HOUSE OF COMMONS

Digital, Culture, Media and Sport Committee

Oral evidence: Online harms and the ethics of data, HC 646

Tuesday 13 October 2020

Ordered by the House of Commons to be published on 13 October 2020.

Watch the meeting

Members present: Julian Knight (Chair); Kevin Brennan; Steve Brine; Philip Davies; Alex Davies-Jones; Clive Efford; Julie Elliott; Damian Green; Damian Hinds; John Nicolson; Giles Watling.

Questions 176 - 235

Witnesses

I: Dr Jiahong Chen, Research Fellow in IT Law, Horizon Digital Economy Research, University of Nottingham, Carly Kind, Director, Ada Lovelace Institute, and Dr Jeni Tennison, Vice-President, Open Data Institute.

Examination of witnesses

Dr Jiahong Chen, Carly Kind and Dr Jeni Tennison.

Q176 **Chair:** Welcome to the Digital, Culture, Media and Sport Committee. This is a hearing into online harms and the ethics of data. We are joined this morning by Dr Jeni Tennison, Vice-President of the Open Data Institute, Carly Kind, Director of the Ada Lovelace Institute, and Dr Jiahong Chen, Research Fellow in IT Law, Horizon Digital Economy Research at the University of Nottingham. Good morning and thank you for joining us. Before I start the first questions, I am going to ask members to declare any interests.

Damian Hinds: I chair the all-party parliamentary group on education technology.

Chair: Does anyone else have any interests to declare? In that case, I will proceed.

This is a question to all three panellists, but I am going to come to you, Dr Jeni Tennison, first. You may have seen that the Competition and Markets Authority has started to recruit experts to look at algorithms in relation to, for example, companies like Google. Do you think that is wise? Do you think that is something that regulators are very capable of or skilled up to do? Do you think it is something that can be applied in other areas of data ethics?

Dr Tennison: Thanks for the question and for having me today. I think that the role of regulators around data ethics and the ethics of use of algorithms and technology in general is extremely important, and they need to be equipped in order to tackle those questions. They need to be intelligent regulators from that perspective. It is the case across different regulators. Although ICO obviously can be the regulator around information in general, there will be different considerations around ethics, the ecosystem that needs to be engaged with, and the kinds of impacts that misuse of data and technology might have in particular sectors. For that reason, different regulators—whether it is Ofgem or Ofqual, whatever regulator it is—do need to have that internal expertise. For CMA in particular it needs to be thinking about competition aspects of the role of digital platforms and the use of data, which are very broad. Yes, I do think it is important for that expertise to be there in regulators.

Carly Kind: Thank you for the question and for having me. We have been doing some work on this at the moment at Ada Lovelace Institute. I think what you are talking about refers to a set of tools that regulators might have to inspect algorithmic systems, and that might apply across a range of applications of algorithms throughout the public and private sectors.

In terms of what regulators need to be able to do that, you referenced capacity and capabilities and that is a huge part of it. That is technical

skills and the ability to interface with data to scrutinise them, but technical skills alone are not only what regulators need to be able to inspect algorithms. They will also need a legal framework within which to inspect those algorithms. They will need the powers to get in there and to assess the algorithms against a certain set of standards. I am sure we will talk more about what those standards might be because the broad notions of ethics will not be sufficient. They will need specific thresholds and standards that algorithmic systems have to meet.

They will need some powers for demanding disclosure of information or demanding access to algorithms. In the private sector, for example, you might imagine Ofcom, if it is to get the online harms remit, will need some powers to demand access to social media platforms' algorithmic systems in order to scrutinise those systems and perhaps also some powers to demand disclosure of information back to the regulators.

Then I think they will need an infrastructure in order to match those technical capabilities with the powers and do the kind of audit that we might have previously seen in the financial sector, that real scrutiny of systems. That is not only about a technical audit but also interviewing, demanding information, and so on.

I think the CMA has proved itself one of the most forward-thinking regulators, but it also is blessed with quite a strong framework within which it can start to look at these issues. The challenge for us is to start to look at where else in society algorithms exist that need to be inspected by regulators and how we start to skill up or tool up those regulators in a range of different ways with the framework and with the capabilities in order to be able to look at them. I personally think there is scope for doing that both with respect to private sector algorithms—social media or dynamic pricing algorithms, for example—and also with algorithms in the public sector.

Dr Chen: Thank you very much. I think there has been a lot of discussion about creating one single regulator in the digital sector to combine the powers of data protection authorities, consumer protection authorities and competition authorities. It is very important that we understand what sort of powers this one single authority should have. In terms of the competition powers, which is what the CMA has been doing, it has already done some work in the digital platform market, and one thing highlighted both in its work and our own research is that competition is a real problem, not just regarding Google and Facebook, but other potential markets at the moment and in the future.

The Government have already made a number of initiatives in the private sector, such as the open banking, open finance and the smart data initiatives. I think it is very important that regulators, including the CMA, are involved in those initiatives early to ensure they have enough resources and staff and skills to scrutinise them.

Q177 **Chair:** Thank you. To follow up with you, Carly Kind, regarding what you

were saying in answer to that first question, how do you think legislators build better trust in the way data is used? Is there any sense that there is robust self-regulation? The reason I am asking that is because we are just about to publish the Government's response to our report into misinformation and a large part of that response effectively relies on self-regulation. I just wonder whether you think that is robust enough and whether we would need perhaps to move to a system—you referenced financial services—like compliance officers in these social media platforms. What are your thoughts on that?

Carly Kind: It is a good question. I think that it is quite clear that self-regulation, particularly in the case of platforms, has failed, and that is evidenced via a number of different factors. One is around the scale and the size of these platforms. Their expansiveness means that they are not incentivised necessarily to meet the standards we might expect from a public trust perspective. That is borne out by what we see in terms of misinformation, disinformation and online hate speech, but also things like bias and discrimination in algorithms. We see pricing in Uber and Lyft platforms discriminating against communities of colour, for example. This relates to the underlying question around the data economy, which incentivises the use of personal data for advertising purposes and it doesn't incentivise necessarily putting the brakes on. I think those actors are not bad actors. Many of them are well intentioned, including the platforms, and they have done a lot to try to address these challenges, but the scale of the problem is so great that having some external accountability mechanism is absolutely imperative in order to start to create an online space that is more hospitable to communities from across a wide range.

We know from work we have done around public deliberation that the public feels strongly about the value of external independent oversight when it comes to any algorithmic system, and that is either in the private sector or the public sector. In terms of private sector algorithms and platforms, Doteveryone's survey from earlier this year found that around 60% of people wanted to see more online regulation, so external regulation of online services, even accepting that that might limit their choice, that it does maybe have a limit on innovation to regulate these platforms more strongly. People accept that and want to see that more, and that is even more stark when you speak to particularly marginalised communities or vulnerable communities; black women, for example, who are 27 times more likely to be abused online than white women and men.

Coming at this issue from the perspective of vulnerable groups, I think the imperative is even stronger to put some regulatory mechanisms in place. I think that has to be done hand in hand with the platforms in order to be effective, and it needs to be responsive. It needs to be agile and I think it should be in the hands of regulators who are able to adapt with the technologies, because we know that technologies are outpacing the ability of legislators to keep up.

Q178 **Chair:** That is a really good point because the truth of the matter is it is

very difficult to envisage a regulatory framework that does keep pace with growing technology. That is a personnel issue as much as anything else because the fact is that anyone who goes and works for a regulator may be slightly out of the loop to a certain extent almost within a very short time. How do we meet that challenge?

Carly Kind: The platforms have a monopoly on excellent talent coming through computer science, for example, but also in other disciplines. That personnel issue is going to be a challenging one. Regulators simply cannot offer the incentives and salary to join that a Google or a Facebook can.

Let's be clear: nobody has the answer for how to regulate online technologies, or perhaps my fellow witnesses do and I don't know about it. Worldwide, I don't think we have seen there is one simple answer to this problem. What we are moving towards is an agile regulatory framework that tries to bring in ethical considerations and puts in place processes rather than hard and fast rules. Things like impact assessment, risk assessment, audits, inspection, the ability for people to get redress, these types of tools and processes can help us evolve a regulatory framework without just being around a hard and fast set of rules that will be outdated when the platforms do their next software update.

Q179 **Chair:** Great, thank you. On transparency, Dr Tennison, should the state lead the way in revealing algorithms, giving people greater access rights to their data that is concerning them? Should algorithms be FOI-able by the public for private institutions and the Government?

Dr Tennison: The short answer to whether the state should lead the way and whether Government should act as a role model is yes. The state has particular controls and powers over our lives, which means that their use of data and algorithms has particular effects on us. They are life-changing effects. They are effects, as we saw over the summer with Ofqual, which determine whether we get into university or not. Those algorithms and that use of data are not something that we can walk away from as citizens. We can't vote with our feet away from the use of that data and algorithms over us. Therefore, it is particularly important that Government in their use of data and algorithms are particularly able to be scrutinised and understood, with particular proactive communication about what is being put in place, with good monitoring and evaluation of the impacts that it is having on particular groups in society and on the outliers and the wildcards, the people who do not fit in with the averages. There do need to be these mechanisms for being able to detect and provide redress for when there are errors in the way in which those algorithms are used. All of those issues have been surfaced over the summer with Ofqual.

The other thing that the state can lead the way on and Government can really demonstrate is that the use of data and algorithms is not the only way to achieve many of the results that we want to achieve. We should be thinking in all these circumstances about the balance between the

human processes and the algorithmic processes that we put into place and thinking of it as a whole system rather than just going all in on using data and algorithms or assuming that data and algorithms will somehow give us a better result.

As for whether algorithms should be FOI-able, I think we have to recognise that transparency around algorithms and data are important but can only take us so far. Even if you have access to an algorithm, you cannot always understand what it actually entails when it is run in real circumstances in the real world. That is why monitoring and evaluating the results of algorithms, and openness about that monitoring and evaluation, the data that comes out, the results of those algorithms, is really important, not only on a global level—so to enable watchdogs and regulators to inspect those kinds of results—but on a very individual level. Individuals should be able to see what information about them was used in order to come out with a particular result. That is particularly the case when it is one that has a massive effect on their lives.

Q180 **Chair:** Dr Chen, do you have anything to add to that?

Dr Chen: Yes. The only comment I would add is that there also need to be debates and consultations before any systems that would make important decisions about people are put in place. For example, we have an anti-discrimination law under which, based on certain factors, it would be illegal to treat people differently. What we are seeing now is that a lot of complex systems are put into place and then we have no idea what factors have been taken into account and how these factors have been accounted for. Decisions as important as, for example, going to university or even financial decisions in the private sector should be subject to prior impact assessment and also consultation with the general public so they can think about whether it is fair to make decisions based on, for example, what postcode I am in, how many steps I am taking per day according to my app, or what I have said or done on the internet. These are very important and, of course, ex post scrutiny is important and making algorithms FOI-able is also important, but I think prior measures should also be put in place.

Q181 **Alex Davies-Jones:** Thank you all for joining us this morning. We really appreciate you taking the time. The Committee has looked extensively at how data are collected on social media platforms but, as the Secretary of State himself has acknowledged, we are living in an increasingly data-driven society. Dr Tennison, how do data collection techniques across the rest of the economy differ from, say, that of social media and are the underlying concerns the same?

Dr Tennison: Gosh, what an interesting question. There are various ways in which data are being collected. For social media platforms, you see people providing information through their posts that then gets interpreted and things about them get inferred from that information. They also have a heavy reliance on behavioural data, so the likes and the

retweets and all those kinds of things that then enable you to infer lots of other information.

In the rest of the economy, obviously the data can be similar. It can be about people. It is about the accounts that you have. It can be about, for example, your energy consumption, which can again reveal things about you, like when you have gone on holiday or quite how warm you like it in your house, that kind of thing. There is that kind of information that you have provided, there is that behavioural information, but there is also a whole bunch of information that is perhaps less associated with you as a person. It might be something about your household as a whole. It might be something about the environment that you live in.

One of the things that we get thoughtful about is when there are, say, sensors outside in our environment looking at air quality or the amount of traffic that is going past. That is not personal information in the same way as information that we provide through social media, but it is still information about our communities. It can still reveal information about our habits sometimes. That to me is the bit that gets more interesting, how that information, which is about our communities and about our lives, gets interpreted and used and the ethics of the use of that information as it then affects our communities, our families and us as individuals.

That was a very good question and that probably wasn't a very good answer, but for me, then, going beyond personal information and looking at the ethics of information beyond information that is just about us is one of the big challenges as we move away from social media.

Q182 **Alex Davies-Jones:** Do you think people are aware of how much data are collected on them and used beyond social media? Personally, I was not aware of all of what you have just told me now.

Dr Tennison: I think that in general people are not aware of the details of what data are collected and how they get used. That is because it is extremely complicated. It is not because it is a failure necessarily even of communication. It is just because it is very complicated to understand all the ways in which data are collected and all the ways in which they get passed around and all the ways in which they are used and might be used in the future. That is a very complex thing, which is one of the reasons why issues like informed consent and how we understand the permissions being given, particularly up front, for the collection and use of data get really difficult.

Q183 **Alex Davies-Jones:** You have just answered my next question for me, which was about how we make these processes—they are very opaque, even with GDPR—more open and make people aware of exactly what they are consenting to with presumed consent, for example.

Dr Tennison: Obviously, up front, transparency and proactive communication is incredibly important, but often people do not clock that until it has a real effect on their lives. As I was saying earlier, constant

and iterative re-examination of the impacts of technology, of the impacts of particular data and algorithms, is essential. I would put in a particular plea for use of more community engagement—citizens' juries type of engagement, which I am sure Carly can talk about at greater length—as a mechanism for constantly engaging with the people in communities who are being affected by these technologies.

Q184 **Alex Davies-Jones:** What are the limits of a model of data collection based on consent or presumed consent?

Dr Tennison: There are a few limits. I already talked about the degree to which we can be informed and the complexity and the fact that it is probably too much for us to really understand what the implications are of data collection and use even right now, let alone in the future, and what it might entail in the future.

The other big limit is this piece about how your consent might not be the only thing that impacts how data are used that are about you. An easy example is when you live in a household. Your partner or even your children having consent for data collection from a particular device that you have in your home will affect you as well. If you have multiple people, if lots of middle-class, middle-aged white women like myself give consent for data about them to be collected, that data can also reveal things about myself because I am part of that group, even if I have myself individually withdrawn consent about the collection of that data. Data is not just about us as individuals, they are also about us as families and groups and communities, and also about the whole set of people who are like us and data that they provide can mean that organisations can get insights into us.

Q185 **Alex Davies-Jones:** Thank you very much. That brings me on quite nicely to my next question. Carly, I think it may be more geared towards you. Dr Tennison has described data as a group of people and this could be colleagues. Last week, I was contacted by a constituent of mine. She has been working as a nurse throughout the coronavirus pandemic. She told me that it has now become commonplace for photographers to covertly take pictures of her and her colleagues at work in the local testing centre. She completely understands that there is media interest in the testing because of the unprecedented situation, but she has now become extremely distressed to find that this image has become on sale as a stock image on a website. At no point was she asked for her consent to this. What are your thoughts on this type of data collection and this type of data harvesting?

Carly Kind: You have some incredibly complex and difficult questions and I am sorry to hear that about your constituent. That must be really alarming for her. I cannot really speak to her legal entitlements in this regard, although I have no doubt that she has some so she should seek some legal advice. More generally what you are referring to is probably this blurred line that now exists between what is in the public domain and what is in the private domain and the control you have over—

Q186 **Alex Davies-Jones:** What protection do people have in their workplace, for example?

Carly Kind: Again, I think there are legal answers to that question that I would not pretend to know the answer to.

More broadly, there has been over the last decade a blurring of what is public and what is private and what control we have over information about us and images of us that exist and end up online and how able we are to control that. A big part of that problem is the lack of agency we have to remove data, to shut down things. The GDPR has provisions that should help with that, things like the right to be forgotten or the right to have your data deleted, or the right to data portability. In the GDPR, there exists a right to take your data from one platform and move it to another as a way to kind of vote with your feet and say, "I am not happy with the service I am getting here, I want to move here".

These types of protections have not really been fully realised yet and we don't really understand the extent to which they can be realised. There are two problems there. One is that most people do not feel empowered to exercise their data rights, either because they don't know them or they don't know how to exercise them, or the barrier for exercising them is quite high because it requires either you go to the ICO or you get a lawyer. That is challenging for most people.

The other part of the problem is that the regulators who should be proactively trying to enforce these rights do not have the capacity because they do not have the resources, they do not have the numbers. The ICO has effectively become the regulator for everything, because data are everything, and it simply just does not have the people power to enforce all these micro violations, which are very important in one person's life but in the grand scheme of things there are tens of thousands of them happening every day.

There are obviously specific problems for this individual and there are specific cases, but we should look at it as a structural problem, I think, which is about getting people to feel that they have agency over their data and the digital literacy that that requires, and also putting the regulatory framework in place to make sure that there is someone looking out for their data rights as well.

Q187 **Alex Davies-Jones:** That brings me quite nicely on to my last question. Obviously, the coronavirus pandemic has now meant that we are all providing our personal data every single time we go to a pub or a restaurant. However, some women have reported receiving unwanted texts and calls by people who are using the test and trace data to harass them, often sending multiple pushy messages asking them on dates and commenting on their appearance. Some have reported feeling threatened, scared, and now they no longer want to provide this important data, and you cannot blame them based on their experiences. The Information Commissioner's Office has made it very plain that this

does go against data protection law, but what can be done to stop these abuses from happening in the first place?

Carly Kind: Another good question. If we were to have a do-over on test and trace, things like the contact tracing app, which now provides a QR code that people can scan in a very privacy-protecting way, is a really good answer to that problem. If we had had that from the start, we might have avoided the handing over of personal data and the kind of wild west of data protection that is being put around it. I think there are technical fixes for this problem if we were to be able to start again.

The problem is that we now have this crisis of public trust as well. It doesn't take many incidents to erode public trust in these systems. It doesn't take many media stories like these poor women or a problem with the contact tracing app for people to start to worry about that. I think the Ofqual algorithm is a good example of that—I'm sure we will come to it later—wherein the amount of damage that has been done to public trust in the use of statistical models by that whole incident far outweighs what actually happened.

That tells us, I think, going forward that there is a very high bar that needs to be met with any new data-driven intervention, particularly when it is put forward by Government or by trusted government institutions such as the NHS. They need to meet a really high bar before that is rolled out. There is no such thing as beta testing on the public when it comes to these new technologies and these new systems because you do risk eroding the public trust and they pull away from these types of interventions when they may be beneficial. I think we all agree that contact tracing is an incredibly important thing to be doing and you can only do it by collecting people's data, that is clear, but you have to put in place a system right from the very start that deserves the public trust, otherwise you risk losing it.

Alex Davies-Jones: Absolutely. Thank you very much, Carly.

Q188 **Damian Hinds:** Can I start, Ms Kind, again with you? It strikes me that this is a sort of multi-layered issue. We rightly get very concerned about the way that some algorithms and formulas are deployed and the implications of that for prices people are charged, the job they can get, how they fare in the criminal justice system; it goes on and on. One layer back, there is the way that data are combined and, as both you and Dr Tennison were saying, the inferences that are made about people based on some things that they have declared and other things that they haven't and did not know they were revealing or not revealing, and sometimes that can entrench stereotypes. It all starts with data collection, with being able to get hold of this mass of data to work on. What evidence do we have that people care?

Carly Kind: That people care about their data being collected?

Damian Hinds: Yes.

Carly Kind: There are some very good public studies that show that they care very much, but I think—

Q189 **Damian Hinds:** Forgive me for interrupting. As politicians we know acutely the difference between opinion polls and what people actually do. Is there any revealed preference reason to believe that people care?

Carly Kind: I would contest the view that you can only tell if people care if they, for example, do not use Facebook. The implication is if you use Facebook you do not care about your privacy. That is simply not the right way to think about things. Facebook has become—

Q190 **Damian Hinds:** Forgive me for interrupting, but there is not using Facebook, there is also just using the full range of options within Facebook that are available or even saying no to the thing that says, "Do you consent to these four different types of cookies?" It is not quite as simple as contracting out of something that many people these days regard as essentially a utility.

Carly Kind: At the basic level, there is a lack of choice about using the tool to begin with. If you want to participate in any community activity, for example, you have to have a Facebook account. What you are saying is correct that you have some choices around the default settings on your applications and you have choices around accepting cookies. I don't have the public survey or quantitative evidence to support how the public feel about that, but I think anecdotally it is clear that people find making those choices difficult. I don't think they find that platforms make it easy for them to opt out of data collection. In any event, a platform like Facebook does not allow you to opt out of all data collection or all inferences. It only allows you to opt out of receiving personal advertising based on that data collection.

There are some structural problems that just simply don't allow you to move out of this data economy, which is everywhere you go, and even if you did not use online services, it would still be the case that when you visit Holland & Barrett, for example, to buy some vitamins, they are collecting data on you and matching that with data that they get from their Facebook lookalike audiences, for example. The whole economy and structure is incentivising the collection of data.

Q191 **Damian Hinds:** Indeed, before Facebook was invented people were doing versions of that. When people got a Tesco Clubcard, they implicitly or explicitly—probably implicitly—traded information about themselves for clubcard points. When they got a Visa card they traded information about themselves for getting free credit until the end of the month. What do we know about how much people know about how data are used and why they are collected?

Carly Kind: Increasingly people know more and they have less agency to do something about it. In particular, the Cambridge Analytica scandal increased public attention and public understanding of data collection, but perversely I think their feeling that they can do anything about it has

declined since that time. That same time coincided with the GDPR coming into effect, and what we have not seen, I think, since the GDPR is a broad-based increase in public empowerment around their data.

Having said that, I do not think we have necessarily the studies that go to that level to really understand people's relationship with data. Mostly this is going off qualitative interactions with members of the public, which the Ada Lovelace Institute does in citizens' juries and similar public deliberation events. People care about their data. They want their data to be used in the right way. Equally, they want their data to be used to benefit them. They want to see data used in the public good and for public benefit. What I don't think they feel they necessarily have is complete agency and control over who has access to their data.

Q192 **Damian Hinds:** Yet our time online is peppered these days with interstitials popping up asking us what we approve of, what we do not approve of. What have researchers done to understand people's behaviour and how they interact with those things? It strikes me anecdotally that in the first two or three days of a new set of regulations people notice the things about cookies and thereafter it becomes just another nuisance in life. What have these academics done to understand how people interact with them?

Carly Kind: I can't speak to any particular research studies. I am not sure if my fellow witnesses have things to draw on.

Q193 **Damian Hinds:** I am willing to open it up to the other witnesses. What do we know about how normal people interact with all these questions about my rights, my data, this type of cookie, that type of cookie, opt in, opt out? What do we know?

Dr Tennison: I cannot recall the studies off the top of my head here, but I can certainly follow up after the meeting. I will say that one of the phrases I have found useful in thinking about this is the concept of digital resignation, so the feeling that you are just resigned to the fact that data are being collected. There is a study, and again I cannot recall the name of it off the top of my head, which looks in more detail at that and how that is applied. If I can come back later, then I can provide some links to those kinds of things.

Q194 **Damian Hinds:** Thank you, Dr Tennison, that would be very useful. I will stick with you for a moment. We have talked about the relevance of data consent. I am also interested in the relevance of the concept of data access. Yesterday I did Google Takeout, which is a fascinating experience where you say, "Tell me everything you know about me, Google" and a warning comes up that says, "Warning: this process may take some time (hours or days)". Mine did not take that long, but sure enough the range and volume of stuff maintained is huge: entire photo collections. It makes you think why bother spending money on cloud backup because it is actually all there, you can just get them all back when you need them, in Google's case from 42 different operating applications. How many people do that? How many real, normal people care about what is stored about

them and what do you think they would think if they discovered it?

Dr Tennison: As far as I am aware from the studies that I have read, a very, very low percentage of people care enough in order to go and find out that information. When they are presented with it, for example, through the kinds of qualitative studies that Carly was talking about or in citizens' juries, then the reaction is sometimes shock but sometimes, "Well, I just assumed that that was the case anyway", that kind of digital resignation piece as I said.

For me, that kind of data access has two kinds of roles. There is the role that is the transparency and understanding both what has been collected about you but also the inferences that have been made about you so that you can point out the idiocies of some of those inferences and perhaps seek to change them. Then the other role of that kind of access is to give you the ability, as Carly described, to be able to move to a different kind of service, taking the data that is about you with you so that they can provide an alternative service for you. Data access in both of those transparency roles and also as a portability role is really important.

Q195 **Damian Hinds:** A similar question to what I asked Ms Kind earlier: on a very basic level, what have academics done to discover who does that, how many people do that?

Dr Tennison: Perhaps I can pass over to Dr Chen for that.

Dr Chen: My colleagues here at Nottingham have done some excellent work with young people. They have been working with groups of youth juries and having very in-depth interviews with them on how much they care about how data has been collected about them and how it has been used. Of course, we see what you have mentioned before about the privacy paradox, so people say they care but then maybe actually they don't. What we found in those interviews is based on how much knowledge they have of these technologies and what they have done when they are using the internet, there is a very strong indication that they actually care. At the same time, there is also a very strong sense of resignation, as Jeni just mentioned.

I think the regulatory or practical implication is that we might be relying too much on individual consent as a regulatory mechanism. We are talking about protecting individuals, empowering them, giving them the autonomy, but also if we over-rely on the individual consent we might be shifting the compliance burdens to individual users. That is why I think there needs to be some technical or legal structure to support people to decide what sort of data are being collected about them and then how they can control the data about them in a way that would respect the fact that they have priorities in life, they have limited energy and attention and time. I think that would be a very big challenge for regulation, but that is something we need to look into in the future.

Q196 **Damian Hinds:** Do you think there is sometimes a bit of a clash of terminology here? When we say data, when folk like you say people's

data, you understand that to mean a very, very wide definition of behaviours and attitudes, inferences and so on, as well as the things that people have just declared about themselves. In my experience talking to constituents, when people talk about "my data" they just mean stuff about me, things that I have told somebody, like my name, my address. It might include things I have bought, but it certainly would not include things I have looked at. They do not think of that as data. I just wonder sometimes if we might be slightly missing the boat here in terms of what people are thinking about. I wonder, too, how much people think that what they do on one platform might affect what happens in a completely different way. Not all 42 of them are substantial for most people, but how many people even know that Google and YouTube are the same company, let alone all the web of commercial interactions that you can make with third parties? How much do people know about this? Sorry, I will open that to any of the three of you who fancies a punt.

Dr Chen: Again, referring to the study my colleagues have done, there is clear evidence that they have a reasonable sense of, for example, how they have been tracked across the internet. What you have done on one website may end up influencing what you see in a different website, especially when it comes to advertising.

Q197 **Damian Hinds:** Yes, for pop-up ads I think that is so "in your face" that it is impossible to avoid. I just looked at something on Google and then I am somewhere else and suddenly I am being bombarded with it. But news serving up would be a different question. How many people do you think understand that the news they see in their news feed might be affected by things they have looked at on Facebook?

Dr Tennison: Again, this is qualitative, but in a piece of research that we did with the RSA, when we consulted people around their awareness that their news feeds were being adjusted. In fact, their news feeds being adjusted was one of the things that they objected to most in terms of the use of data about them affecting what they saw on social media or what they were recommended. For example, recommendation engines like Netflix telling you what films to watch next they found more acceptable than their news feeds being adjusted based on their revealed preferences through their behavioural interaction with the service.

Q198 **Damian Hinds:** Thank you. I am almost done, but if the Chairman will indulge me with one last question on a slightly different tack, you talked a little earlier about the right to erasure or the right to be forgotten. Thinking particularly of kids growing up these days and when they become young adults, for all of us there are embarrassing things about our youth. Thank God for most of us here we never had to worry about them seeing the light of day. Given all the technology that is available, bots and so on, crawlers across the web, is it realistic to say you should be able to utterly erase something from the internet? Obviously, you can't if it is in somebody's own hard disc, but something that is on the web, however much it has been distributed, ought it to be possible to search and destroy so that a kid who becomes a young adult can erase

something entirely from their past?

Carly Kind: You are suggesting, and I think that is right, that it would be impossible to do it absolutely, but that is no reason not to at least create an obligation on the part of data controllers to make it possible. If you think about the extent of linkage that does happen between data, as you will know, Mr Hinds, there is the existence of something called data brokers, whose job it is to amalgamate data on individual people across the web, taking cookies, taking social media data, taking offline shopping data, and putting it all together into a unique, identifiable characteristic for one person. Their sole job in life is to track an individual uniquely as they move across the web, and there are data brokers such as Oracle that claim that they can do this on up to 5 billion consumers in the world.

There is a mechanism for tracking an individual's data as they exist across the internet, and thus there should be a mechanism to at least try to rein some of that in should there be a legal right to make that claim. If there should be a legal right to make that claim, it should exist vis-à-vis children. The fact that it would be logistically hard to effect absolutely should not mean that there should not at least exist an obligation to enable kids to be able to do that.

Dr Tennison: The fact that data are very hard to absolutely get rid of is one of the reasons why only focusing on data collection as the mechanism for protecting people from harms from data is not sufficient. It is why we would advocate for looking just as much, if not more, on how data then get interpreted and used and affect our lives as on the collection aspect of data.

Q199 **John Nicolson:** I would like to explore how much human bias can leak into apparently objective algorithms. Am I right in saying that algorithms are really only as good as the data that they are based on? In other words, if it is a team of white males who are inputting the algorithm, it is likely to produce an algorithm replicating their perception as the norm. Is that fair, Jeni?

Dr Tennison: There are two aspects there that you are touching on. There is the aspect of the data that are coming in and the limitations that all data have in terms of how they were collected, who they might be about and where there might be biases within that data. Then there is the aspect of the teams and the groups of people in the organisations who determine how the data are used to create an algorithm or use that data to create an algorithm. Again, their biases and ways of thinking about the world, the things that come into their heads as being problems and the things that do not come into their heads as being problems are going to determine what they look at when they are building that algorithm. Biases and problems with the results of algorithms can arise from both those sources.

What is needed is a much more critical view of data, the data that are coming in, and a much more critical view of the processes that are used in order to create technology to get an outcome that has a proper view of

the degree to which we should trust, for example, the result of one of these algorithms and, therefore, what we should do in order to help people, provide them with the redress mechanisms that they need, and enable them to correct errors in the system as a whole.

Q200 **John Nicolson:** Are systemic racism and structural inequality built into some of the algorithms that big companies, for example, use?

Dr Tennison: Yes, in many cases the type of data that get collected about different groups of people, the degree to which there is detail about different groups of people in the data that we have, the fact that sometimes data are collected—for example, if you look at data about stop and search, then it is biased towards saying that more black people need to be stopped and searched because historically that has been the case. Yes, all the structural racism that we have gets built into the data that we have as well as that structural racism being built into the way in which teams create algorithms and technology. Carly would probably have very good things to say about this.

Q201 **John Nicolson:** You mention black people in that context and maybe I can put this to you, Carly. I was fascinated and immensely disturbed to read that when Amazon's facial recognition surveillance software was tested inputting members of the Congress in the United States, in 40% of cases where a congressman's picture was put up, he or she was falsely matched to someone on a criminal database. That is despite the fact that only 20% of people in Congress are black or minority ethnic. That is shocking.

Carly Kind: It is, and that is the most simple example as well of algorithmic bias because that is an inaccuracy issue that is based on how the algorithms are trained—

John Nicolson: I like simple examples because I can understand them.

Carly Kind: Can I give you a slightly more complicated one, which Jeni alluded to? This will be of interest to the Committee because it is about the use of algorithms in healthcare, which is a potentially exciting area to be using algorithms.

In the United States, an algorithmic system that was used to assess whether or not patients would need further care and should be recommended for further care was found to be discriminating against black patients and saying that black patients did not need further care when white patients did in otherwise factually equivalent cases. The reason that that algorithm was biased was not on the basis of the accuracy of the data that were being inputted to train the algorithm, as your facial recognition example was, but because it was trained on the basis of health insurance data, which showed which patients spent more money on healthcare. Because white patients spend more money on healthcare—they are better insured across the board—than black patients, it was recommending that white patients get more healthcare.

That is a much more complicated example to show that this is not only about getting equivalent amounts of data in from different communities but also about the types of data that are collected on different groups and how they are fed into the algorithm and then how they generate the outcomes.

Q202 **John Nicolson:** I remember working in the United States Senate and becoming unwell at one point with the flu and being told not to go to the closest hospital to the Senate because it was disproportionately African-American and, therefore, I would get less good healthcare because there was less health insurance money being pumped into that hospital and that I should travel to a different one. We have this problem here, though, haven't we? We know that the Home Office has already dropped a racist visa algorithm, which was making it tougher for black people to get visas than white people. How do you know if you yourself have been subjected to and discriminated against by a racist algorithm?

Carly Kind: On the visa streaming algorithm, algorithm is probably a slightly strong word. It was quite a simple equation that was being used that certain countries just were not fast tracked. But I take your point. The protections that we need to put in place there are around, first, the ability to know when you are subject to an algorithm at all, and that is not the case currently. We know, for example, at least 60 local authorities in the UK are using algorithmic systems to do risk scoring around certain communities that might be vulnerable to falling into homelessness, for example, or children needing care. They are using algorithmic systems to flag those families and that information isn't public. There is not widespread understanding of where algorithmic systems are used in local authorities and where they are not.

So, first, you need to know when you are subject to an algorithmic decision. Secondly, you need the ability to scrutinise that. You need to be able to apply for a human explanation of that decision, which does exist. That right does exist under the GDPR in certain circumstances. Then you should have the right to remedy that as well and to query it.

The other side of that coin is the people using the algorithm need to understand it and be able to explain it to you, and that is not something we necessarily see in all cases either. It may be the case that a local authority or another public sector agency in the UK is using an algorithmic system that may or may not be biased and it cannot really tell because it is either a black box system that it has acquired from a private sector entity or there is simply not the technical capability to do that type of testing. It is possible to do a bias audit of a system and understand if it is delivering fair outcomes, but you need the capabilities and powers to do that.

Q203 **John Nicolson:** That is absolutely fascinating because that is something that I certainly did not know about and something that is very important for our Committee's work on this particular subject. It is fascinating. So, one has no way of knowing whether or not one has been subjected to a

racist or discriminatory algorithm at the moment, even though it is possible, to use your interesting phrase there, to do a bias audit. Even if not in an individual case, collectively we could discover there is discrimination against black people or Scottish men or whatever it happens to be. That is possible, you are telling us?

Carly Kind: Yes, technically it is possible. It is possible at a simple level of understanding if the outcomes are fair and equal for people from different demographics.

There may be another level of discrimination that exists around algorithms, which is that they are often used on groups that are historically subject to discrimination; for example, people receiving benefits, asylum seekers and migrants, people subject to overpolicing. That is an area in which algorithmic systems and data-driven systems are disproportionately used and that is a separate problem around discriminatory algorithms.

At the point of whether the outcomes of an algorithm and the decisions that it delivers are fair and unbiased, it is possible to test that technically. You do need access to the system, so it should be possible to do by the individuals deploying the algorithmic system, for example. I think Jeni wanted to come in on that point.

Q204 **John Nicolson:** Jeni, please do come in, but again I had an interesting experience of this once because I discovered that I was being surcharged by more than 200% for health insurance on my mortgage simply because I am a gay man. I was not told that I was being surcharged; I only grew suspicious because I thought that the bill was so large. I phoned back and presented myself as a heterosexual rather than as a homosexual and my quote was substantially reduced. They were not interested in lifestyle, somebody living in a monogamous relationship, they were only interested in the fact that I am gay, which was interesting to find out. White men do not often find themselves discriminated against, so it was interesting to see such blatant discrimination. Jeni, you wanted to come in on this.

Dr Tennison: I just wanted to point you to a piece of work that we have done at the Open Data Institute with the Legal Education Foundation around the collection of data about protected characteristics and other indicators in order to assess whether algorithms are complying with the equality duty or are biased themselves. We looked at whether in digital services information about protected characteristics, about gender and ethnicity and so on, gets collected, because you need to collect them for the purpose of assessing whether different groups are being discriminated against. In many cases, they are not collected. Of course, you need to collect them in ways that are sensitive, that do not then feed into being used in the algorithms themselves. But in order to test, as Carly was talking about, whether there is bias in the results of an algorithm you need to collect the data about who is going through it in the first place and that detection can only happen, as you have already

discussed, at that global level. When you are the individual going through the algorithm that is very hard to tell.

The other mechanism we should be exploring here is that of mystery shoppers for these kinds of services. They need to enable third party organisations to try out what happens when they use the name Mohammed rather than the name John when going through an insurance claim, for example, in order to see what those results are, just as you did with trying out using different sexualities to see whether that would have an impact. There is the mystery shopping aspect to enable watchdogs to do that kind of scrutiny but also, and particularly in a Government case, there is a duty to be able to test and demonstrate whether the system is treating people equally and that can only be done with the collection of that kind of data about those people going through it.

Q205 **John Nicolson:** Very clear and very interesting. May I move on to ask you a question about a different topic on Brexit? The European Court of Justice has ruled against the retention of data for intelligence purposes. The UK's Digital Adequacy Agreement is therefore now, as I understand it, in jeopardy. What will be the consequences for the United Kingdom if we crash out without such an agreement? I am told by constituents that it would be very bad for business. Perhaps you could tell us why.

Dr Tennison: All I have seen from every organisation that is dealing with data says we need to have data adequacy within the UK in order to enable the digital services that are grown in the UK to flourish, survive and be able to grow outside the UK.

Q206 **John Nicolson:** Very quickly, lots of people who are watching this in their millions might not know what data adequacy is.

Dr Tennison: This is the assessment of whether or not the data protection in the UK over data about European citizens is adequate to protect that information about European citizens. The European Union only grants that adequacy to third party countries that can demonstrate that the data about their citizens will be protected, will not be exploited by that country and will not be used in order to harm those citizens. That is the idea. The use of data by intelligence services and the powers intelligence services have to access that data is one of the main reasons why third party countries do not get adequacy.

The various relationships that have been set up with the US are ways of trying to secure that adequacy so that the big US companies like Google and Facebook are able to take and hold data about European citizens in their servers in the US. As Brexit comes into action we will be a third party country from that aspect, from that point of view, so those same considerations about the treatment of data about European citizens come into play here.

That is important because any digital services provided by organisations here, if they want to service European citizens, will need to be able to hold the data about those European citizens and so it is very necessary

that organisations here who are either digital services or any organisation that is servicing European citizens should be able to hold that information themselves.

Q207 **John Nicolson:** It is yet another disadvantage of Brexit and an unforeseen consequence because the Government wants us to be able to continue operating with the European Union but we will not be able to. Am I right?

Dr Tennison: Until and unless that data adequacy is granted then that is the case. However it is possible for third party countries to get data adequacy. It is just that it has implications for the rest of the ways we are able to deal with data here.

Q208 **Chair:** Dr Chen, I noticed you raised your hand a little bit earlier. Did you want to add anything to any of the questions that John put?

Dr Chen: Yes, a number of points. We can start with the last question about Brexit. As Jeni just said, by the end of the transition period we will be leaving the EU as a third country so data transfer from the EU to the UK will be subject to further restrictions and there are strong reasons to believe why it is unlikely that the UK will get an adequacy decision, or if the Commission was willing to give the UK one it might be subject to legal challenges.

There is very extensive analysis on the compatibility of the UK data protection regime with the EU. One reason my colleagues believe it is very hard for the UK to get an adequacy decision is the immigration exemption under the Data Protection Act 2018. That goes back to John's point earlier regarding getting a visa based on algorithms. Under the GDPR, data subjects do have the right, for example, to access their own data and to challenge the decisions or object to the processing in general but there are also restrictions.

When it comes to things like immigration, according to the DPA 2018 individuals would not be able to access their data or the authorities have the right to turn down their requests. That is one aspect of how the UK system might be incompatible with the standards set out by the GDPR.

Q209 **Chair:** Just to clarify, you are effectively suggesting that the immigration law from 2018 bakes in a lack of compatibility between the EU and the UK and therefore that makes it more difficult, or you suggest maybe nigh on impossible, for adequacy to be granted.

Dr Chen: That is the analysis and conclusion by some of the leading academics in our country and I have strong reasons to believe that is the case. Also as John mentioned, last week the Court of Justice of the EU handed down another important ruling regarding mass surveillance and the bulk data collection power provided for under the Investigatory Powers Act is another reason why it might be difficult for the UK to get that adequacy decision.

Q210 **Clive Efford:** I apologise in advance if there are some noises coming

from behind me. There is a building site near me that is making some really loud noises so I apologise if that happens. When I started as an MP not every MP had an e-mail. We communicated with text messages and bleepers and things like that. Things have moved on today. The post came and it was a massive post because most things were done by post. Now we have an avalanche of e-mails every day. I get complaints from constituents that they have not heard from me and when I say I do not seem to have had anything from them they say, "Yes, I tweeted at you and you did not respond".

Things have moved on enormously so how do we as legislators make sure that people are empowered to be able to deal with this? With technology moving so fast and so much data being gathered on them. How do they understand the ethics and rules that enshrine their rights but also create a framework where those people who are after their data, trying to hunt it down over the internet, should operate? Do we do enough to educate people?

Carly Kind: I am no expert on digital literacy but I think it is fair to say we could invest a lot more in growing a digitally literate population from a young age, including not only the ability to use technology but also to critically use technology. It can help with things like misinformation and online bullying if we are able to build a digitally literate generation from school level.

There is a charity I am involved with called Glitch that does things like digital citizenship training where they go into schools and teach children how to be good digital citizens online, which is not only about protecting your privacy and understanding privacy settings but also about what online stability looks like and how to behave well online.

We have always as a society struggled to adapt to new technologies. That is not a new thing that has come with the internet. What has changed is the speed of change and the pace and scale at which things change. Inevitable the technology is going to be ahead of Parliaments and lawmakers like yourself, which is why building a culture of ethics is important both in the private sector and the public sector.

Legislation might not be able to keep up but we can do a better job at building a coherent understanding of what public legitimacy for technology looks like and what standards companies have to meet in order to enjoy a social licence to operate and enjoy the public legitimacy of their users and consumers and that is why organisations like the Ada Lovelace Institute and others are trying to work not only with lawmakers like yourself to develop regulation but also with the private sector to try to develop a common understanding with them about what technologies should do to benefit people.

It cannot only come from a reactionary stance. I think it has to come with building a common understanding with these companies. One of the reasons that is incredibly difficult is we are seeing lots of tech development coming out of Silicon Valley in the US, which is not

necessarily something we are able to control. It is not as if we can have a national conversation with companies over here, it has to be a global conversation as well. As China also enters the technology market, increasingly we are challenged by the fact that Chinese technology developed in accordance with Chinese understandings of ethical principles is going to be available as well.

It is a big challenge to have this global conversation about what moral principles technology should adhere to in the absence of legal restrictions but we have to try. There are good things like the global partnership on AI that the UK is involved in that are trying to achieve that.

Q211 **Clive Efford:** Is there a basic set of rules or guidance that people should have that will be transferable no matter where the technology goes? Is there a basic set of things that people can apply?

Carly Kind: At a high level, there is general agreement on what the basic principles of data ethics are. Essentially data ethics are about the moral principles of right and wrong that we should apply to the use of data in technology. I am sure you know there have been scores of data ethics frameworks and principles and guidelines written in the last three years. I think at the last count it was somewhere around 100. Essentially they all centre on six or seven general principles that everybody agrees. Things like non-discrimination, justice and fairness, privacy, security and safety, accountability, transparency, human control, autonomy in human agency and explainability as well.

Making those tangible in practice is the challenge. Translating these high level principles that everybody generally agrees on and putting them into practice is where the devil is in the detail and that is where we see different cultural and national applications differ. I would say we are still on that journey towards a common understanding of what fairness means in the context of technology. What does bias mean? What does privacy mean? These are huge concepts we have to try to make tangible and I would say we are heading in the right direction but there is a lot more work to be done.

Q212 **Clive Efford:** How do data co-operatives and data trusts operate and do they offer a way forward that would protect people's rights over their data?

Carly Kind: I will let Jeni answer that.

Dr Tennison: Thanks, Carly. Data co-operatives and data trusts are types of what we at ODI call data institutions. They are organisations that are set up to steward data on behalf of the community and often that can mean gathering data together from a whole bunch of actors and then sharing it more widely in restricted or well-governed ways. Or it can mean things like just having a whole bunch of organisations and people collaborate around the creation of data sets that are for common good purposes. For example, OpenStreetMap is a free to access map that is edited a lot like Wikipedia is edited.

These data institutions sit in the middle of a number of relationships between organisations that hold data where we might want to have broader access to that data in order to spark research and innovation. The data users who want to do that research and innovation and need to get hold of that data and the communities that are affected by the use of that data—so the people the data are about or the communities that use the tools that get created—sit in the middle and can act as this neutral third party to make decisions about how data should be shared and accessed and for what purposes they should be made available.

Your question was do they give us a way forward? They do. Data institutions have existed for a long time and there are many of them that we rely on as distributors of information. I would point to UK Biobank, for example, as a data institution of that type. What are being looked at now are these new forms of governance within those institutions like co-operative models, using trust law and fiduciary responsibility to set those organisations up. This is a very active area of research across the world, how these organisations can come into play, how to make sure they are trustworthy and how to make sure they are sustainable so we can rely on that data. I would be happy to circulate more material if the Committee is interested in that area.

Q213 **Clive Efford:** Thank you, we would be very grateful if you did. The Government has released this National Data Strategy last month. Do you think they get the tension between the desire to create growth and the need to create a trusted data regime? Do you think they get that balance right?

Dr Tennison: One of the things about the National Data Strategy is it seems to try to set up a dichotomy between innovation and responsibility as if they are competing when in fact we can do both and we should do both. Having responsible processing of data and responsible algorithms is not only the right thing to do but it is absolutely necessary to win trust and to get adoption of those technologies. The issue for me is where it is set up, these two things, as if they are competing or incompatible with each other when they are both completely necessary. We can have innovation and growth with technology and we can do it in a responsible way and that is what we should be aiming for.

Q214 **Clive Efford:** Do you think, for instance, the Government could encourage over-collection of data or inappropriate use through its strategy?

Dr Jeni Tennison: I don't think the National Data Strategy sets out to do that. It is important we have the right kinds of mechanisms in place to provide the right level of oversight, enforcement and protection. This is not just a one-off thing that Government does enacting the National Data Strategy, it is an ongoing set of activities around the use of data and algorithms.

Q215 **Clive Efford:** We have not asked much about the application of artificial intelligence or how data are used. Algorithms are human creations so

they are as effective as the humans that put them together. But with the runaway trolley argument about moral judgments being made by an algorithm controlling, say, a vehicle, just how much do we abdicate the decision-making to an algorithm in a moral judgment like that where you have to choose?

The example I use is the automated vehicle at Greenwich where they were demonstrating smart cities. Somebody put a chair in front of a vehicle and it ran it over. They said if this was a four year-old child then that vehicle would have run over that four year-old child but had it spotted the child, the choice was to deviate one way and go head-on into another vehicle or mount the pavement and hit a group of people at a bus stop. Given that there is human decision-making in the algorithm or whether it is a vehicle driven by a person, how much do we use technology and how much should we let it go to make judgments like that?

Dr Tennison: This comes back to the point I made earlier on that when we are looking at these systems we need to look at the system as a whole, not just the people decisions and algorithm decisions but how those combine and interact to get to the outcomes that happen. Those examples aside, when we look at real examples we will be presented with one-off grey area decisions about how we balance the responsibility between those two, the humans and the technology, and it will iterate and evolve. We will not know what the impact of those things will be until we find the end. What is important is that iterative reflection on "Did this go wrong?" How can we adjust what happened to help it to readdress those balances?

Q216 **Clive Efford:** What I found challenging about it is whether the split second decision is made by somebody who is involved at that split second or by somebody who wrote the algorithm somewhere else completely detached from that situation. Morally, what is the right way forward? Does anyone have an opinion on that? It is one that always challenges me when I read about artificial intelligence.

Carly Kind: It baffles a lot of people. There are a few things to say. One is lots of the focus around ethics have become focused on this trolley problem around self-driving cars and it is a bit of a distraction because self-driving cars are not going to be on our roads any time soon and this type of problem, as Jeni said, is going to be a real edge case, a real grey area. It is not that we should not use it to focus the mind on some of these challenges but we should also not get distracted about it, given there are applications of AI that are every day deciding not to deliver job advertisements to black women or to Jewish people because they are based on racist formulations. There are real live AI applications now that raise ethical issues so we should not get too caught up in that.

Having said that, we should also remember that in many of those edge cases, computers will be better decision-makers than humans in some instances and we should be able to understand and identify when a computer will be able to act in more of a split second than a human,

which is probably the case in some of those drastic cases. But, even if it is demonstrably better than humans, we still have to trust it. Your instinct is still, "It makes me feel queasy" and that is completely understandable. Most people feel like that.

How do we build the trust in the system? Lots of that will be about understanding it. As long as it is something foreign to us, it is like an AI over there doing something, we are going to feel queasy about it, so how do we get normal people understanding how artificial intelligence works, to demystify some of it and normalise some of those things? That is about education and it is also about bringing people into these types of deliberative processes like citizens juries to interrogate and understand these and that is how we can start to build public trust. To reassure you perhaps, we are not going to be faced with that problem tomorrow but we definitely need to start understanding how we can become comfortable with it.

Q217 **Kevin Brennan:** In a way we are already facing those decisions. The Boeing 737 Max aeroplane crashes in recent years, where automation and computer technology was designed in such a way that human beings could not switch it off because computers are supposed to be better decision-makers in a crisis, led to hundreds and hundreds of deaths so it is already an ethical massacre on our doorsteps.

Could you put your hands up if you have a smart speaker in your home? We have one of the three. Can I ask you, Dr Chen, why do you not have a smart speaker in your home?

Dr Chen: The short answer is I do not feel the necessity and also I do not feel comfortable having a smart speaker that might potentially record everything I talk about or the conversations I have with my guests. But I understand and appreciate the fact that it is useful for a lot of people so the question is not really about whether one should or should not have one but what sort of safeguards or mechanisms you have to be sure people trust these technologies.

Q218 **Kevin Brennan:** You do not trust them, do you? That is what you just told us.

Dr Chen: Not 100%.

Q219 **Kevin Brennan:** Why not?

Dr Chen: Because it is still a little bit too complicated for me to understand how it works exactly. Like I said, I do not see the necessity in my everyday life so after that balance in thinking I decided I did not need it.

Q220 **Kevin Brennan:** You are a research fellow in IT law at Horizon Digital Economy Research at the University of Nottingham and you feel you do not understand whether or not this technology might be recording your every activity in your home and somehow or other misusing that data. Is that correct?

Dr Chen: I suppose I would have the skills to figure it out but, like many other fellow end-users of these smart technologies, I do have priorities in my life and I decided not to spend all those hours just to read all the Ts and Cs and policy.

Q221 **Kevin Brennan:** You are a very wise man, Dr Chen. Some time ago Google sent a number of Members of Parliament one of these smart speakers through the post, as a little gift, and I decided not to activate mine and to put it somewhere it could not hear me. Do you think I was wise to do that?

Dr Chen: I don't know enough about the details so I don't think I can comment on that.

Q222 **Kevin Brennan:** Carly Kind, you did not put your hand up either, what is your view about these speakers?

Carly Kind: I suppose if I interrogate it beyond what Dr Chen said, which I agree with, it is that I do not trust the companies that make them. I know that companies like Amazon and Google have over the years expanded the way, and changed their terms and conditions about the way, they use data and I would not quite trust that the current understanding about what they might do with that data would remain the future understanding, that they would not change their terms and conditions going forward. That is probably what it comes down to for me.

Q223 **Kevin Brennan:** Dr Tennison, you did put your hand up. You have one so you clearly do trust these companies and you are not worried at all about where they might use the information on your smart speaker at home.

Dr Tennison: For me there are two things. First, there is a limit in my estimation to what Amazon—I have an Amazon Echo—could do with that information that would be problematic for me and, in particular, because I am in a very privileged position the risks of misuse of that information I think will have very few consequences for me and my family. For me that judgment goes in a different direction. But again, I completely understand people who decide not to. To go back to earlier conversations, our ability to understand and make informed decisions about the use of information that is collected, and including the potential use of that information in the future, is very limited. That is why individual decision-making about these things is itself limited and why we can't only rely on individuals in order to get the protections we need, either as individuals or as society.

Q224 **Kevin Brennan:** Did you see the story in today's *Times*, "Woman hacked Alexa to scare off ex's new girlfriend"? "A management consultant allegedly hacked into her former partner's Alexa account to harass him and his new girlfriend, a court has been told. She is accused of using the device to switch the lights at the man's property on and off and order the woman to leave. She also allegedly hacked into the man's Facebook account and posted nude photographs of him". Obviously this is about a

breach of security, but I suppose my question is ultimately this: do you think there are enough safeguards in place for these sorts of devices that are being put into people's homes, Dr Tennison? You have one in your home; do you think there are enough safeguards in place in relation to them?

Dr Tennison: This is a question in general about a whole range of technologies we have in our homes, not only these smart speakers. The kinds of protections that there are around the IOT technologies—internet of things technologies—has been called into question many times. They are frequently unable to be updated with security updates, for example, which means that there are often risks with simply having one of those technologies in your home when there are third parties who might attack with it. It is really important to get the basics of security around those technologies right. That level of compliance is one level, and then we also have to think about the level of ethics around the use of that data and so on, so the broader implications about the use of that data.

Q225 **Kevin Brennan:** Thanks. Dr Chen, can I ask about this concept of the smart city that we hear so much about. Does that concern you at all, that a citizen is effectively unable to opt out of living in a smart city and the implications of that?

Dr Chen: Yes, absolutely. In our area of research there have already been a lot of discussions about implications of being monitored or having your data collected wherever you are in a smart city. The real concern here is when data are collected they can be repurposed for other uses. There is no way we can tell in a smart city what has been collected about us and later what such data may be repurposed for. A lot of the data have been collected initially for well-intended uses, but let us say in some circumstances it could be used for purposes that were not originally envisaged, and there might be unintended consequences. Yes, absolutely. That is why we need debate and discussions before we start to build smart cities.

Q226 **Kevin Brennan:** George Orwell wrote "Nineteen Eighty-Four" back in the 1940s with this idea of the state monitoring you in your own home, and then the internet came in and everybody thought, "There is so much information available the state can no longer control information and, therefore, people will be freer as a result". It seems to me the pendulum has swung back in that debate. Do you think this technology makes totalitarianism essentially, ultimately easier, rather than more difficult?

Dr Chen: While state surveillance is a very important consideration, and I do think having a city capable of collecting data everywhere is a concern for having a totalitarian regime, I also think we need to be aware that there are also private-sector considerations here. These infrastructures in smart cities will also make it easier for business to exploit our data.

Q227 **Kevin Brennan:** What do you think is the greater threat to the individual liberty in this instance—and perhaps this is geographically based—the

state or the private sector, in relation to their ability to recognise your face as you walk down the street, to know exactly where you are, who you meet, where you go, when you got on the bus, how many steps you walked in a day, so on and so forth? Ultimately, which is the greater threat, do you think?

Dr Chen: To me they are equally concerning. To some extent the commercial practices might be even more worrying, especially when businesses are teaming up with state actors. We have already seen that in some countries. Some commercial uses of algorithms, data and monitoring techniques might end up being used for public purposes. I think we need to have debate on both fronts. "Nineteen Eighty-Four" is a very useful metaphor to initiate discussions but we also need to think about the private-sector implications as well.

Q228 **Kevin Brennan:** The Government are developing a secure-by-design framework to set minimum standards inside the security for smart devices. Does the focus on cyber security address concerns about smart technology, or are there further ethical questions about smart technology that need to be considered when designing regulation? Perhaps I can ask Carly. Is that a question you might be able to tackle?

Carly Kind: I can speak at a high level, though I would not claim to be an expert in smart technologies. There are a couple of things I wanted to say. One is that we should be careful about the allure of smart cities tempting us down this techno-solutionist pathway: technology solves all the problems that our cities have. There are a great many problems our cities have that can't be solved through smart technology, and we should not overinvest in the tech at the risk of underinvesting in things like roads and other infrastructure, things that really do impact people's lives, the high street, and so on. That is a broad warning.

Jeni made the point about the general insecurity of smart devices being a real problem. Therefore, any investment the Government want to make in improving security standards in smart technologies is very important, given just how many there are in the internet of things these days. Clearly, as Dr Chen said, it is not only a question of security, it is a question of data use, data privacy, and the ethics of how the data are used as well. It is a combination of those various things. It cannot be solved through security alone.

Dr Chen: My department has submitted a response to that consultation on secure-by-design. One thing we have identified is that the regulatory framework set out in that legislative proposal focuses very much on the technical-security concerns but has not fully addressed the more human-centric factors in cyber security. For example, we have seen how smart technologies have been exploited by abusive family members to abuse their family members. Those concerns are not fully debated and addressed in those proposals. Going back to your previous question regarding the secure-by-design regulatory proposal, that is another thing

we really need to think about, not just the technical safety or security issues but also the human factors in the smart environment.

Q229 **Kevin Brennan:** There is a report out today by the 5Rights Foundation called "Building the digital world that young people deserve". It is setting out some of the issues. It points out there are 1 billion young people online and that basically the digital world is not optional. It is becoming not optional for any of us, but it is not optional particularly for young people. It goes on to say that effectively digital services and products should treat young people according to their age and take account of the needs of young people from the design stage. It talks about the need for the Online Harms Bill to put all this into it and make sure that regulated services under the Online Harms Bill conduct regular child-impact assessments, and so on. My colleague Damian Hinds spoke earlier about the ability to look at everything you have been doing online and look at your digital record through Google Takeout. We are having a discussion about ethics here. Should parents be able to look at that for their children? Who shall I ask? Who would like to have a go? Put your hand up if you want to have a go. Nobody. Yes, go on, Carly. Well done. You got there first.

Carly Kind: In a previous life I did some work with UNICEF on children's rights online. UNICEF's position has been that until the age of 18 parents know what is best for their child, and there is a lot of leeway for parents to make decisions about their child's life in their child's best interests. At a young age that probably does involve some monitoring of online activity, and I think there is good evidence to suggest that some parental controls can be really effective in helping parents do that. As children get older, they are entitled to more privacy, more freedom of expression, and children rely on the internet for things that their parents should not know about, for example, researching things about their sexuality. That is not something that their parents should intrude on. There is obviously a fine balance to be struck. I think parents should have some ability in their early years to be able to monitor their child's online life.

Q230 **Kevin Brennan:** Until what age, would you say? We are going to make laws about these sorts of things, so I would be interested in what your view is.

Carly Kind: Yes. The GDPR makes a distinction at the age of 16, and in the US it is the age of 12. The Children's Online Privacy Protection Rule says that children under the age of 12 should not have data collected on them and that parents have a right to consent up until the age of 12. That is slightly higher in Europe. It is very difficult to set a specific age; as every parent knows, every child is different, right? We need a framework that is going to respect that difference between children. The research tends to show that children these days are savvier than their parents, often, so it is not necessarily the case that the parents are going to be the best protectors of the child. Maybe educating the child to be the best protector of themselves is the right way forward.

Q231 **Kevin Brennan:** I would certainly agree with that. If there is a swimming pool nearby, you should teach your child how to swim, not just put up a sign saying, "Danger, swimming pool". That is absolutely correct. But possibly what you are saying is that parents need more education in this area, which we all know is true. Finally, I will ask people about the idea of a digital Bill of Rights, which is an idea that has been floated from time to time to consolidate people's rights to data and internet access. Do you think a digital Bill of Rights is needed? If so, what should be in it? Dr Chen?

Dr Chen: I think that would depend on what we are expecting from a digital rights Bill.

Q232 **Kevin Brennan:** Sorry, Dr Chen, I am not talking about a digital rights Bill. Just to be clear, a digital Bill of Rights, if you like, is a different idea. In other words, not a piece of legislation about digital rights but basically something that tells people exactly what their rights as citizens are in relation to their data in the digital world.

Dr Chen: There are already existing rules in different sectors of law that give people rights regarding their digital lives: data protection law, the consumer protection law, even competition law. The idea really is: what is the added value of having a separate digital Bill of Rights? You can argue it is good to codify all these rights into one piece of legislation so people have a clearer understanding of what sort of rights they have, but you might want to specify some of the rules that are specifically designed to tackle some of the issues exclusively in cyberspace. My point is that the biggest priority is perhaps enforcing existing rules before we have yet another separate set of rules. Enforcement of the data protection law in this country has been criticised by many scholars. It is a good idea to start discussing the details of a digital Bill of Rights but at the same time we should not let that distract us from the existing, ongoing issues around enforcing the rules we already have in place.

Carly Kind: I broadly agree. There are existing rights, they apply online, and we do not want to distract from that. We should focus on enforcing existing rules. But if there could be some type of Bill that would help people more easily understand what those are and claim them, that would not be a bad thing.

Q233 **Kevin Brennan:** Dr Tennison, do you broadly agree with the other two witnesses, or do you have a different view?

Dr Tennison: Yes, I broadly agree with it. It is a useful way of framing a conversation that highlights where those rights exist and how they conflict with each other in certain circumstances, and also to maybe identify some places where there are some gaps that could be filled through other kinds of mechanisms. But they are only one tool within the larger box that we have to put in place. Particularly, I would reiterate Dr Chen's point about enforceability of anything like this. I would also highlight that in the negotiations there might be around a Bill of Rights, we might end up with a lower bar than the bar we want to have. There

are risks in articulating those kinds of Bills of Rights, as well as potential benefits, you just have to be aware of.

Q234 **Kevin Brennan:** Thank you. Final question, where would people be most surprised to find that an algorithm is being used to make decisions about their lives?

Carly Kind: It is a good question. It is one that we have had from journalists quite a lot since the Ofqual algorithm, which is: tell us about more scandalous algorithms that are out there. The reason we do not all immediately come up with the answers for you is that there is just so much opacity about where algorithmic systems are being used. We all know about the social media ones, we know less about where they are being used by local authorities, for example, or other public sector bodies.

Q235 **Kevin Brennan:** The answer is: we do not know, but we think we should be told, basically?

Carly Kind: We wish we did, yes.

Chair: Thank you, Kevin. That concludes our session. Thank you to our three witnesses today.