HOUSE OF COMMONS

Science and Technology Committee

Oral evidence: Algorithms in decision-making, HC 351

Tuesday, 14 November 2017

Ordered by the House of Commons to be published on 14 November 2017.

Watch the meeting

Members present: Norman Lamb (Chair); Vicky Ford; Bill Grant; Darren Jones; Neil O'Brien; Stephanie Peacock; Graham Stringer; Martin Whitfield.

Questions 1 - 88

Witnesses

I: Hetan Shah, Executive Director, Royal Statistical Society; Professor Nick Jennings, Royal Academy of Engineering; Dr Adrian Weller, Turing Fellow, Alan Turing Institute; and Professor Louise Amoore, Durham University.

II: Silkie Carlo, Senior Advocacy Officer, Liberty; Dr Sandra Wachter, Lawyer and Researcher in Data Ethics, AI, and Robotics, Oxford Internet Institute; and Dr Pavel Klimov, Chair, Law Society's Technology and the Law Group.

Written evidence from witnesses:

- Royal Statistical Society
- Royal Academy of Engineering
- Professor Louise Amoore, Durham University
- The Alan Turing Institute
- Liberty
- Oxford Internet Institute

Examination of witnesses

Witnesses: Hetan Shah, Professor Nick Jennings, Dr Adrian Weller and Professor Louise Amoore.

Q1 **Chair:** You are miles away, I am afraid. We can just about spot you in the distance. Welcome. I realise some of you have had trouble getting into the building with queues and so forth. We are very grateful to you all for coming along. It would be good to start with you introducing yourselves, saying who you are and where you are from.

Dr Weller: Good morning, everyone. My name is Adrian Weller. I am a Turing Fellow at the Alan Turing Institute, where I co-ordinate a group looking at issues of fairness, transparency and privacy of algorithms. I am also a senior research fellow at the University of Cambridge, both in the machine-learning group where I work on machine-learning systems and algorithms themselves and also at the Leverhulme Centre for the Future of Intelligence, where I lead a project on trust and transparency of algorithms.

Professor Amoore: Good morning. I am Louise Amoore. I am a professor of human geography at Durham University. I also currently hold a Leverhulme major research fellowship on the ethics of algorithm.

Professor Jennings: I am Nick Jennings, a professor of artificial intelligence at Imperial College London and I am also here representing the Royal Academy of Engineering.

Hetan Shah: I am Hetan Shah, executive director of the Royal Statistical Society. We care about data and statistics in algorithms. I am not a statistician myself; I am more of a policy person, but representing our 6,000 fellows.

Q2 **Chair:** Excellent. As we ask questions, do not feel you all have to answer everything if you do not feel that you have anything particularly important to add to what others have said, but can I start off? I am conscious that basic algorithms have been around for a long time, but we are interested to understand more what is changing, how this is developing and what is new, and what the key things are on which we need to focus in our advice to Government and others. Perhaps, Dr Adrian Weller, you could start, but others come in as you want to.

Dr Weller: Certainly. We are at a time where we have seen dramatic increases in the capabilities of algorithms and they are being used in more and more important areas throughout society: areas such as making hiring decisions; making decisions on whether to grant a loan to somebody; even in the criminal justice system, perhaps, to help determine how long someone should be locked up for. Areas that may seem innocuous, such as deciding which adverts should be targeted toward us, can have an impact on us. It would be interesting to get into some of those subjects.

I should emphasise that a lot of the improvements that we have seen in algorithmic performance are due to increases in the availability of data, which is likely to continue and has privacy concerns that we might get into, and because of increased computing power. Both are likely to continue even if we do not get further improvements in the algorithms themselves. So, it is a very important time to be thinking about the societal implications of these algorithms.

Q3 **Chair:** Their use and the areas in which they can be used are likely to expand significantly.

Dr Weller: That is correct.

Q4 **Chair:** Are there other contributions?

Professor Amoore: May I say something about the specificity of the type of algorithms that we are talking about, because this is a huge area? Specifically, in thinking about the relationship to decision, the key change that I note in my research over the last 10 years or so is a shift from what we think of as rules-based algorithms—where we might identify the criteria. To put it simply, we might say the decision to give someone a mortgage could be, "If this occupation with this income, then this"—if and this, then this. The conventional way we think about algorithms historically is this series of calculations that produces a decision. That is what we are used to.

The way that is changing, with the availability of multiple forms of data and big data, and, as you say, increased computational power, is to show the algorithm large volumes of data—in something like image recognition, it could be 15 million images—and allow the algorithm to identify patterns in that data, which is called clustering. You cluster the attributes and patterns, and decide whether this person should have a mortgage based on multiple vast volumes of data patterns of other previous people to whom you have lent money.

Q5 **Chair:** To what extent is this extended approach already being deployed?

Professor Amoore: It is already being deployed quite extensively. As an explanation for that, it is a very attractive prospect, in a sense, because in all sorts of domains of life it promises to do more with less. If it is about policing decisions, it could be how you allocate a much more scarce policing resource to the frontline. You need a more finely tuned targeting system. What those machine-learning algorithms are promising, which is what we are talking about in the decision-making changing, is that they can tell you what the optimal targets are in your volume of data. They could be voters for an election or they could be possible customers for your mobile phone service. But at the level of the machine learning the problem is the same. How do you identify an optimal target? You might be targeting them with advertising a particular product or you might be targeting them with a particular political party in an election. For me, it is a huge area, but the critical point is the expansion in the use of a

particular sort of algorithm that is doing some of its own learning and that changes how we think about decisions.

Q6 **Chair:** That offers, potentially, enormous gains in the efficient use of resources, in particular, but also carries with it real risks.

Professor Amoore: Absolutely, yes.

Professor Jennings: You are right to identify that algorithms have been around for a long time. Any computer program that exists is underpinned by an algorithm. Algorithms have been around for as long as computers, and in fact many algorithms have pre-dated computers as people have thought about solving problems. The space of algorithms is vast. What is new, which is what several others have alluded to, is a particular class of algorithms that goes under the broad area of artificial intelligence or machine learning. They are a different family of algorithms, because, previously, in much of computer science you specify how you solve a problem. As Louise said, if this happens to be true, then you take this action, or if this, and so on. You specify how to solve the problem. With the machine-learning algorithms, you define how to learn about solving the problem, not actually how you are solving the problem, which makes them very much more complicated to analyse.

Having said that, the computing power that we now have available means that we can do very complicated and complex algorithms for a variety of tasks that have nothing to do with AI and machine learning, but still make decisions because we are able to have that complexity. The decision- making of algorithms is something that we are going to see increasingly more of because we have ever more data from our world, from our mobile phone devices and sensors in the environment. We cannot expect humans to process that volume of data. We will increasingly rely on algorithms to do that, and the way in which the algorithms interface and interact with people—are they an adviser and do they take the decisions for themselves?—is a really central aspect.

Hetan Shah: I would agree with everything that my colleagues have said, particularly around the importance of machine learning. As a bit of context here, we at the RSS see this as slightly akin to what happened in the debates around GM foods. We need a public licence to operate with all sorts of innovation, and I think we are in a space now with algorithms, artificial intelligence and machine learning where there is a risk that, with all the possibilities here that science is giving us, if we get the trust wrong, if we get the democratic accountability wrong, we will lose that licence to operate. It is really useful that the Committee is having this discussion about how we can open up the transparency and the accountability in this area to maintain public trust.

Q7 **Chair:** We will pursue some of those things later on, but where are the potential most positive and most transformative impacts to be seen, going back to your points, Louise, about the potential for this in the efficient allocation of resources and so forth? Let me have some thoughts

on that, and then Nick and others.

Professor Amoore: In my research I have also been looking at medicine and health. You can see in the relationship between machine learning and decisions there, for example, in an area such as oncology, again that problem of how you anticipate ahead what the optimal treatment pathway for a particular sort of tumour might be. With cloud computing, with the availability of data and great expansion of the capacity to analyse that data, the surgeon and the oncologist are able to make a decision with the support of a tool that allows them to look at what were the outcomes for many tens of thousands of other past cases exceeding their own hospital, their own country, their own research group, and thinking transnationally. In a sense, what we are talking about is also situated in the benefits of big science. This is not a question of saying that these are good or bad things, that there are evil or good uses to which this is put, but it is essential that we understand the implications of that for the areas that we value in society.

To go back to those oncologists, for example, they would say the algorithm is giving them a score. At the level of the computer science, it is really between 0 and 1. It is a probability that this particular treatment pathway has an 85% chance of being more successful based on what we have in the past data. Some of those oncologists express concern because they say they are not able to open the model to see how it arrived at that. For the hospital, the benefits are risk management, in a sense, because you say, "We have this availability of data; we can say that these seem to be the optimal pathways; these are the decisions that you ought to take." But in the research that we are doing we are noticing that that does potentially make it more difficult for a human being to make a decision against the grain of the algorithm, if that makes sense.

We have seen it for a long time in the UK, even, for example, in algorithms in border controls. Can the border guard decide something that goes against the risk score? The risk score is just the output of the algorithm, and if you adjusted the algorithm the risk score for the individual would be different. In thinking about algorithms and decisions, we know that there are benefits, but it is a question of thinking about what the implications of those might be in terms of the broader benefits and costs for our society. Yes, it goes across different areas. It is not that we would want to say we should use them here and not here, but we need to understand the context and think about how the model is producing its output. We might come to that in a moment perhaps.

Professor Jennings: One great power of this technology is that it is application agnostic. You can apply these sorts of algorithms to any aspect of society. You can pick your favourite, most preferred area, which might be education, health, social benefit, banking or advertising, and all those can be impacted in a positive or potentially negative way by this class of algorithm. They are able to help personalise things to you as an individual and to support decision-makers if done properly. You can really

get benefit from these, but, if they are not used properly or they are not developed in a proper way, then there are also risks in relying on algorithms that are not fit for their particular purpose. They are entirely generic in their technology and the benefit around which we will see applications.

Dr Weller: If I may take up the point of your question, we should not lose track of the fact that there are tremendous opportunities here, which is what you are asking about. There are great possibilities to improve many areas of our lives, not only to personalise aspects of our lives but also dramatically to improve efficiency. We have talked about healthcare, as Louise has mentioned, where there are many exciting potential areas where you may be able to gather data across many individuals and thereby see patterns that allow us to solve diseases that we would not have managed to solve before.

Q8 **Chair:** Sepsis I hear talked about as such an area; 40,000-plus people die of sepsis every year, and AI and algorithms could be used potentially to reduce that death toll.

Dr Weller: That is absolutely right, or even perhaps cancer or other diseases, but to do that will require gathering data, which has privacy concerns. But still there are tremendous opportunities, also to improve efficiency of lots of things such as transport or inventory management. While we think about the possible concerns, which are very important and we are going to discuss them, we should also bear in mind—the example of a border guard is a good one—that, while we know that human beings can exercise some common sense and wise, broad judgment when analysing a problem, which is difficult for algorithms to do currently, we should also note that humans are sometimes themselves a little bit arbitrary in the way they make decisions and can be somewhat biased, and also can be very difficult to understand. So, there are many advantages as well as concerns.

Hetan Shah: Following on from colleagues, there are lots of positive examples being used right now. In international development, people are using satellite imagery to look at new ways of measuring development and to find sites of modern-day slavery. There are drones being used to stop elephant poaching in some African countries and so on, and this is using machine learning right now for very positive humanitarian purposes. We are going to see the rise of autonomous vehicles, which is going to change the landscape probably in quite positive ways with regard to road deaths and so on. It is best to understand this as a ubiquitous technology and to think of it almost as a public infrastructure in many respects. From a policy perspective, that is the approach to take.

Q9 **Chair:** Designing the framework within which this develops is the critical thing.

Hetan Shah: Absolutely.

Q10 **Darren Jones:** I want to try to dig in a bit to the issues of bias and fairness, and let us take them in turn. On the question of bias, with the more traditional category of algorithms you have described, it is easy to understand whether you might have human bias because you are deciding what the if, and, but and the equals are. When you are putting in the factors, if you have bias about what types of outcomes you want, that is easy to understand. I struggle a bit where these algorithms are machine learning and taking decisions for themselves based on data. Where, if at all, does bias come in? Is it in the dataset; is it in the way in which you design the algorithm; or is it in fact in the way that the algorithm then makes its decisions while it is machine learning; or is it all three?

Professor Amoore: It is all those things. Let me begin with a very brief but dramatic example, which might help to clarify things. There was a very well-popularised case recently of a young black researcher at MIT, Joy Buolamwini, who is a software engineer working with facial-recognition algorithms. She realised that the most widely used algorithms did not recognise her black face and so she made a white mask. Her point was to show that these algorithms had learned through a dataset—what we call training data. They had been trained to identify the patterns in a facial geometry using predominantly white faces. It is an obvious example, but at the point of the training data it is a really important moment where bias can enter the algorithm.

The point I would like to make is that bias, in effect, for me is intrinsic to the algorithm. Let us put aside for a moment what we think as bias in law or society, where we would say bias is prejudice; it is giving favour to one thing or one person over another. In a sense, in order for an algorithm to operate, it has to give weight to some pieces of information over others. The way that a neural network learns is that each layer—which they call a hidden layer—of the neural network passes its data, or its calculation, on to the next layer. It has to give a signal of what weight to give that. It is a probability. Every layer has a probability that this piece of information should be weighted more heavily than the other.

Many algorithm designers have described to me that bias for them is a tool that they use in the algorithm itself. It will always weight some things, some people, some images, some pieces of voice for voice-recognition software, more than others. If we begin from that point of saying, yes, the algorithm operates with something that we might call bias, that would probably help us to understand what to do about it and how to think about where it can have a position in our society. I am concerned by the notion of being able to extract bias—take it out and take it away. In all of the areas that I have looked at, you could not have an unbiased algorithm.

Q11 **Darren Jones:** Building on that point, I can see, for example, how, if you were doing targeted advertising, you might wish to bias certain demographics of people; that makes sense. However, if, for example, we

wanted to update the Equality Act to deal with discrimination in algorithms, how would we ensure that people who are designing these algorithms, when they are making these weightings, are not weighting them against the law on issues of gender and decisions on promotions or pay rises?

Professor Amoore: It comes right back to the problem of the notion that we do not begin with rules but that the algorithms themselves generate rules. Research has been done in the United States in the area of algorithms used to optimise or predict the outcome in a trial—custodial versus non-custodial sentences—that showed that, even where race as a category was removed from the input data, the algorithm still learned characteristics, or attributes, that we might say are in breach of the Equality Act, because they use what we could call proxies for race. They learn patterns in past patterns of crime or they learn patterns in postcodes, for example, or even now, with increasing so-called open source data, social media data, Twitter and Facebook data, they learn patterns in that sort of data that we might say has had a racialised outcome in the algorithm.

To summarise, it is extraordinarily difficult to imagine a framework for software designers to remove the bias at that point. We should begin from the position of saying it is there, and then identify what we do about it, how we respond to it and how we regulate it.

Professor Jennings: I would slightly contradict Louise here. As a computer scientist, I do not think algorithms that are written are inherently biased. A machine-learning algorithm will learn according to a set of functions that you give it. There are many flavours of machine-learning algorithm: some of them are neural networks and some are different sorts of networks. That will learn according to what you want to prioritise and what you, as the designer of the system, want it to learn—so what you reward. That is intrinsically unbiased.

When you have trained it on data, and if you have done that in a poor way, then the instantiation of the algorithm—the algorithm after it has learned and been trained—may well give biased decisions if you have given it biased data, but the actual machine-learning algorithm is not inherently biased toward anything. It is trying to identify and maximise some particular function according to what you are going to do.

The challenge for the algorithms as to bias is that they are not always well trained because people do not always understand exactly how they work or what is involved in training. The training of a machine-learning algorithm well, effectively and in an unbiased way is still very much part of research in this area. It is not well known. There are some guidelines and good practice, but lots of people are not necessarily aware of that or necessarily inherently aware of what the issues are with bias. Therefore, you end up with poorly trained algorithms giving biased results.

My final point is that a key part of machine-learning algorithms, both now and in the future, is to make sure that they are able to give a degree of confidence about the sort of outcomes that they are saying. It should not be just, "Here is the answer." You want it to be able to say, "Here is the answer. I am not very confident in this," or "I am really confident in this." That is a key part of that interplay and interpolation of the algorithm.

Hetan Shah: An interesting area is recruitment, where companies are increasingly using algorithms to decide whom they should hire. The training data is really key. If you say, "Here are all my best people right now, and can you get me more of those?" and then you get lots of old white men, don't be too surprised because that is what was contained in your original dataset. Interestingly, you can use algorithms to flip things the other way round. If the goal is that you want to try to screen out human bias, which we know is present in recruitment practices, the behavioural insights unit—the nudge unit—has set up a completely new algorithm called Applied, which randomises the order in which you see questions; you never see the candidate name; you do not know their gender and so on; and through all these things the algorithm is helping you to recruit in a less biased way, as it were.

The intent is firmly built in from the outset, thinking about the training data, but all these things can be used in either direction. Nick's point that the community of data scientists may not yet be very aware of this is a really important one. Not only are there regulatory matters to think about, but there are codes of conduct, training and ethics that can be built in. The community of people working in this space is relatively small, so we do have an opportunity to get in there from the start right now.

Dr Weller: My colleagues all make excellent points and I fully support everything in particular that Hetan just said, but, going back to the original question, first, we need to be very careful about what we mean by bias. We have been using it in different ways here. Do we mean unfair discrimination against certain kinds of people, or do we mean that a model is more likely to pick a certain kind of person than another? These are quite different things. If we train an algorithm to try to detect a terrorist, you could say that it is biased toward detecting terrorists, but that would be an odd use of the word. What we typically are concerned about here is algorithms that are unfairly picking on certain types of people, in the sense of discrimination, and that is what I think we should try to focus on. Using the word bias is sometimes a bit confusing.

We have had some discussion around how this can happen with machine-learning algorithms that are learning from training data, but, just to be clear, it can also happen without any training data. It can happen with those old-fashioned expert AI systems. I might design a system with the best of intentions thinking that it was going to pick out the best people to hire and either intentionally or unintentionally it might be discriminatory in hiring certain types of people who happen to just fit in with the kind of person I was thinking about at the time. It is not only

a feature of data. It is also about who is building the algorithm, how they are doing it and what they have in mind. You can get unintended bias or intentional bias even without any data. This is one reason why we have seen many people call for having a diverse set of people involved in the industry, which is very important. We need to have representations across society in designing these algorithms.

I have just a few other points, if I may. One is that, as Hetan mentioned, there are different kinds of problems you can get when you are learning from data. If you are looking at hiring data, as he quite correctly points out, the only truth you can look at when you are training your algorithm to try to do the right thing is to look at what people in the past have done. People in the past might well have behaved in a discriminatory fashion and so it is going to learn that pattern. You could then try to examine the extent to which your algorithm is following that and try to remove it, and that is interesting, but to some extent you are a bit stuck there.

The next kind of data I want to mention is where there are some subtle biases that may be difficult to foresee. I am thinking here about banks making loans. When a bank is using an algorithm, typically it is thinking that it wants to have some threshold level of certainty that an individual is going to be able to repay a loan. Perhaps it wants at least a 98% chance that someone is likely to repay this loan. They will train their algorithm based on looking for people who have at least this probability. When they do that, if they happen to be looking at a particular person who comes from a demographic where there is not much data, perhaps because there are not many people of that particular racial background in a certain area, they will not be able to get sufficient certainty. That person might be an excellent risk, but they just cannot assess it because they do not have the data. This is sometimes called a thin-data problem, and it is an example of a subtle problem that can come about in data. Also, I want to say that there are other sorts of data that—

Q12 **Chair:** Of course, this problem exists now, as we have said, without the use of any of this technology.

Dr Weller: This is exactly right. Then there are other kinds of data where we can make great progress because there is a kind of ground truth of reality from which you can learn. If you are thinking about insurance companies deciding how much premium they should charge to different sorts of people for different kinds of activities, they are very well motivated in the commercial world to try to assess that probability precisely. There is data on what actually happens subsequently, to which they are trying to train. They are not only training to copy what a person has judged to be sensible but they get the data of what actually happens. That automatically pushes them in the direction of being less biased.

Q13 **Darren Jones:** Very briefly—we have already touched upon it in using different language—I want to talk about this question of fairness. If we look at non-algorithmic situations, we obviously have regulators that

regulate certain industries where we think there needs to be a level of fairness for consumers. I have worked in the energy and telecommunications sector. I had to deal with Ofgem and Ofcom in the corporate world. They would set rules that businesses would have to apply and abide by, and if there were concerns around outcomes they had the power to intervene and to ask for information and explanation, and we would either agree to change things or they would force that upon us. Do you think that model works in an algorithm situation?

Hetan Shah: I think an outcomes-focused framework is a very good one, because, particularly with machine-learning algorithms, trying to peer into the black box is quite opaque. It may be possible; there is a debate in that space, and making things more transparent may help. But if you are in a regulatory situation, where you can just monitor outcomes and then point back, without having to look into the black box, that is an extremely helpful way of doing things. In a sense, there are discussions about fairness, but, of course, we have different notions of fairness in different contexts, so the clearer a regulator or other body can specify what that metric is in an outcomes-based way is very helpful.

Professor Jennings: As to regulation for it, my observational thought is that it should fit within existing sectors in regulation. If the algorithm recommends an energy tariff to you, it is relatively straightforward to take the usage data of your own electricity, learn that pattern and then for an algorithm to make a recommendation to you, which says, "The best tariff for you is the following." If there are biases in that, that should be done through the energy regulator. I do not see much sense in having an algorithm regulator that would look at algorithmic decision-making, because it would be forever running up against whichever domain you are talking about. An algorithmic regulator who talks about unfairness in energy would end up having to talk with the energy regulator all the time. It is in sectors. Algorithms get deployed and used in sectors, so that is where I think the regulation should come from.

Professor Amoore: We might disagree again. I think you said it was application agnostic, and I completely agree with you. At the level of the design of the algorithms and the market, in marketing IBM Watson or a particular approach to analytics, there is a desire for it to be mission agnostic or sector agnostic, because that is the way to maximise a return for those algorithms. I have looked at this in a lot of detail in terms of how a particular algorithm designed in one sector travels to another sector. For example, with an algorithm that was designed specifically for casino fraud, you might say there is a lot of very relevant data to train your neural network for casino fraud because you can look at transactions and you can model what normal and abnormal transactions might look like, but when that algorithm moves out of that space into another, identifying somebody who is more likely to commit a future crime, it means that the algorithm itself is, as you say, application agnostic. It is very important that we think about questions of regulation and oversight

across sectors because of the way that the algorithms themselves are travelling across sectors and being applied in new domains all the time.

Q14 **Chair:** Are we seeing plenty of that already?

Professor Amoore: Yes, I think so.

Professor Jennings: A lot of the algorithm machine-learning frameworks are taken from one application area and used in others.

Dr Weller: On this point, I agree with Louise. We are seeing these algorithms applied broadly across society. There are challenges in how to think about these issues related to rapid progress in data science and artificial intelligence, which are likely to continue, and I think we in the technical community need to play our part to help communicate effectively what is going on so that we can help to form rules that are legally and technically feasible. But it is moving forward quickly, and I think there is good sense in having an overall stewardship body to help advise on these issues.

I have one quick comment on fairness. Fairness overall is something we all like, but it means slightly different things to different people. People in the technical community get very excited when thinking about how exactly we can define it: well, it must mean we want equality of this particular thing. Now that is good because we can optimise our algorithm with the constraint that we want to enforce equality of this particular thing. You do get that, and of course that is great, because, yes, you can look at the outcome and you can achieve what you are asking for, but that might not represent the broad-based sense of fairness that we might want as a wider community, so it is important to keep discussing these issues.

Chair: I am conscious that time is pressing on. If you are okay, Darren, we will turn to Vicky.

Q15 **Vicky Ford:** My questions are about transparency, because lots of the written submissions we had stressed that the algorithms should be operated in a transparent way. Is that feasible, especially when the algorithm itself is learning and changing?

Professor Amoore: I will be bold and say no, I do not think so. The algorithm designers that I interview and follow in the development of specific algorithms that they are going to put out into the world consistently say that the model is in part opaque to them. I am going to explain what I mean by that. Let us put it in a context and say that somebody is working on a particular sort of neural network that a Government—not the UK Government—is going to use for detecting possible immigration infringement. The person designing the neural network is watching the output signal for the algorithm change over time. They have a target output that would be to identify a certain number of people at a certain time of day, for example, at the border. They want to know how they can optimise the output to the algorithm. They do that by

adjusting the weightings within a particular layer of their algorithm, and they can explain all of that and can show how that works. What they cannot do is explain how that shifts.

In a very well-known algorithm such as AlexNet, which is used all the time for image recognition online—"Is this a dog, a cat or a human face or not?"— there are 60 million parameters. I often remind myself of that when I see the idea of transparency. What would it mean to make 60 million parameters for a single algorithm transparent? Would that be useful to us, to legislators, or would it be useful in a court of law? What would we do with that transparency if we could have it? If we could see those 60 million parameters in a short space of time, what would we do with them? What is needed is not transparency but accountability. We should be trying to hold them to account and really push on this notion of accountability. They must be held to account, but we should begin from the position of saying that transparency is not possible and is probably not desirable.

Hetan Shah: Transparency is a slightly slippery word. I would support that on a narrow definition of transparency. In New York City at the moment there is a Bill going through asking for all city agencies to publish the source code of their algorithms. That will be an interesting thing. In fact, the courts have just forced them to publish some of their DNA evidence algorithms. Even on a narrow transparency basis, people think there might be some use there.

It might be worth looking at credit-scoring agencies, which is a mature market that has been using algorithms, and some people in those spaces say that they think they can explain their algorithms, but I am a sceptic.

Slightly wider than narrow transparency would be explainability. Can the person tell you why the algorithm made its decision? Probably why is too hard, but there is an approach, which in fact one of the members of your next panel has supported, which is, what would it take to have changed the decision? What is the counterfactual, as it were? If it was a man applying for the job rather than a woman, and you would have given it to them when you would not have given it to them when it was a woman, that tells you something about the algorithm. That is different mode of transparency.

There is a wider transparency that is useful, which is, "Can you show us what training data you used in the first place to train these things on?" This is one area where I would ask the Committee to give power to the elbow of public sector organisations that hold datasets. One example is that the Royal Free Hospital gave a lot of data to Google DeepMind. It was rapped on the knuckles by the Information Commissioner about this. My view on this is that, clearly, they were trying to do useful stuff, but they were seduced by the magic of the algorithm company. What they did not realise is they were the ones with the really important thing, which is the dataset. Over time, you are going to see more private sector

providers springing up who can provide algorithms, but the public sector have the magic dataset, on which they have a monopoly. When they are transacting with the private sector, they should have more confidence and should not get tied up into exclusivity contracts, and they should ask for greater transparency from the private sector providers to say, "Open up so that you can show people what is going on with this evidence."

Q16 **Chair:** The data held by the public sector has value as well.

Hetan Shah: It has extraordinary value and it will increase over time. As is often the case, the public sector has a slight lack of confidence in this area and thinks the magic lies with the private sector.

Q17 **Vicky Ford:** I am sorry, but I am not understanding clearly enough. I understand that having 60 million different datasets or bits of information going into facial recognition is not particularly useful. I also hear you say that being transparent about the dataset might be very useful. What was your point about New York City?

Hetan Shah: There is a Bill. I was just pointing to the fact that this is something that regulatory authorities are thinking about. It is not a settled matter and they have said, "Publish all your source codes for all algorithms that"—

Q18 **Vicky Ford:** Even though that might be the 60 million.

Hetan Shah: Yes.

Q19 **Vicky Ford:** Then what we hear from, say, Google in their evidence is that, even if we were completely transparent, it would allow spammers to act in a way that undermines the product. Is there truth in what they are saying? Does it undermine the algorithm if we force it to be public?

Professor Jennings: I will come to that. As to transparency, it is impossible to say that all algorithms will be transparent and understandable—agreed because of the large parameter space. It is possible to say some algorithms can be made transparent and understandable, so there is not a blanket answer to that.

As to misleading algorithms, in terms of getting them to learn the wrong thing, there is a term for that, which is adversarial machine learning, which is where you know the way a machine-learning algorithm works and so you try to dupe it to believe something and come to a particular set of conclusions; then you can exploit the fact that you know that it has been mis-trained. That is an area of study these days. So there is some truth to that. It is quite hard to do, depending on the size of the dataset that you have, but it is a plausible thing.

I would like to echo what Hetan said about the value of the data that Government held. To my mind, as a machine-learning and AI researcher, the machine-learning algorithms are not the clever part of this. Their application to the data is the really important thing to have here.

Q20 **Vicky Ford:** Who should be making a decision about transparency? Should it be a regulator? Who should be responsible for deciding whether the company that has tried to adapt its algorithms should be making it public or not?

Professor Jennings: If you want to have a degree of consistency, you cannot leave it to the companies, because they will come up with their own particular bespoke solutions that work well for them.

Q21 **Vicky Ford:** The data protection commissioner, for example, has said that she thinks they could play a bigger role in the development here because of the interface there.

Professor Jennings: Being clearer about what it is being trained on and what are the limits of the algorithms are the crucial bits to get and to understand here.

Q22 **Vicky Ford:** Let us take the example we were given that women were being shown fewer instances of ads for high-paid jobs than males when they were looking online. Clearly, if it was you or me choosing them, we would be breaching equality legislation. Is the company that derived the algorithm responsible if the algorithm breaches competition law or equality legislation? Who should be responsible?

Professor Jennings: The people who are using the algorithm. The algorithm might be an open-source general piece of machine-learning algorithm that anyone can download. They might have taken that algorithm and put it out for everyone to use. The people who are devising the adverts have then trained it to deliver particular behaviour. It is not the original developer of the machine-learning algorithm because that could be an arbitrary researcher who has said, "Here is a machine-learning algorithm. You can use it and download it for free." It is freely available and anyone can take it. It cannot be them. It has to be the person and the organisation who have trained it and are making those recommendations.

Dr Weller: In many cases I would agree, but we need to be careful. One point on which I would disagree with Nick is that I do not know that you can have blanket rules across everybody. There are very often trade-offs among issues of transparency, privacy, performance and other issues, and it is very important to examine context to determine what should be done in each specific situation. We are just at the beginning of trying to establish overarching principles. I am involved in trying to help form workshops to analyse these issues in specific domains and am very happy to talk more about that if you like.

Q23 **Vicky Ford:** Do you feel as if you are acting in a bit of a policy vacuum at the moment of not knowing who is taking responsibility and making decisions on transparency?

Dr Weller: Yes, but I do not feel that is an urgent problem. It is urgent that we should think about how to start tackling these problems, but it is

a bit too soon to say, "This is the solution." There are a lot of subtleties here, and we need to bring communities together and think carefully about how we want to ensure that we get good outcomes for everyone broadly across society. That may depend on the specific application. As to the example you give with ads, there are clearly issues about the extent to which internet platforms are platforms or whether they need to be held responsible for what they are showing. There have been long histories of newspapers taking adverts targeted at specific people, and generally we have thought that that is okay. At what point is it not okay? These are somewhat subtle and difficult issues that we need to discuss carefully.

Q24 **Chair:** You say it is a bit too early to have solutions, but there is also an urgency in that this is developing rapidly, is it not? So, we need reasonably soon, one would imagine, to create the framework that does address these issues?

Dr Weller: Absolutely. I think it is urgent to form a framework or an advisory group or some body that is urgently going to think about these issues and make some recommendations. I do not think we are ready to give the solutions, if that makes sense. Can I take two minutes to add something?

Chair: One minute.

Dr Weller: We have heard about different kinds of transparency. We have heard about accountability, which in some settings is very important. We have heard about contestability, which in some settings is very important. If you have an algorithm that tells me that I should be locked up for five years, I would like to understand how they came to that conclusion in a way that allows me to disagree with it. That also has an aspect of justification. You also need to be careful because, if you demand an explanation from some entity, they might provide an explanation that is misleading—a word that Nick used before in a different setting.

A well-known book seller recommends products to us online and it used to be that you could ask it why it gave that explanation to you. I think they have stopped doing that now. If you think about it, when they give you an explanation, they usually say it is because you liked some other book and you bought some other product, but we know that they are doing something much cleverer than that underneath the hood. The reason they are giving you for that explanation is really optimising it in order to get you to click through and buy the book. It is not a faithful representation of what it is doing. That is another issue that needs to be thought about carefully.

Hetan Shah: On your point about where the regulatory framework needs to go, with the GDPR coming in—the General Data Protection Regulation— and the Data Protection Bill, data protection does give a useful framework. We need a strong Information Commissioner for that.

Case law will help, so where we have tried to regulate new technologies in the past—

Q25 **Chair:** Case law is haphazard.

Hetan Shah: It is, but it will evolve over time. There are already other regulations, equality laws and so on, which will develop. We are also involved with a new deliberative council that the Nuffield Foundation—an independent grant-making body—is setting up, which is to start thinking about issues around privacy and consent, and so on, and it will be able to start making recommendations in this area.

The final point is that I agree with what Nick said earlier that a sectoral approach is really important. While bodies such as the Information Commissioner will be able to look across the piece, I would recommend that you, as Chair of the Committee, write to all the major regulators and ask what they are doing to think about this right now, because, at board level, everyone should be thinking about it.

Chair: That is interesting, yes.

Q26 **Vicky Ford:** That is really helpful. Finally, you mentioned New York. I know the European Commission has looked at anticompetitive issues in algorithms. Is anybody else looking at this?

Dr Weller: Bulgaria passed a law last year in a similar vein to the one mentioned about New York, so it will be very good to examine what is happening there to make open-source algorithms used by Government.

Chair: Bulgaria.

Dr Weller: Yes.

Q27 **Graham Stringer:** I think you have partly or mainly answered the questions I was going to ask when Vicky asked questions, but if something goes wrong with a system who should be accountable? Should it be the designer or the operator? How different is that if it is a self-learning algorithm?

Professor Amoore: The question of accountability is the nub of the issue. Perhaps it might be helpful to think about examples where things have gone wrong. In the health system, for example, in experiments using algorithms for triage, in the case of identifying particularly high-risk cases of pneumonia, the neural network learned in the data that underlying asthma was correlated with positive outcomes. Of course, it learned that because the hospitals were treating the people who were arriving with asthma as urgent cases. I am only using that to illustrate it because it emphasises the problem that this is not about causality; it is about correlation. That is different. The nub of trying to push on the accountability is to ask questions of how something could have been decided otherwise and whether that correlation really is a sound basis.

While I agree with this notion of an independent body, several years ago one of my research teams was in the national border targeting centre when the data on citizens was being analysed for border risk, and the Information Commissioner's Office was also going into this place in Manchester where this was happening. But at that point they only really had oversight of the data—how the personal data was being used.

My slight concern about our discussions in regulation or oversight is whether it is enough to bolt this on to existing data regulations or the GDPR, because we are also talking about an architecture or infrastructure of calculation and not just a question of data. At that point, the ICO was only able to ask questions about how the data was being used and not the form of analysis that was taking place, which is what my researchers observed.

There is currently a gap. There is a deficit there in what should happen, but there is a question also of legislation in terms of the public good. We may say that there are areas of our social, political or economic lives where we might want to say there is no place for algorithmic decision-making. In extremis, you might wish to say that there is no place for inference or correlation in the criminal justice system.

Q28 **Chair:** There are plenty of existing biases in the criminal justice system, are there not?

Professor Amoore: Yes, there are, but when you have an existing bias around a piece of material evidence, even DNA, or a photograph, or a CCTV image, or the evidence that has been given by an eye witness, there is always the possibility of this cross-examination and the questioning: "How did you arrive at that judgment?" Together with my researchers in the past, we have worked with barristers representing people in trials where part of their case has hinged upon an algorithm's arrangement of the risk that this person may have been involved in this particular sort of activity. That was arrived at through the behavioural profiles and models built on past people and not that individual.

Chair: It is very dangerous.

Professor Amoore: Those barristers were asking us whether we could explain to them the broad logic of how these sorts of algorithms profiled their client's patterns to match with these high-risk patterns. As to holding to account, there is also this question of how—

Q29 **Graham Stringer:** Can I take you back to your example? If the algorithm triages somebody and the patient dies when they otherwise would not have died, the coroner's court will want to place responsibility. Will the responsibility in that situation that you gave us be the designer of the algorithm or the medics who followed the algorithm? That is the key to accountability, is it not?

Professor Amoore: Yes. That is the nub of the issue. Referring back to something that Nick said about "off the shelf"—that this could be an

algorithm off the shelf; it could be openly available—that is the problem with this notion that we should train designers with a more ethical sensibility or even the idea that you publish source code. In a way, the horse has already bolted. These things are already freely available and almost untraceable as to where they are. Holding the designer of the algorithm to account in that way will always be an impossibility, even if we thought in terms of something like corporate responsibility, where we imagine the notion of the firm as being responsible for corporate manslaughter. We have examples in law of something other than a human being being held to account for a decision, and I think this is exactly the nub of the issue. It is acknowledging the extent of algorithmic adjudication in our lives and saying that this means that our existing rights to privacy, for example, or to freedom of assembly, and lots of other already-enshrined rights, are inadequate to deal with the consequences of some of these things we are looking at.

Professor Jennings: There is not a simple answer to this, which is probably what you might have guessed in the first place. This is true for all software systems, whether or not they do machine learning. If you develop a software system for some form of life-critical system, it need not have any learning in it. A software engineer has designed it and a company has delivered it. You have an interplay between whether the software has done what it said it was going to do—that it is possible to explore and investigate from the development side—and then it has been used in a particular context, and it may have been used correctly or it may have been used incorrectly. You can end up with different answers to these depending on your circumstances. As I say, that is true, independent of whether the algorithm is learning or not. It is a general issue for all software systems.

Hetan Shah: I reinforce the earlier point, which is that, therefore, you need to think about this regulator by regulator and sector by sector, and what the standards are that currently apply for accountability within that sector. It may not look all that different, but you should be asking each of these regulators whether they have thought about this question.

Q30 **Graham Stringer:** That would be a good question. Generally, how would one of the regulators prioritise between the different algorithms that there are out there so that potentially those with the biggest impact were dealt with first?

Dr Weller: It is a great question on the issues of accountability. They are terribly important. I agree that they very likely need to be examined case by case. We do have an existing legal framework for dealing with situations where you need to assign accountability appropriately. For example, if you have a car and the brakes go, is it the car manufacturer or the brake manufacturer, and how is that dealt with? What is a bit new here is that, by their construction, these algorithms often are learning from data themselves and doing things that are very difficult foresee. Foreseeability is a challenge. We may want to assign strict liability in

certain settings, but it is going to require careful thought to make sure that the right incentives are in place to lead to the best outcome for society.

Chair: That is fascinating; thank you very much.

Q31 **Martin Whitfield:** We have the algorithm, which, in essence, mines the data. My interest is in that massive dataset that sits underneath—and the dangers, the opportunities and the value of that dataset. You talked about training data and you talked about the risk of bias being picked up by the algorithm from bias in the dataset. Could you just expound or explain a little more about the difference between a training set of data and the whole set of data that the training set has been extracted from?

Dr Weller: One difficulty is that perhaps the training data you happen to have is not representative of the entire test set you are going to go out to examine. The assumption that it will be is what is very typically assumed in academia when we are building our algorithms, yet in the real world it is often not the case. As one example, we want autonomous vehicles that will perform well no matter what they come across in the real world, and of course they are bound to see situations that they have not seen before. That presents all kinds of interesting challenges.

Coming back to an example we talked about before, but a slight variant of it, we heard about a situation where people with dark skin were perhaps not treated fairly. There was another well-known example a few years ago where an algorithm that had been trained by Google to recognise images was recognising people with dark skin as being a type of gorilla. That was because the training data they would look at did not have many people with that kind of face. You can unintentionally have that kind of effect because your training data does not capture all the effects you would like to consider.

One other issue about data that we have touched upon—and I will mention it quickly—is that the data in many cases powers your algorithm. It is a key component. We are seeing in many areas that big companies are releasing the source code of their algorithms because they can do that freely. They know that really the power is because they have all that data. There are powerful network effects to this data. Once you accumulate a lot of this data, it can be very difficult for new entrants to come in, and we have already heard talk about people saying it would be great if we could try to make datasets publicly available, while appropriately respecting privacy.

Hetan Shah: Sometimes if you are training an algorithm and that is not your area of expertise, you will use a code library, as it were, and so, if we could improve the code libraries that people are tending to use so that they are more diverse, that would be very helpful.

This is not in quite the same space, but I saw a paper in *Nature* the other days saying of science datasets in general that only 3.5% of them have

African and Hispanic data in genetic datasets. So, if we are thinking about healthcare, for example, how might this start distorting things?

A final example is on language, which is an interesting area. I typed into Google Translate, "She is an engineer. He is a teacher," translated that into Turkish, which is gender neutral, and then translated it back, and it came back as, "He is an engineer. She is a teacher." Even the way that language is assigned with an assumption about gender is really interesting in this.

Q32 **Martin Whitfield:** To take that point slightly further, in a sense we have a need for the transparency of the training dataset, and are we within the foreseeable future able to assess the value of the algorithm from the training dataset that has been used and could there be criticism given to an algorithm because of the training dataset chosen?

Dr Weller: Certainly, yes. Ideally, you would have access to the training data, but very often there will be difficulties in doing that for privacy reasons or because of the proprietary nature of the data. There are methods that people are starting to develop, and I will give a very quick overview of how people are approaching interpretability. One is that you restrict your model class to very simple models that might perform well enough in your area and in which you can understand your own interpretable, so that is great. If you are beyond that and need to use a complex model, you are then in the domain of trying to use different algorithms that are going to be trained to try to explain what the first algorithm is doing. Many of these algorithms will do things such as perturb the inputs to the initial algorithm and tell you the sensitivity to different features. That is very useful. You might be able to ask questions, as you will hear about later, about what is the nearest test point, which would give a different answer.

Another approach is to look at the training data. This is recent ongoing work, an emerging field, but there are people who can help to tell you that the reason why you are getting this particular answer for this particular test data point is mostly because of this particular training point, and then you could try to request that. You may not have access to all of them, but you could try to focus on certain parts of the training data.

Q33 **Martin Whitfield:** Because time is pressing, to go back to that concept, we heard in the genomics inquiry about the value of the NHS data and the huge power that they have because they hold this data, albeit with the protections and things like that. Would you like the opportunity to expound on that element of it—the value of the data holder over the algorithmic designer, and how we should leverage that, particularly if we look at the NHS, given our almost unique set of mass data that is held?

Hetan Shah: I would not say much more than I said earlier, which is that the public sector really needs to recognise that it has some power in those negotiations. As you say, the NHS datasets are unique and very

powerful, but even administrative datasets that we hold in DWP, HMRC and so on are very powerful, and the algorithmic providers cannot do anything in public policy without the datasets. As I say, there will be more and more of these providers, so let us make sure that we set high-quality standards and that, when we are doing the procurement from the public sector, we recognise the position of power that we are in.

Professor Amoore: I would not want us to leave the room with the impression that the data is entirely separate from the algorithm in this instance. Many of the cases that are perhaps most pressing in society can be the same organisation. For example, there is the ongoing ICO inquiry into Cambridge Analytica and thinking about the way, almost, algorithms end to end identify who might have the propensity to be a swing voter, and to identify and target those individuals for a particular tailored news feed, for example, or for their Facebook feed to read a particular way. In many cases, this continual testing and refining of the algorithm is an ongoing iterative process. It is not only the case that the model is trained and then it goes out into the world. It is continually modified and adjusted in the field in practice, and that happens in lots of different areas, not only in the commercial area of advertising. We need to take into account that the algorithm is continually refined and adjusted in line with its almost real-time exposure to new data and new things.

Q34 **Martin Whitfield:** Is that not the proof of the need for the transparency of the dataset that it is using—not just its training data but its real-life, real-time data analysis?

Professor Amoore: Yes, but I suppose the question is how you get at that real-life process. Let us make that real and say that somebody is training a particular sort of machine-learning algorithm for a particular purpose. Those people who are concerned with the computer science will often talk about how they go to speak to the operations team, or for the client, if you like, about what is optimal in this space. What is useful? Is this useful?

There is some useful information in the submission that the Durham constabulary made in their Hart algorithm, because they do disclose what the training data was. They name the algorithm. They say it is a random forest algorithm, which may seem like an abstract thing, but the GDPR itself talks about an informed expert, who would be able to explain what a random forest algorithm is. I am sure we all know what it is. But that is useful information in making judgments about the effect that that decision-making tool that Durham uses has on people on the ground.

They also talk in their submission about how they have to adjust the parameters. They realise that a false negative could be a very risky thing, because that would mean not arresting somebody or keeping them in custody when they potentially had a high risk of reoffending. They talk about how they adjust the parameters in their algorithm, which means they tolerate a few more false positives—people who perhaps should not

have been kept in custody for an extra 24 hours, let us say. That is interesting.

They are even talking about how their own operational needs adjust the parameters in the algorithm. I want us to recognise that it is not just the algorithm and the decision. It is an iterative process, back and forth, with new data, with the test data, with background and foreground data, which is explained very well in the submission of the Institute of Mathematics that you have had as well.

Q35 **Chair:** Louise, did you question Durham police's use? You talked about sectors where it may just be inappropriate to use algorithms, criminal justice being one of them.

Professor Amoore: Again, it is not a question of saying it is good in this case or it is not good in this case, but accountability would mean that we are able to raise questions about the logic of that principle and we would be able to see whether it has particular impacts on the life chances of certain individuals. They disclose themselves that it does almost geographically profile the area that that police force covers in terms of the probability or the likelihood of reoffending, which matches exactly the research that we see in the United States in terms of recidivism. This logic becomes ingrained or inscribed within the algorithm, and once it is there it is very difficult to escape, and it is much more difficult for a defence lawyer, for example, to challenge it in court.

Chair: Understood.

Professor Jennings: Can I briefly come back to your data point and its value? It is a mixed landscape. Some data is best openly released and is shareable. We have an Open Data Institute in this country; we lead in our open data standards and what we do with those. That is fantastic for some classes of data. We would not want to release all our data under those sorts of mechanisms. For more sensitive data that cannot just be released and shared for the common good, we need to find more sophisticated ways of sharing that data in a way that protects privacy but lets others get into it. A monopoly of some big providers on particular sorts of data is not a healthy place to be in.

Q36 **Martin Whitfield:** I have a slightly mischievous question. On the social media platforms we hear about bots and artificial cases, and creations like that. We have fake news. Do you anticipate a danger in fake data affecting algorithms simply by the volume of fake accounts that are opened and closed? Do you think there is a risk of that or is that far away still?

Professor Jennings: No. That is very much here and now. There are many sources of fake data out there. If you have a number of bots learning over them and interacting over them, it is very difficult for them and sometimes for people to figure out what is and is not the fake bit. That inevitably happens. Your machine-learning algorithm is based on the

quality of your data. If there is fake data there, you cannot tell it is fake data and you do not exclude it, it is going to influence what you do.

Q37 **Bill Grant:** In noting the existence and improvements to the data protection and the Information Commissioner—putting them to the side, but noting they are there—we have, on the one hand, technology companies advocating self-policing, self-regulation and shared principles. On the other hand, we have some research institutions suggesting there should be an overarching regulation to deal with algorithms going forward. It is also suggested that self-regulation or shared principles may be quite weak and that if you overregulate there may be restrictions. Where would the panel favour going forward—self-regulation or an overarching regulation?

Dr Weller: Large companies certainly have an incentive to try to get ahead of regulation. Clearly, they are trying to protect their positions, but I believe they see it in their best interests that they want to try to move forward in a way that makes the public comfortable. I believe that it is dangerous to regulate too much too quickly. We may get things wrong, and that will stifle innovation and lead to trouble. As I mentioned before, I agree with the Royal Society and the British Academy report that suggests we should have a body set up to advise on the stewardship of these issues and that that is a pressing issue.

Q38 **Chair:** Do you all agree with the sense that Adrian indicates—that we need to establish a process as a matter of some urgency to come to a conclusion about the right model and we need to get on with that? Am I interpreting you correctly, there, Adrian?

Dr Weller: Yes.

Q39 **Chair:** Do the rest of you agree with that?

Hetan Shah: I think so, and it is where we are headed. In a way it is not just your conversation, but the Information Commissioner has published a paper on big data analytics and it is the first data protection body in the world to have done that, so we are ahead of the game there. The Nuffield convention on data ethics I mentioned earlier has been established as an arm's length body, and then it was in the Conservative party manifesto to have a data use and ethics commission of some sort. That is now something that Matt Hancock is considering, springing from the Royal Society and British Academy recommendation of a stewardship body. My understanding is that, if such a body were created, its role would be to scan the landscape and say, "What is missing? How should existing regulators change?" and so on. This idea of reflective processes is something that is beginning to happen.

Q40 **Bill Grant:** Am I sensing that there is not a desire for overregulation or regulation at all in going forward in algorithms? Should it be left to the companies? Is that what you are suggesting?

Dr Weller: No, I am sorry. If I gave that impression, it is not what I meant. Many people point to issues that seem troubling and problematic—for example, the way the social media can affect elections and targeted ads pointed at particular people. Having said that, it is very easy to have a knee-jerk reaction and say that we need to do something, but it is very hard to think carefully about what exactly should be done and how that interacts with other parts of our society. It is important for us to think carefully about what should be done before doing something too quickly.

Q41 **Bill Grant:** Can I have a supplementary? You mentioned two things—elections and public—which has prompted a question, and not least of all where we are located today. Going forward, will algorithms enrich or endanger democracy?

Dr Weller: We really need to strive together to make sure that algorithms are going to enrich all of society and help everyone move forward together, and not leave anyone behind. That is going to require some careful thinking. I do not have any quick answer, I am afraid.

Professor Jennings: There is not a right answer. It is like asking whether web pages promote or kill democracy. They do both. Algorithms have the potential to do both. People will try to shut down democracy and give particular messages to particular targeted folk, and others will try to make it more open. There is not a one answer for all algorithms; there will be both.

Q42 **Vicky Ford:** I want to take specific actions, and I liked your specific action to write to all the different regulators and find out what they are doing in this space. I have heard that the Information Commissioner wants to take certain actions. You have mentioned the Nuffield bioethics committee, and we have heard the Alan Turing Institute suggest we should set up one separate regulator. Are you suggesting we should set up or bring together a panel? It is clear you say there is not a one-size-fits-all solution for all algorithms or all responsibility. What other actions do you want us to take?

Professor Amoore: One action around that idea of a committee or a body should be to acknowledge also the interdisciplinary nature of these kinds of technologies. It is not only a matter of so-called data science, mathematics or computer science. Without wishing to pre-empt the ICO inquiry, Cambridge Analytica, for example, is quite profoundly engaged also in psychology. This is about behavioural modelling of what somebody's future actions or intentions might be, and we can see that in both the domain of election processes but also criminal justice, for example. The reason why this body would need to be independent and interdisciplinary is because this is not only a matter of the mathematical arrangement of algorithms. It is also about imagining what some sort of future might look like and thinking about somebody's future actions.

Q43 **Chair:** Are we talking about both a body to design what the framework

should be like but then a body to monitor on a continuing basis? Is that right?

Professor Amoore: Yes. At the moment, where would one go to seek redress? I know you are going to talk about the question of the GDPR with the next panel and whether it is possible already to seek redress from an automated decision, but I see a gap there in our current frameworks of exactly how one would seek redress for an outcome or a whole series of outcomes, because often this is algorithms acting on other algorithms—not just human decisions and algorithms, but algorithms interacting with one another. How would one seek redress from that chain of events? At the moment there is quite a lot of emphasis placed on the human in the loop, but that would be an important question. Who should that human in the loop be who has oversight? To go back to the point, that would need to be an interdisciplinary independent body, which is able to understand those processes.

Professor Jennings: The role of such a body is both to work with existing regulators in particular sectors and to challenge them—what about this and what about this?—and you have heard a couple of those sorts of queries going back, but also to identify in more of a proactive way that this is what these new technologies are bringing to society. It is a very broad range of things. What needs to be done at a societal level? Where do we need to act? Where should we not act, because people stifle things? That is the role of the committee, I think.

Hetan Shah: I have given you two recommendations already, to write to regulators and tell the public sector that they have good datasets.

Q44 **Chair:** We will take you up on your suggestion.

Hetan Shah: I have two more recommendations. One is that the Information Commissioner should be strengthened. Part of that is for Government to sort out its funding model and part of it should probably set up an advisory committee on this area now because it does not have that. My fourth one is to tell universities that they should be embedding data ethics into their courses in this area for data scientists.

Chair: Brilliant. Thank you all very much. It has been an incredibly fascinating session. Thank you.

Examination of witnesses

Witnesses: Silkie Carlo, Dr Sandra Wachter and Dr Pavel Klimov.

Q45 **Chair:** Welcome to all of you and thank you for being here. Could you introduce yourselves briefly and say where you are from?

Dr Klimov: Good morning and thank you for this opportunity to appear in this inquiry. My name is Pavel Klimov. I am a solicitor. I am here on behalf of the Law Society. I am also general counsel working in the IT and digital media sector.

Dr Wachter: My same is Sandra Wachter. I am a lawyer and researcher in data ethics. I work at the Oxford Internet Institute at the University of Oxford and I am also a Turing research fellow at the Alan Turing Institute in London.

Silkie Carlo: Good morning. I am Silkie Carlo from Liberty. I am a senior advocacy officer and I work primarily on our programme on technology and human rights.

Q46 **Chair:** Thank you very much. Have your respective organisations been undertaking analysis of how the use of algorithms is changing, how it is developing, and, if so, can you tell us something about that? Perhaps, Pavel, you can start.

Dr Klimov: Certainly. In the Law Society we are closely monitoring these developments, in particular in the works of various committees, and I am chairing the technology and law reference group. It is true to say, as the previous panel was saying, that the traditional view of the algorithm as linear instructions significantly changed over the last five years, and now we are talking about this neural type of algorithm, machine learning, where the data is given to the algorithm and based on that data, in various noughts of that process, it comes to a certain decision and is then translated to the higher layer. At a certain point in time, it would appear that it is not that easy to cascade down that logic and, in other words, unravel why certain decisions were taken at different levels. That is where the issue lies, in our view. We are essentially dealing with a situation where humans may no longer be in control of what decision is taken and may not even know or understand why a wrong decision has been taken because we are losing sight of the transparency of the process from the beginning to the end.

Q47 **Chair:** Thank you. Sandra, perhaps you can deal with how widespread the application is already, because I am conscious that in the previous discussion we talked about what is needed. I am interested in understanding how important it is to be getting on with this and what we need to be doing now. How widespread is the application?

Dr Wachter: It is incredibly widespread. Basically, every decision that can be done by a human can now be done by an algorithm. The applications are wide-ranging, starting with financial trading and the stock market. Banks use them in order to decide if somebody should get a loan; insurance companies use them to decide whether someone should be granted insurance and what the premiums would be like. They are used in health, for example, as well. The decision of whether somebody should get hired, fired or promoted can all be done with an algorithm. Should somebody be admitted to university? In the criminal justice sector, granting parole, arresting someone and determining sentencing—all of that, which was done by a human—can now be done by an algorithm.

Q48 **Chair:** Silkie, has your organisation been looking at how widespread the

use of algorithms now is, and what is your assessment of how it is going to change over the next five years?

Silkie Carlo: Yes. There are a number of areas in which, already, we have concerns about the potential human rights impact of the use of algorithms in decision-making. Naturally, primarily, our focus has been on uses of algorithms by the state; so we have been looking at the intelligence agencies' and law enforcement uses of algorithms. Obviously, there is a lack of transparency around use of algorithms by the intelligence agencies, but knowing that they are in the business of collecting bulk data and in the practice of bulk surveillance is a real issue for us and we think there ought to be some transparency of the algorithms that are used by the security and intelligence agencies.

We have also started looking more recently at the uses of algorithms in law enforcement. For example, we are really interested in the use of Hart by Durham that was discussed earlier, and also, increasingly, in biometric forms of surveillance as well, such as the use of algorithms in facial recognition and the kinds of issues that that presents, not only in privacy interferences but potentially raising concerns about discriminatory policing that could be shrouded by the algorithm.

Q49 **Chair:** Professor Amoore suggested that there may be some areas where algorithms simply should not be used, and she mentioned the criminal justice system. Do you have a view on that?

Silkie Carlo: Our view is that, where algorithms are used in areas that would engage human rights, they should be at best advisory. We have a concern at the moment that there is an exemption in the Data Protection Bill in this area. The GDPR gives us the right not to be subject to significant decisions that are purely automated, but there is an exemption for intelligence services processing and law enforcement processing, and that is exactly the area where we think—

Q50 **Chair:** This gives carte blanche in a way to those areas.

Silkie Carlo: Potentially, yes. That is exactly the area where we think this protection is really important. We are lobbying for an amendment to the Data Protection Bill that would ensure that any decisions that engage human rights would not qualify for that exemption.

Q51 **Vicky Ford:** Can you say that again? The Bill gives an exemption for what?

Silkie Carlo: The GDPR transferred into the Data Protection Bill gives individuals the right not to be subject to a purely automated decision that is a significant decision or has legal effects, but there are exemptions in the Bill.

Q52 **Chair:** And in the GDPR.

Silkie Carlo: The GDPR allows member states to draw their own exemption. Our exemptions have been applied in a very broad way for

law enforcement processing and intelligence servicing processing in particular. That is concerning. The amendment that we are lobbying for is one that would remove that exemption where a decision engages human rights.

Dr Wachter: I would totally agree. The problem with the GDPR is that it has some good intentions there. Ultimately, the decision-making can be allowed when three of the bases are filled. Either I give informed consent and then automated decision-making is legal; or it is necessary for the performance of a contract, and then it is legal too; or, as was just mentioned, if member states allow that. The problem is that in the GDPR there are no clear indications what safeguards need to be in place to make automated decision-making based on law lawful. That is a problem. It could be a higher standard than explicit consent, or it could be a lower standard. There is a large margin of appreciation for the member states.

As it is currently in the Bill, there is room for improvement, but, for example, the House of Lords amendments have proposed something to close those things, at least in some respects. For example, it was proposed that, if an automated decision is being made in member state law, the individual should be informed about the outcome, which is very important, which is something that is not currently legally required in the GDPR—so, that will be better actually; and when you get informed about this decision you should have the possibility, basically, to contest it—to say you want the decision to be reconsidered or you want a decision that is not solely based on automated processing. As it is currently proposed by the House of Lords, this would set a higher standard and close one of the loopholes that we currently see in the GDPR.

Q53 **Chair:** On the other side, where do you see the most transformative positive impacts of the use of algorithms? Do any of you want to comment on that, or are you only looking at the negatives?

Dr Klimov: No, not at all. It is the reality of our lives that we use technology, and technology helps us. It is not so much the creation of technology that is an issue; it is the usage of it—to make sure that we put it to good use and put the right checks and balances in to ensure that we do not turn it into a weapon against ourselves. The previous panel identified various elements, from medical research to the financial sector, where algorithms help us on a day-to-day basis.

I wanted to add to the issues that have been highlighted in terms of automated decision-making. We have a concern with the GDPR that it still operates with these notions that were created 20 years ago, or even earlier, where all the data cycles were more or less closed. People would give their personal information to a controller; a controller would process it, apply it and make a decision; and now we operate in open systems where snippets of information are gathered about us on a pretty much continuing basis. Each snippet of information may itself not even be personal data to which the GDPR applies, and these automated decisions may be made on the basis of these various pieces of information

gathered from different sources. We may not even know about this decision being made against us. How effective is that right to object or to get someone to reconsider the decision with a human element, where one does not even know that they have been excluded from certain services or been denied certain opportunities because the algorithm placed them in a particular category or labelled them in a particular way?

Q54 **Chair:** Are you suggesting some sort of right to know that that is how the decision was made?

Dr Klimov: I am suggesting that the debate should be broadened beyond the individual versus my data and how it is used. It is about data management. It is not necessarily that consent is something that should allow people to do it and if I do not consent that they should not, because that essentially directs efforts by people who want to use it to get this consent. That is what has been described as a transparency paradox where we are now bombarded with these cookie screens; you click on through various layers and essentially consent to anything and everything under the sun that can be done to your data without anybody knowing. That as an effective means of preventing abuses and the things that we talked about may no longer be as effective in the current open data management environment.

The debate should be shifting towards what is a fair use of data, and, irrespective of whether it qualifies as a definition of personal data or not or whether there is consent or not, if you hold a swathe of data about groups of individuals, particular individuals and general society, what you can and cannot do with that data and what are the principles that should be applied there.

Q55 **Stephanie Peacock:** I have a question for Sandra to begin with. The Oxford Institute told us that algorithmic decision-making can harm individuals in new and unique ways. Could you elaborate a little on what the new harm is that they are causing?

Dr Wachter: Yes. The previous panel mentioned that the difference between traditional algorithms and now machine learning and more complex systems is that they are unpredictable, and if they are unpredictable you cannot foresee the outcome, so it might be that you start discriminating against certain people whom you do not even know that you are discriminating against.

We need to think about maybe a more refined harm taxonomy. What are the actual ethical and real-world problems that could arise if we cannot predict the outcome from the beginning? That is something that is new. The other thing is pervasive inferential analytics. Even if I collect data that seems neutral, I can just predict or infer a lot that I was not previously able to. It is the postcode example. It seems like it is very neutral data, but we know now that where you live also allows a lot of insight into your socioeconomic background and all of that. The proxy

data and what we can read from the data is new, and that could lead to new harms and new kinds of discrimination.

Q56 **Stephanie Peacock:** That makes sense. We talked with the last panel about and touched on the legal frameworks. Would you agree that the current legal framework fails to provide a remedy or offer redress for people who have suffered harm?

Dr Wachter: Yes and no. What I think is very good is that the GDPR is going to give individuals a right to contest decisions, express views or obtain human intervention if a fully automated decision with a significant effect is being made about them. That is very good, and it is the right step forward. The problem that I see is how this would work in practice.

If I want to contest a decision because I think it was discriminatory or wrong, I would need to understand what happened in the decision-making process; so I need some kind of explanation. I would need to know why I was denied the loan, why I did not get the promotion, why I had to go to prison, and then contest it on that basis. A right to explanation is not yet legally binding in the GDPR. It is mentioned in a non-legal binding provision, recital 71, which is guidance; it is not legally enforceable. In fact, it was actually proposed to come in a legally binding text but was not adopted during trialogue. You could say there is an idea there, and obviously the legislators want to give some kind of explanation or at least offer the opportunity to give an explanation, but the question is whether it is actually enforceable. I very much doubt that a right to contest the decision will be very meaningful if I do not know what I am contesting.

Dr Klimov: The right, in terms of being informed, is not a new one. It is in existing legislation and in a directive, but from my research and knowledge there are no cases, or at least no published cases, where individuals effectively exercised that right. I do not know if anyone else knows of any. It is good on paper, but whether it works in practice is a different question.

Dr Wachter: I agree. There are very limited cases where the right that you were referring to is the right of access, which we have in the new GDPR, and we had a similar right, and do still have, in the data protection directive, which means I can go to data controllers—that is everybody who holds my personal data—and ask them what kind of data they hold about me and what they are doing with that data, and ask them to give me access to the logic involved of the processing that they are doing.

We have a similar right now in the right of access as well. We wrote a paper where we looked at existing jurisprudence. There were very limited cases, but there were a couple of cases in Austria and Germany where people went to court about credit scoring. They were denied and wanted to know how their credit score came about. The access rights did not give them much. What the courts were willing to give them was very much limited because of trade secrets. This was an effective barrier to fully

exercise that.
One of the essential things is that very often trade secrets and IP rights are hindering transparency, even though it might be technically feasible, because those are the two things that usually go hand in hand. Either I cannot explain a system because it is not interpretable or I do not want to because I fear that my competitors would pick up on that.

Q57 **Stephanie Peacock:** What tools do you think viewers, readers or indeed perhaps parents need in order to identify content that is created by an algorithm?

Dr Wachter: For example, there are guiding principles in the GDPR that say fairness, transparency and accountability. The problem is that what that means is not really defined, but if you look at the framework in general it wants to make you aware of what is going to happen. The problem was already mentioned with regard to all the terms and conditions; nobody is going to read that; there are hundreds of pages, and nobody is going to go through all of that and be informed. It would be helpful to give very concise, easily understandable information such as, "Okay, this is what is going to happen on that web page; this is what is going to happen with your data." Then you can make an informed decision whether you want that.

For example, in the current GDPR it was suggested that the notification should not be pages and pages, but maybe icons—just an icon that says, "Okay, we are collecting your data here. All the data that we are gathering will be encrypted,"—very short information. If you want to learn more, that is possible too, but it should give you a very easily understandable overview of what is going to happen to your data while you are visiting a service.

Q58 **Stephanie Peacock:** What do you think the role for the Government is in this?

Dr Wachter: In terms of transparency, it would be very good maybe to come up with icons. The GDPR says that the European Commission is now tasked to develop such icons, which would be very helpful, for example, for children as well. Explaining an algorithm is challenging to computer scientists, let alone lay people, even children, but all those people are using those services, so you need to give them something so that they understand what they are consenting to. Consent is a very critical issue. We need to think about how we can get the message across very quickly, very sharply and precisely to a variety of audiences.

Q59 **Stephanie Peacock:** To conclude to the whole panel, do you have an opinion on whether there are wider cultural and civic issues that emerge from algorithmic-created content?

Dr Klimov: Certainly, algorithms do not live in isolation. They train on the data that we produce. They reflect, to some extent, our behaviour and the biases that come with that. There are certainly opportunities for bad actors to use those algorithms to weaponise them for their

advantage. It is like any other technology in our lives that can be put to good and bad use. Especially educating from a younger age about the pluses and minuses and risk associated is something that has to be high on the agenda.

Q60 **Darren Jones:** On this issue of explanation, lots of companies will have quite complicated contracting structures with suppliers and different processes and different parts that connect my customer database right through to how I place ads on someone's Facebook page. Are companies in a position to explain what is happening? Do they know?

Dr Wachter: It really depends on how you define "explanation." What was mentioned before was interpretability: is it humanly understandable and can I explain it? Can I explain the why—why did you arrive there? That can be very challenging and maybe not always humanly understandable, but you can think about explanations from a different viewpoint, because if I am concerned that I did not get a loan I might not care so much why, but I want to have something that helps me to contest it, to understand it even better, or to know what I can change to get the desired result in the future.

For example, recently—I think a week ago—we released a paper together with Dr Brent Mittelstadt and Dr Christopher Russell, both also Turing fellows, where we looked at the possibility of counterfactual explanations. This would be an idea of giving somebody information about how a decision was reached without explaining the black box. That would help you to understand why you did not get the loan. It would tell you what would have needed to be different in order to get the loan and give you some grounds to contest it. If I give you the explanation that you have been denied a loan because your income was £30,000 but if it had been £45,000 you would have got it, I can do various things with that information. First, I would understand what happened there and what kind of criteria they had used in the decision-making process. I could say, "Well, actually I make more than £30,000." So, I could contest it on that basis, and it would also give me an indication of, okay, I might not have £45,000 now but I could re-apply at a certain stage and maybe then get the loan. All of that would be possible without explaining the complex system but just giving some indication. You can think about explanation in various ways without necessarily attempting to open the black box.

Dr Klimov: Also, it gets more complex where it is not a binary decision whether you get a loan or not, but whether you are offered a service at a particular price point, say, or are directed to one resource or a different resource, which in many cases is happening now as well in the digital marketing space in particular. Then it becomes much more complicated; it is not one black box—it is multiple black boxes that are involved in multiple pieces of information that get gathered to come to this ultimate decision, which gets escalated. There, it would probably be very difficult to pinpoint one person and say, "Explain everything to me." Here, we would probably be more reliant on looking at the input and output data

and comparing it with the common-sense logic. Then, hopoefully, over time, when those practices are reviewed more and built on, we can weed out bad practices, biases and discrimination from the systems.

Q61 **Darren Jones:** On that point, GDPR introduces this idea of pseudonymisation of personal data rights so that you can hash it. I am interested in the point that, again, if you use social media platforms as an example, ultimately, you are targeting an advert at an individual based on characteristics about that individual. Do you agree that pseudonymisation removes the right to protect your own personal data if it is therefore being targeted at a person at the end?

Dr Wachter: Yes, and it is even worse than that, because data protection only applies if personal data is being processed. However, we are making, ultimately, a decision that uses anonymised data. Anonymised data does not fall within the scope of data protection. We also use data of other people to make decisions about individuals. Profiling is a perfect example. We have heard this already with employment decisions. I train my algorithm based on historical data and find the perfect candidate. That is all based on someone else's data. That is not my data; so, actually, data protection law does not apply to that. This is the question of: is data protection law actually fit for purpose for machine learning and AI? I would be very hesitant to say that. The problem is that it is not so much the fault of the GDPR, because the GDPR was not designed to govern machine learning and algorithm decision-making on a broad scale. It is just some legacy that we held from 20 years ago and it was not really updated that much. But we need to think about a new definition of personal data, because decisions will be made based on profiles and anonymised data, and data is then generated from the training data, which is not always considered personal data. Then you have none of the safeguards.

Q62 **Darren Jones:** On the question specifically of article 22 of the GDPR, there are two questions I am interested in your views on. The first is that it talked about having a legal or significant effect. What in your mind is a significant effect?

Dr Wachter: That is a very complicated and unanswered question. In the recitals themselves, on which they give guidance, in recital 71 they give only two examples: e-recruiting purposes and credit card applications. Those are the only two things. The article 29 working party recently published guidelines on that and they say it is an unsatisfactory answer but a very sensible approach, because they say that what "specific effect" means is context specific. It will depend on the individual circumstances of the individual. If I apply for a loan but I am very well off it might not have a significant effect on me, whereas an application for a mortgage could have substantial ramifications if I do not get it.

The guidelines explicitly mention targeted advertisement, for example, and how ads, if they are specifically shown to you, could affect your way of choosing products, or ads for positions. If I am filtered out, then I do

not know that I could apply for a job and therefore would be discriminated against. It will be very context sensitive and it is very hard to define.

Dr Klimov: In some respects it is also a bit circular because one suggestion is that you also have to decide on who the target is, particularly if it is a vulnerable person. They give one example of someone who is addicted to betting and who is in financial difficulty. They are more likely to be tempted if you show them betting ads continuously. It predisposes that you have profiled somebody already, that that person is vulnerable and keen on getting his finances right. That article is precisely designed to prohibit that. Where do you start?

Yes, it is a notion. They suggest it has to be something serious, akin to denial of a legal right or certain fundamental services or access, but we will need case law to decide on that.

Q63 **Darren Jones:** We were discussing briefly in the previous panel the role of regulators and whether we needed to use the current sectoral regulators that exist or we needed a cross-sectoral approach. On the question of a case-by-case assessment of what is a significant effect, do you think the regulators are currently fit for purpose in being able to take that decision when intervening?

Dr Wachter: It will kind of depend. For example, in articles 35 and 36 the GDPR is going to obligate data controllers who are dealing with profiling and automated decision-making to make a data protection impact assessment, thinking about the possible risks of their application beforehand. This is a very sensible approach. I guess it would help us to identify the new risks that come with data analytics and then just work from there. I very much agree that a one-size-fits-all solution is probably not possible and not sensible. I think we can agree on overarching principles such as fairness, transparency and accountability. Everybody will agree on that, but we have to trigger it down to the individual use cases and sectors, because fairness means something different for autonomous cars when algorithms are being used, whereas fairness in healthcare means something different, and probably fairness means something different in the criminal justice sector. So, even though we will agree on certain principles, we need to look at specific applications and work from there, but having more foresight would be helpful, to think about the possible risks beforehand, and then decide what needs to be done before we have even deployed them.

Q64 **Chair:** Do the overarching principles that you talk about sit there and get applied by the individual sectors, or is there a body that has responsibility for protecting those overarching principles?

Dr Wachter: The idea would be that, if you have something like a data stewardship body, then you would agree on transparency, privacy and accountability, but the question of how it will apply, for example, will not just be a data protection problem; it will also be a problem of the laws

that will apply to that sector. Fairness means something in data protection and machine learning or computer science, but fairness and non-discrimination mean something completely different in labour law. But if we are making recruiting decisions on that, we need to ensure that we satisfy those requirements as well. It would have to be a very holistic approach when we regulate or when we look at the specific sector. It will not just be data protection. It will be much more.

Chair: Understood.

Q65 **Darren Jones:** I have two very short, quick last questions. The first is that, looking at article 22, it says that data subjects have the right not to be subjected to a decision, which—and correct me if I am wrong—in my understanding means you have the right to "not"; so you can remove your consent after it has happened. The ICO guidance says that data subjects have to give explicit consent in order to be profiled. Is the interaction with the data subject before the profiling happened or afterwards?

Dr Wachter: Yes. It is a very great point, something that we pointed out in that paper, and I think it has now been clarified thanks to what the ICO said and what the article 29 working party guidelines said. You are absolutely right. The way you read it, it could be both. It could either mean I have the right to object once it has started, or it is a prohibition and there will only be a couple of exemptions when automated decision-making will be legal. As it stands now, the ICO said in the article 29 working party that it is a prohibition. Therefore, there is a general prohibition of automated decision-making unless I give explicit consent, there is a law that allows it or it is necessary for the performance of a contract.

Dr Klimov: I think that is also maybe the right place in that it is a prohibition rather than a right of objection, but that forces businesses that use the data to get that consent up front. It forces this endless policy. Now they said, "Okay, we need to have a layered approach. We have icons and then you scroll down and up," but it does not resolve the issue because, essentially, instead of having everything on one piece of paper, you now have to click, click, click to get to the bottom of that, and I do not think there will be that many people who—

Q66 **Chair:** They will just click through without thinking about the implications of what they are doing.

Dr Klimov: Because those policies were so long and nobody ever read them, now we have to give certain information up front, and then if you want to learn more you have to click further, but essentially it creates a situation where you have very limited information given to you up front and if you want to get to the bottom, rather than reading from top to bottom on one page, you simply have to go and click through these icons to understand how your data can be used. Usually you will be directed to

third-party privacy policy as well because the data may be shared with third parties and so on. In term of getting yourself more informed—

Q67 **Chair:** It is not a particularly informed consent.

Dr Klimov: Unless someone is keen to read it through to the end.

Dr Wachter: In terms of informed consent, yes, but, since automated decision-making is also possible without informed consent, transparency is essential. The fact that I know my data has been gathered, that it will be processed, and that it will be shared with someone else, if I cannot object to that, I should at least know about it, and that is what the notification to this wants.

Q68 **Darren Jones:** This is my very last question on the amendment in the House of Lords with the human rights piece. Was there any thought given to any rights that might be lost around those derogations if we withdraw ourselves from the EU Charter of Fundamental Rights compared with just the Human Rights Act?

Silkie Carlo: The aim of the amendment is in the current environment in which we are working, and we were envisaging, taking the ways in which algorithms are being used currently by law enforcement, what kind of things could happen with this exemption. For example, with Durham's use of Hart, it is frequently cited that a really important protection is the officer's discretion. The Hart tool is advisory, and, clearly, with this exemption, an officer's discretion would no longer be necessary as part of the process. There, potentially, you have someone's right to liberty at stake, so this is a grave issue.

With the use of facial recognition, one key concern we had, particularly with the decision to use it at the Notting Hill carnival two years in a row, is the potential for discrimination. We have seen in the US, when facial recognition software was tested for accuracy biases, it was found that the software was more likely to misidentify female and black faces. Again, in this instance officers cite as an important protection the fact that they themselves check the matches that come up. The officer's discretion is still an important part of the process that protects people's liberty, that protects people from undue interference and from being discriminated against. I am not entirely satisfied that the officer's discretion does all those things in that instance, but the point is that, with the current exemption in the Data Protection Bill, it seems to us that theoretically it would not even be necessary to have the human check on the decision made by the algorithm.

Q69 **Darren Jones:** In the EU (Withdrawal) Bill, once we remove ourselves from the EU charter, article 8 says that citizens have the fundamental right to data protection. Do you see that not being obliged to be subject to article 8 of the Charter of Fundamental Rights causes us a problem after Brexit?

Vicky Ford: Chair, can we come back to algorithms, because we can go

on about data protection for a long time, but I am aware that—

Darren Jones: This is about—

Chair: Can I have a quick answer to Darren's point?

Silkie Carlo: We have been reading it across with the HRA, yes.

Q70 **Vicky Ford:** I am sorry, but I do want to get back to the question of algorithms. Some companies have been very reluctant to disclose their algorithms. They say they have invested in developing these algorithms, they are intellectual property and that they should be protected by copyright and trade secret protections. This is particularly an issue for our first witness here, for Pavel from the Law Society. Do you see that there should be the need to protect copyright, or should the right to explanation trump that?

Dr Klimov: The issue of transparency is important, but it should be meaningful transparency, because, if we mandate simply that all the companies have to disclose source code of the software they use for this algorithm decision-making, that will certainly make a big impact on business. It potentially can stifle innovation; it will create a higher entry barrier for start-up businesses in that area; and it will also potentially give access to that information for bad actors, because they will also be asking for that information and they will see how they can utilise it for their benefit. Whether it will give meaningful understanding for the people who may feel affected in a wrong way by automated decision-making, I am not certain. The issue should be that there has to be a logical explanation, so that people, whether by themselves or with the assistance of a specialist, can understand why a certain decision is being taken and in what way. That can be based on the input and output data with certain knowledge. It can be based on some further testing being done.

Q71 **Vicky Ford:** That is all about transparency, but should copyright and intellectual property law also exist with algorithms? Should you be able to patent an algorithm?

Dr Klimov: It should be to the same extent that it exists today. If it qualifies for the patentable algorithm, it can be. If it is simply an idea that is not expressed or the other conditions are not satisfied to be patentable, then I do not think there should be specific exceptions made for that, because it is a decision-making algorithm as opposed to any other algorithm.

Q72 **Vicky Ford:** Some people have suggested algorithm auditing. Do the panel have views on algorithm auditing? Is it workable? Can we have a system of algorithm auditing that does not also hold back innovation, because we know we want to support the innovation? Are there any other tools that we could give to the end users to help them feel they have more control over the underlying algorithm?

Dr Klimov: When people are confronted with a black box, they will not understand. There are lots of black boxes in our lives that we trust, and we trust them because they work or we know that the input data and output data makes sense. You can talk about TV sets and the Coca-Cola drink. We do not know the formula but we still drink it. Auditing it again, there has to be a way of verifying it. Whether that means that you have a third party going and unravelling or reverse-engineering the software that is used very much depends on the circumstances of what decisions are being taken by that algorithm. If they are particularly intrusive of human life, there might be a case where, if things go wrong, they have to be mandated to open up, maybe in a controlled environment, showing how the algorithm works and where the problem areas may lie. I do not think there is one solution that will fit all the scenarios.

Dr Wachter: Trust can only be established if we have the opportunity to challenge decisions or to hold someone accountable. I do not think anybody is going to blindly trust anyone. We only trust systems; we only trust humans—

Q73 **Vicky Ford:** That was not my question. My question was, are there any other tools that we could give end users, such as auditing or any other tools, to help them?

Dr Wachter: Yes. Auditing, broadly speaking, could be something that is very helpful. If we have the problem of trade secrets and revealing too much, having a third trusted party that would have the power to investigate further would help to balance those things. If I feel discriminated, I could go, for example, to the ICO, and the ICO could then investigate the code, if possible, and see if there is actually something wrong with the system. We could also think about certification schemes. So, before we deploy algorithms, for example, in highly sensitive areas, such as criminal justice, we give seals, or certification schemes, to those algorithms so that they have to be tested for a certain time. This could help increase user trust.

Q74 **Chair:** That requires some sort of body that can do the certification.

Dr Wachter: Yes.

Q75 **Chair:** Is there any comment from you, Silkie?

Silkie Carlo: I would echo the other comments that were made and say that I think that that kind of auditing process, particularly by an expert body, is exactly what is required for innovation, and I do not think that would be seen as holding back innovation. But, as algorithms are used increasingly, particularly in the public sector, we also need at the same time those frameworks for accountability and auditing to be developed alongside the algorithms that are being deployed.

Vicky Ford: It is interesting you mention the ICO because, to me, that is part of the role of the data controller and Information Commissioner as to how that information is being used as part of that discussion. Thank you.

Chair: Vicky, you are saying that that body potentially performs that role.

Vicky Ford: The Information Commissioner is meant to be making sure that data is protected and not abused. That is why the Information Commissioner has suggested that that office should have a role in the development of how that is used and analysed in the real world.

Chair: Thank you very much.

Q76 **Bill Grant:** In processes or systems that rely on algorithms, should these processing systems be totally autonomous, should they have various levels of autonomy, or do we need, dare I say it, a human element in this loop? Would the autonomy of the algorithm trump or override the human element? Are they two separate issues, or do we need them integrated or intertwined?

Dr Wachter: Again, I think it will depend on the sector where they are deployed. There might be some sectors where it is perfectly fine not to have a human in the loop, and it might be that we want some human control after all because the area in which the algorithm is operating is so sensitive. If you think about criminal justice, for example, do we want an algorithm to make the decision without any human involvement, or would we rather have a judge maybe using an assistive tool to help them make the decision? It could be more like going hand in hand, human decision-making with algorithmic decision-making, if the area is potentially sensitive. Both could work, but in highly sensitive areas we should not eliminate the human from the loop. In the current legal framework, it would mean that as soon as a human is in the loop the safeguards would not apply, and that is very tricky because sometimes we need humans in the loop, or we want them in the loop because we feel safer; but at the same time that would mean that the right to contest the decision or obtain human intervention or express your views would not apply any more because it is not solely automated. So, there is some room for improvement to amend this to say you should have safeguards when the decision is either solely or predominantly automated, which would leave room for having a human in the loop if necessary.

Q77 **Bill Grant:** In certain instances you are suggesting that the final decision would be a human decision assisted by an algorithm to arrive at that, which should be a better decision. If bias and prejudice are embedded in the training data for that algorithm that informs or populates the algorithm, whether that embedded bias or prejudice is known or not known and it remains opaque, do these algorithms risk widening the existing inequalities in society in ways that are invisible or obscure? Maybe I have not explained it well, but could they get it wrong? Could they discriminate if they are not populated or informed properly as to what is being asked of them?

Dr Klimov: They can. They can magnify it and they can also show those biases that exist in society, which we may not be aware of consciously.

There was an interesting example, which I do not think was mentioned in the previous panel, where a study suggested that women are less likely to be shown ads for higher-paid jobs. That was explained by some sort of bias that was built into the algorithm, thinking there were jobs for men and jobs for women. A subsequent study revealed—or at least it claimed—that that was a by-product of the, again, automated decision-making algorithm, which put a higher premium, a higher price, on female, as they call it, eyeballs being shown the advert in that age group than for men.

A marketing advertising company will target females, and the price for showing ads on the female screen is higher than for men. So, an algorithm that was meant to rationalise this span for this particular ad for higher-paid jobs was selecting men so that they spread it more widely across the population. Essentially, women were put in a category not because they are less likely to apply for a higher-paid job—and actually evidence showed that they are more likely to click on the ad than men if it is shown—but because, whether it was human-introduced bias or algorithm-made bias, the suggestion is they are more likely to spend money on the ad and therefore they want to be shown the ad, or the price to show the ad to them would command a higher price tag. By definition, if you want to spread it as widely as you can, then you send it to the cheaper audiences, who happened to be men in this instance.

You can see that it works both ways. The bias built into society gets translated into an algorithm that can be magnified by an algorithm applying that knowledge, that training data, and using it with a different consequence. That will highlight to us that this is the way we think about how we put labels on various groups of people.

Q78 **Bill Grant:** I have a supplementary on that one. What steps could or should be taken, if my terminology is right, when you are training or populating an algorithm to ensure that we do not have obscure biases or prejudices? What checks and balances can you introduce to prevent these possibilities?

Dr Klimov: It is just trial and error in my mind. Also, as humans, we have biases, and, arguably, if we train it with those biases we will only make the problem worse. But, in those instances, the more we put the study into that, the more we run those things and analyse them, as was done in this case, the more likely that will come up with those issues and they will surface, and we will be able to take corrective actions.

Q79 **Chair:** Sandra or Silkie, do you want to add anything?

Dr Wachter: Yes. As to what can be done to mitigate biases, there are a couple of things we could do. First, it is very important that we educate the people who are designing those systems so that they know about the responsibilities they have for society and what the future application means for society. The second is that the coding community needs to be more diverse. If we only have white male coders, of course the systems

are going to be biased—that is not a surprise—so we need more diversity in that respect.

Also, as to disciplines, I see a very strong trend on only looking at computer science and the STEM subjects at the moment. We are talking about questions of fairness, transparency, bias and discrimination, which are ethical and societal questions. The social sciences and the humanities have to have a role in that because they have been tasked with those questions for centuries. We need to have an inclusive discussion about those problems because that will help us to get rid of the biases that exist in the dataset.

Q80 **Bill Grant:** I have a tiny supplementary. I come from north of the border. It is only a wee Scotsman who would ask this. What is the lifespan of an algorithm and how often is its journey reviewed?

Dr Wachter: I did not hear the question.

Q81 **Bill Grant:** What is the lifespan of an algorithm and how often should its life's journey be reviewed? Do you make an algorithm just now and it will last for 20 or 30 years and should it be reviewed in five years, or is it going to fall off the shelf in two or three years?

Dr Wachter: It depends. I think it was said in the previous panel that we have very simple algorithms, where it is just a simple decision tree that we follow, which you can use for a very long time unless your business model changes. Now, with the more complex systems, they update as they go. So, with every input data they update the actual profile; they can be very fast-changing as well. It could also be that we have new regulatory frameworks coming into place and therefore companies need to adapt their business models. There is no binary answer to that. It will depend on a lot of circumstances.

Q82 **Martin Whitfield:** Dr Klimov, you raised the question of costs and price to companies. Do you think there is a role for strict liability in respect of loss that occasions from an algorithm's use, which would, from the injured party, remove the necessity of trailing this back to the company, a dataset, a software designer, and would that perhaps focus the attention on more transparent, explainable, understandable and more responsible algorithms?

Dr Klimov: Yes.

Chair: Keep your answers tight as we are coming close to the end.

Dr Klimov: Again, unfortunately, I have to answer as a lawyer. It depends on the particular circumstances and the way that the algorithm was developed and used. We are more and more talking about open systems where datasets get collected from different sources, where they get combined. It might be the situation where one good example will be automated self-driving cars, where you can say this is the person who developed the car and it is a product, and then a policy decision can be

taken as to whether or not strict liability should be given to the manufacturer. In many cases in other areas, these decisions are taken based on the off-the-shelf algorithm, some customisation may—

Q83 **Martin Whitfield:** There is a person has used it and there is a person who has suffered as a result of its loss.

Dr Klimov: Again, I do not think that strict liability will answer all the circumstances in all the cases. I think that potentially may put the users—and in many cases innocent users—of this algorithm at the risk of having to answer for the losses that, on the normal application of legal principles, common law principles, they will not be liable, just because they use that tool. If it was a different type of tool, the manufacturer may be liable, but I think strict liability is a very powerful weapon and has to be applied.

Q84 **Chair:** Sandra or Silkie, do you have any thoughts on that?

Dr Wachter: In order to develop a system of any kind of accountability, you would need to develop standards first, because you can only punish someone if they did not follow certain standards, and we currently do not have any coding standards. I would start from there because, even if you go to court and you try to prove negligence, if you do not have any standards you are not likely to win. We should have a standards approach first and think about how we design those systems properly before we can think about strict liability.

Q85 **Chair:** I want to end with two quick questions. Going back to the Durham example, we talked about their argument that there is always the safety of the officer exercising discretion after the algorithm has done its work. Has there been any analysis of whether that discretion is exercised in practice, or is the force of the outcome of the algorithm's work so strong that officers tend not to contradict the outcome of the algorithm? Does that make sense? Have we done any analysis of whether, in those circumstances, the person ever exercises the discretion to reach a different conclusion?

Silkie Carlo: I have requested that data from Durham constabulary and have not been given it, but that is exactly the kind of analysis that needs to be done to look at how meaningful the human involvement is in the process.

Dr Wachter: Yes, similarly, I do not know of any data but I think that is exactly the question that needs to be asked and also the question, am I going to be punished if I take a different decision? Is there a moral responsibility to follow the algorithm or is there a moral responsibility to not follow it, and what is the threshold? Is a 70% likelihood to reoffend enough to say you have to stay—

Q86 **Chair:** The individual exposes themselves quite—

Dr Wachter: Yes.

Q87 **Chair:** We had a discussion with the first panel about what we should be doing now in response to these really quite dramatic emerging forces. The suggestion was made that we should be establishing some sort of body to start the consideration of what framework we need to get in place and that we cannot just leave it. These decisions are being taken in real time in the Durham police force and all over the place, and it has emerged much faster in the criminal justice system in America, as I understand it. What should we be doing? Do you agree that we need to be setting up a body now to come up with recommendations for Government to implement?

Dr Klimov: I would say yes, and that will bring expertise from various fields—the legal profession, developers, businesses and social groups. That is just the reality of life and it affects all of us.

Q88 **Chair:** Is there some urgency to this, do you think?

Dr Klimov: Yes, there is, and they say that 90% of the data that exists today has been created in the last five years, so you can see it is exponential. It has been used at enormous scale. It is only going to progress exponentially and we need to get on top of it very quickly.

Dr Wachter: I totally agree. An advisory body is exactly what is needed, but it has to be equipped with the proper resources, the proper expertise and the proper authority, and it has to be multidisciplinary. It has to have social sciences and the humanities there. It is also important for it to have a multi-stakeholder approach, because we usually just look at academia, Government and the private sector, but we also need to look at NGOs and, for example, consumer protection groups, because those are the representatives of the people who are going to be concerned with those problems, so they need to have a voice on that too.

Silkie Carlo: I would agree, and also there needs to be a growth of expertise in regulatory bodies across a variety of sectors to be able to look at the growing use of algorithms in those areas. I have already pointed in some of my comments to the areas in which algorithms are already being used where there is a lack of transparency and accountability that Government could address a lot sooner.

Chair: Brilliant. Thank you all very much indeed for your time. It has been a fascinating session.