somebody else should look at this who will also mark the work, but in the meantime, have you, for example, released the data, which you presumably have, on how much flagged content was then shared, and how many people saw the fact checks?

Karim Palant: Some of that data will be in the dataset that we have shared. We know at a macro level, for example, that it gets 80 per cent less reach, on average, when it is fact-checked.

Lord Harris of Haringey: You are now saying this publicly. Have you said it publicly before?

Karim Palant: Yes.

Lord Harris of Haringey: Okay, that is fine: I suspect most people would like to know the answers to some of those questions without having to wade through squillions of megabytes.

Karim Palant: Absolutely. To be blunt, I think it is perfectly understandable that people do not take that number and run with it, because it is us saying it and why would we not say it? It is important to have that independent scrutiny and audit, but as I said the Reuters Institute has been really effective in publishing some numbers: I think it showed a very substantial reduction in the amount of misleading or fake news during the 2019 election. It gave evidence on that to the Lords committee inquiry into the future of journalism, and I recommend looking at that evidence. It told that committee that there is a dearth of independent research in this space. It has tried very hard to do this work and I think it is quite useful.

Lord Harris of Haringey: I have one final question. In your written evidence you told us that content was referred to third-party fact-checkers through a combination of technology and human review. Can you tell us what the proportions are?

Karim Palant: I will look into that and get back to you.

The Chair: Baroness McGregor-Smith has a quick question to which a yes/no answer might suffice.

Baroness McGregor-Smith: It is just a reflection on the challenges Facebook has on privacy and how far you can protect the individual. Has Facebook ever thought of turning the model on its head and saying, "We are going to charge our users for accessing Facebook, ditch all advertising and keep their privacy completely whole"? Is there not a danger in continuing to say, "We are really good, we are amazing, we know how to moderate"? If I buy a bottle of water, I know what I am buying. Would you not then have a better contract with the public, ultimately? Have you guys ever thought about it?

Karim Palant: It has definitely been thought about. I know our chief executive was asked about this numerous times in front of Congress and I refer you to his answers rather than mine. I am conscious that he is probably closer to those discussions than I am.

Baroness McGregor-Smith: What are your reflections on it?

Karim Palant: I know that he has said, and it is a very valid point, that lots of media, lots of services and lots of tools are funded by advertising and they provide a real service for people.

On the question whether people understand and feel comfortable with that trade-off — "I get this product, I do not have to pay for it, I may be on a low income and not be able to pay for it, I get this product for free and in return I am shown advertising" — I think that a lot of people are very happy with that trade-off. In many cases, they are quite happy to have advertising, especially if it is well targeted. They do not like irrelevant advertising but they quite like relevant advertising, so I do not think most people see it as too big a cost or a burden, but I totally accept that it is something that people have raised before and I refer you to my boss's answer.

Q309 **Lord Lipsey:** Until recently, I was deputy chair of Full Fact, your main fact-checking partner in this country, and we very much welcome that partnership. What I cannot for the life of me understand is why you will not fact-check politicians' utterances. Full Fact has a wealth of experience in doing this and it is very rarely controversial with the politicians involved, even when they have been very naughty. Your steering clear of this whole field seems to leave a gaping hole in your defences.

Karim Palant: This is a really tough one. You are right to say that fact-checkers have routinely fact-checked politicians in the abstract; they have looked at politicians and said, "That is false", but they have then generally posted that on their own website: "I am Full Fact and I am saying that what this person said does not stack up. We have looked at the evidence" and so on. Equally, when a politician posts content themselves, in their own voice and in their own space — in an advert or on their own page — and says, "This is what I think, this is what I believe", to mark that in some way as false, to downrank it or put a screen over it and say, "You cannot see this until you have clicked to say you acknowledge that this has been fact-checked", would be quite an extreme measure.

To be very clear, we deal with all content that is not directly from politicians or political parties. One thing that Full Fact and other fact-checkers asked for was a very clear definition of who was in and who was out, which is important. However, when it comes to a political party or a candidate in an election who is representing a political party — that narrow group of people who are ultimately accountable for what they say to the media, to the electorate and to their competitors in the election — do we want to get involved and say, "We are the referee. That is right and that is wrong" in political debate?

We do that for every other piece of content on the platform, so if a news story were to make a false claim about a politician, we would. However, the judgment call is that when you have politicians debating the issue during an election, for example, putting that fact check front and centre on their content is not the right approach. That is partly to do with the effect of doing so. If you mark every politician's piece of content, if you

down-rank it every time they post this claim, that combination of things that says, "We are going to intervene in this debate in the following way and say 'that is false'", goes beyond anything that has ever been done in political debate before.

It is not done in media interviews. At least, it is done by the journalist but in a questioning way; there is no big pop-up that says, "This politician has done this, that or the other." It is not done on posters and it is just not the way we have traditionally done political debate in this country. In fact, we do not have a referee; we have Full Fact, which is very impartial and a valuable partner, and Reuters, which we also partner with.

However, ultimately, we do not generally have a referee in political debate in the UK. I know that you have taken evidence from the Advertising Standards Authority and others which suggests that maybe we should; that is a matter for the UK to decide. If there were a referee, that would be a different question, but there is not.

Baroness Morris of Yardley: We are not talking about what politicians think and believe and trying to take that down, which is what you said. We are talking about when they give a straightforward factual inaccuracy or tell a lie. That is a much more limited thing. We do it for every other provider of information, for example if the BBC told a lie or perpetrated a lie made by a politician. Why do you stand out? Even Twitter does this now.

Karim Palant: There is nothing to stop Full Fact saying, "Politician X posted this on Facebook and it is false", just as there is nothing to stop it saying that what a politician said on the BBC is false. It is perfectly within its rights to do that. What we are saying is that we will not put that into the tool so that, when that pops up on Facebook, that politician gets a flashing message saying that the claim they have made is false. That would mean us overlaying on what the politician says. They might be the Prime Minister or running to be Prime Minister or a local MP, and we would be overlaying something on top of what they said. Before the electorate could see what they had said, we would be telling them that what the politician had said was false. I think that is quite extreme.

Baroness Morris of Yardley: Other people seem to do it. You talk about overlaying, but why should they have first bash at telling a complete untruth without you taking the opportunity that your fact-checking network provides to see what somebody objective says about it? This comes from America, does it not? An American free speech philosophy is permeating what happens in this country, to our disadvantage.

Karim Palant: I would engage in the thought experiment about how it would play out in reality and I am not sure that I agree with you. It is clearly controversial and there are clearly people who disagree with that judgment call. In reality, if you imagine that a major name in politics posts something on their Facebook page and very few people see it, but when they do see it, there is a message saying, "Full Fact says this is false", and that is the main way that that politician usually

communicates, it is very different to a person doing an interview and somebody else subsequently writing that what that person said was not true. As you know, we work a great deal with Full Fact to share those fact checks and ensure that lots of people see them.

By the way, if a politician shares a news story that has been debunked, they get the same treatment as everybody else. They cannot share debunked news stories and get away with it. Equally, if they said anything that violated the community standards — hate speech, incitement of violence or misinformation that led to, for example, health issues during the current COVID-19 crisis — that would not be fact-checked but removed.

It is a balance. When somebody says, "This will cost this much", and another politician says, "Actually, no, it won't cost that much, it will cost this much", and then Full Fact says, "This person is right, this person is wrong", should a big message saying that the information is false flash up every time the person who was wrong makes that false claim? That is not how we have done political debate in this country before. It would be a big change.

Q310 **Baroness Kidron:** In your own evidence, you say that dealing with false news and misinformation is a top priority for Facebook and talk about the huge importance of Facebook in political engagement. You cannot say, "We haven't done it before", because you are describing a new world order in which you target people, spread information and so on. Surely the essay question is not about whether we have done this before and whether it will be ridiculously heavy-handed.

In these new circumstances, when 85 per cent of UK MPs have a Facebook profile and are using that tool to reach the electorate, I agree with you that they have a greater responsibility to look after the information they share. However, when they step over a line on something factual — as Baroness Morris said, we are not talking about what they believe — it is surely irresponsible for Facebook to spread that misinformation. I take your point that the example of politicians disagreeing about the cost of something is somewhat different from the absolute, outright misinformation that some politicians have given.

I am concerned about the democratic process when it is the person who goes closest to the line — I was really interested to hear you use that phrase — and is the edge case that gets the most benefit from your system, in which, as we all know, content spreads faster the more outrageous it is. There is the danger — some people would say it is already here, but let us say that it is not here yet — of a system being created in which you deliberately allow the most extreme voice to get the biggest reach, with no check whatever. That is also a democratic deficit.

Karim Palant: It is clearly not a clear-cut decision one way or another. There are clearly very good arguments on both sides. That was our decision. In reality, although it feels like there are two sides to the decision, in practice, if you did it, I think it would quite clearly feel worse

the other way. It is an empirical question. In reality, if a politician could not communicate online without prior fact-checking, things would be very different.

Baroness Kidron: I do not think anybody is suggesting prior fact-checking, to be clear. I also do not think — forgive me, Lord Lipsey, if I am speaking out of turn — that anyone is saying that this must be done by Full Fact. This may be a role for the Electoral Commission or the ASA. We have to be open about what the limit is and who finally holds the hand. You asserted that, in this new world, there is no room for truth-telling when the person speaking is a politician. I am not sure that I can accept that assertion.

Karim Palant: The assertion I would make is that, for as long as we have had democratic debate in the UK, this has been about accountability on what politicians say and transparency about who they are. The latter is a really important point and it is something we have really focused on. If you are speaking as a politician in the UK, especially through paid media, you have to tell us who you are, where you are and that you are UK-based, and you have to be transparent about which organisations you are speaking on behalf of, and so on. That transparency is hugely important, but the accountability comes from other politicians, journalists, civil society and so on.

As an example, if you wrote a very big number on the side of a bus, you are accountable for that number, just as much as if you said it in an interview or on Facebook. Clearly, there was scrutiny from civil society and the media during the referendum and other elections when big numbers were being thrown around. That scrutiny was real and legitimate. The outcome is that, ultimately, the voters decide.

Baroness Kidron: I will stop but forgive me for saying that that is a little old-fashioned. I work with a lot of children. Most of what they see about the news happens online; they are probably not looking at the television, free newspapers or newspapers behind a paywall to check for that accountability. Now that we have a system in which people get an increasing proportion of their information online, does not the online space have the same responsibilities that traditional media and broadcast had? That is my point. You cannot have this new system without some of the responsibilities of the old world.

Karim Palant: It is important not to imagine a golden age when everybody agreed that what the newspapers wrote was a perfectly fair representation—

Baroness Kidron: Absolutely. I would 100 per cent agree with you on that, for the record. I am making the point that you are saying that there are these other sources which ensure accountability but, increasingly, parts of the demographic are not accessing those and are seeing news only through the mediation of platforms.

Lord Harris of Haringey: You are essentially creating — or this has

been created, if you like — immunity for politicians from some of the processes which would apply to other people. Therefore, my question is: how do you define a politician?

Karim Palant: We would say it is somebody who is a candidate for election, an elected representative or a political party.

Lord Harris of Haringey: You said, “or a political party”. If you are a leading politician, you personally might be very careful about what you say, but there might be a large number of crazy — to use the technical term — members of your organisation to whom you give absolute licence to say the most bizarre things. Are they covered? They are representatives, because they are members of your party. They might even be elected parish councillors.

Karim Palant: They are not covered by virtue of being members. A member of a political party does not speak on behalf of that political party. I can think of a good many political parties right now—

As for an elected parish councillor, I would have to look into where the line gets drawn. The intention is not to create a big lacuna in the rules here. We are very clear about what we are trying to do, which is to say that political debate needs to be the primary way in which politicians are held to account for what they say, because that is what democracy tends to do. If, on the other hand, there are players in that debate who are not accountable for what they say, who are using online tools for the spreading of misinformation, then we want to make sure we have systems in place to stop that. When somebody is clearly a politician — we are clear about who they are — and they are seeking election or representing people in Parliament, for example, they are accountable for what they say.

Lord Harris of Haringey: It would be very helpful if you could write to us with your definition. For example, what level of elected politician do you have to be? Is it a local councillor? What if you are representing your party? If, for example, I am the secretary of the youth branch of the ward’s political party, there may not be many other members who are of the same age group, however I am the secretary of the youth branch and am therefore speaking; am I then a politician? What is your definition?

Karim Palant: I will triple check, but I am fairly certain that your latter definition would not be covered.

Lord Harris of Haringey: But there must be a level, because if you are the chairman of the national executive of a political party, presumably you do have a degree of authority.

Karim Palant: It is less about individuals and more about the page of the party, for example, that would be counted. I would have to double check whether that applies.

Lord Harris of Haringey: But within political parties there are all sorts of pages created by local branches and organisations, some with more

care than others.

Karim Palant: Absolutely.

Lord Lucas: Is there not a middle way between overlaying, saying, "This is false" and doing nothing, and just making it easier for people to find a serious challenge to what been said? As Lord Harris says, you tend just to get competing craziness following contentious statements and it can be quite hard to find the serious argument against what this politician has said. Would it not be helpful if you made that easier?

Karim Palant: With news, it is certainly the case that we make recommendations about other sources of news. I will have to look at how we do it for political parties and how that works. It would clearly be sensitive, for a whole host of reasons, as you can imagine, but I will find out whether we do it specifically for politicians and whether, when you engage with one politician, we might suggest to you other politicians and other related content. I will get back to you on that because I do not 100% know.

The Chair: Karim, you have been very patient. Everything about this Committee that I have learned over the last however many months is that this is in essence about trust. Lady Kidron is quite right; this is a new world we are moving into, a world where, whether you wanted it or not, you have colossal responsibility. It may well be that you are too large. That is a separate conversation for a different committee, but it is possible.

We need to know that you are doing everything you can to build trust, as opposed to undermining and destroying it. The Committee is here to see how digital technology can be used to build a more trusting society, a more politically effective society, if you like, that will not be undermined and eventually maybe even destroyed by it. That is a responsibility that you and your colleagues and your boss have and that is why we have taken as long as we have with you: we are looking to you to do better.

Karim Palant: Would you like me to respond?

The Chair: We are looking to Facebook to do better because we need you to do better. The present situation is very alarming to many of us.

Karim Palant: I agree that there is a very real pressure on us to do a great deal to continue to improve. Having worked on the 2017 and 2019 general elections in the UK, the transformation in the infrastructure that is in place to tackle things like abuse of Members of Parliament, misinformation and transparency around political advertising was a huge step forward and made a huge difference. I know that a lot of Members of Parliament came back from the 2017 election feeling that they did not have any way of contacting us directly, in the right way, to talk to us about online abuse, for example. That is something we worked a huge amount on.

I absolutely concur that there is a lot more we can all do in this space. I

totally agree with you on that, but I would point, for example, to the transparency tools we have put in place on political advertising and the coverage during the general election, when people scrutinised what people were doing, the adverts people were placing, and asking: "Who are these people? Why have they posted this advert? Where is their funding coming from?"

Those are the questions that journalists were asking. I personally, as a British voter, really hope that that will drive change in the way the regulatory system works, so that we can get a clearer and more trusted system in place for that kind of online activity. I think that is really important and I totally share your goal.

The Chair: Thank you.