HOUSE OF COMMONS

Science and Technology Committee

Oral evidence: Algorithms in decision-making, HC 351

Tuesday 12 December 2017

Ordered by the House of Commons to be published on 12 December 2017.

Watch the meeting

Members present: Norman Lamb (Chair); Vicky Ford; Bill Grant; Clive Lewis; Stephen Metcalfe; Carol Monaghan; Martin Whitfield.

Questions 89 - 207

Witnesses

I: Martin Wattenberg, Senior Staff Research Scientist, Google AI team; Charles Butterworth, Managing Director for UK, Europe, Middle East and Africa, Experian; Dr M-H Carolyn Nguyen, Director of Technology Policy, Microsoft; and Nick Pickles, Head of Public Policy UK and Israel, Twitter.

II: Sheena Urwin, Head of Criminal Justice, Durham Constabulary; and Professor Kate Bowers, Academic Director, UCL Jill Dando Institute.

Written evidence from witnesses:

- Google
- Microsoft
- Sheena Urwin, Durham Constabulary
- Durham Constabulary
- UCL Jill Dando Institute

Examination of witnesses

Witnesses: Martin Wattenberg, Charles Butterworth, Dr Nguyen and Nick Pickles.

Q89 **Chair:** Welcome, all of you. Thank you very much indeed for coming along. I would be grateful if each of you could introduce yourselves briefly. Just say who you are and where you are from. By way of guidance for this session, with a panel of four it can sometimes take a long time to get responses from everybody, so don't feel obliged to answer everything. If you feel that someone else has given an adequate answer that you largely agree with and you have nothing further to add, that is fine. Try to keep your answers succinct so that we can get through everything we need to get through in the hour session. Let's start with introductions.

Nick Pickles: My name is Nick Pickles. I head up public policy for Twitter in the UK and Israel.

Martin Wattenberg: My name is Martin Wattenberg. I am a senior staff research scientist at Google AI. I am based in Cambridge in the US. I am very grateful to have the chance to talk to the Committee today. I co-lead a group called PAIR, which stands for the People+AI Research Initiative. The goal of the group is to think through, research and design ways for people to interact with machine learning and AI in a better way. I hope to provide a technical and scientific perspective on the issues today.

Dr Nguyen: My name is Carolyn Nguyen, director of technology policy based in Washington DC, working on artificial intelligence as well as issues related to the digital economy. I applaud the UK Government's effort. You are always ahead. You may already be aware that you are holding this hearing five hours ahead of exactly the same Senate Committee hearing on algorithms. Your industrial strategy is also ahead in the creation of the Office of AI, because today in the Senate there is the introduction of a Bill for the creation of a robotics federal commission. I applaud your efforts.

Q90 **Chair:** Excellent. You haven't been invited to both hearings.

Dr Nguyen: The power of technology.

Charles Butterworth: My name is Charles Butterworth. I am the managing director responsible for the UK, European, Middle East and African businesses for Experian.

Q91 **Chair:** I will start with a few questions. On the first question, I would be grateful if all of you could answer. Can you give a brief description of how algorithms are used in your various products and services?

Nick Pickles: Twitter is probably the simplest place. Twitter started as a reverse chronological platform, so the timeline was displayed first in the most recent tweet. Recently, we have started to use algorithms in how

we present tweets, but the content you see is tweets you would have seen anyway, so we are not taking content from other people's timelines and putting it into yours; we are taking content from your timeline and perhaps changing the order in which it appears, hopefully showing you the most relevant content first. Importantly, we have the ability to turn off that function, so if users want the non-algorithm version they are able to select that as well. That is the primary way we use algorithms.

Q92 **Chair:** Are people aware of this?

Nick Pickles: Yes. One of the challenges we have is communicating to our users the range of tools and options available. Earlier in the year, we made changes allowing people to see how their data were used for advertising, for example. It is very granular; you can go down to the very specific labels applied to your account to see whether you are interested in politics or football. That is a lot of information, but we think it is good to provide it, and we regularly communicate with our users both in the app and by email to inform them of changes as they are made. We also use Twitter to tell people when we are changing, and our users are certainly familiar with using Twitter to communicate back to us when we make changes, and share their feelings and views on them.

Martin Wattenberg: Google is a company built on software, and all software is in a sense implementing algorithms. An algorithm is essentially a series of very precise instructions; it is a very general concept along the lines of a business process, so in a sense everything we do is informed by algorithms. It is worth talking about the range of algorithms that we use. Many of them do incredibly mundane-seeming things; they convert data from one format that one module likes to use to another format that another one likes to use, making all sorts of minute decisions.

At the same time, some of the things we could recognise as an end-user as bigger decisions are done by algorithms. For example, when you use Google Maps to find the best route from one place to another, an algorithm does that; when you use the Google Translate site to read something in another language, that is an algorithm. But I emphasise that, because the concept is so general, algorithms pervade pretty much everything we do.

Dr Nguyen: The way Microsoft looks at machine-learning algorithms is driven by our focus to use these things to amplify human ingenuity. I want to focus on three parts of how we use algorithms today. One is further to develop cognitive services, which means increasing and improving our ability in language processing, speech recognition and text recognition. How does that help? We developed an application called Seeing AI, which enables people who are visually impaired to use a mobile app that allows them to see and hear a description of what is around them, as well as to look at product codes so that they can see exactly what the product is. It will also allow them to look at menus, so

that when they go to restaurants they can see exactly what items are included. That is one part of our capabilities.

Another part, because we believe in the ability of AI for innovation in terms of bringing to market new products, is that we include algorithm capabilities in our Azure cloud platform. The idea is that this will enable developers, as well as small and medium enterprises, to use and incorporate AI in their own products. Thirdly, we use AI in cyber-security to improve the security and capabilities of our own platform and software, as well as our data centres.

Charles Butterworth: At Experian, over time we have amassed a significant amount of data. Fundamentally, we use algorithms to help make sense of and provide insight in those data either to individuals or businesses. The best known usage is in the provision of credit-scoring. As such, we do not make decisions as a company; we provide credit scores and insight to allow lenders to make better-informed decisions about individuals when they provide credit, or we provide them to individuals to help them better understand how lenders would view them and how they could improve their credit score and so on.

It is not just algorithms in the context of credit-scoring; it would be in the context of preventing fraud and helping identify fraudsters. It would also be authentication, helping businesses and public bodies authenticate people as they come through digital journeys, and it would be in the context of helping companies and RAMS prospecting in the marketplace to make the right offer to the right person. There is quite a broad use of algorithms in that space.

Q93 **Chair:** There is a lot of debate about the potential to create or exacerbate bias or discrimination. Charles, I guess that in your line of business there is inevitably concern about decisions that may be made as a result of the analysis you have done. The point is made that it can reduce the discrimination or bias of which all of us are capable. What is your assessment of the risks there?

Charles Butterworth: There is always a risk not necessarily in the use of algorithms but in the people who use the algorithms, so there is the potential for bad outcomes.

Q94 **Chair:** In your case, you pass on your analysis to customers.

Charles Butterworth: That's right.

Q95 **Chair:** Isn't there a risk that they automatically take the advice from the analysis you have done rather than exercising human judgment on it?

Charles Butterworth: Over the years, particularly in the credit industry, there has been a significant amount of work done to create best practice, principles and ultimately regulation. Starting with the OFT, there are principles for scoring that are picked up by the FCA, so Experian is regulated as a financial services industry. Not only are we subject to the

laws of the land around discrimination on race or sexual orientation—the legal aspects—but we have the FCA principles and regulations, which govern outcomes for consumers at the same time. From our perspective, there is a substantial way of working that we have established over many years to ensure that there is no unfair discrimination in the algorithms used in our industry. That is the way I would answer your question about how we make sure we don't have it. Your point is correct: ultimately, unintended or unconscious bias can be avoided through the use of algorithms as well, but, on the broader question about how we make sure, it is through self-regulation of the industry, and oversight and supervision by the regulatory bodies around us.

Q96 **Chair:** Are there other contributions?

Dr Nguyen: With respect to bias in the use of data, an algorithm is based on the kind of data it is being used to train on. An algorithm will reflect what is already in a dataset, so, if the dataset contains inherent or existing bias within society, the algorithm will reflect that. That inherent bias is what we call potentially representative bias.

There is a second kind of bias, which is allocative bias—in other words, how the algorithm will be used in terms of the results, recommendations and so on. As a technology company, the way we look at it is that, as long as we can help to develop tools that point out the lack of representation within the algorithm, it will at least air, or bring awareness of, potential issues in the use of the algorithm. Making that clear to the people who will be using the algorithm can help to bring about what decision is needed as a company, because bias depends on the culture of a society. As a global company, we have to be very mindful of that. As long as we can share that information, we believe it will help to address the issue.

Another part of it is to make sure that multiple perspectives, insights and so on are represented during the development, deployment and operation of the algorithm, to mitigate some of the challenges that may be represented.

Martin Wattenberg: Recently, a very interesting case was reported by ProPublica and others about the use of algorithms in decision-making in the criminal justice system. I think it illustrates many of these issues in an interesting way. The idea is that we have a system that creates recidivism risks algorithmically, and you can point to many possible issues related to that. One is that you could ask whether the training data were biased. It was based on rearrest rates rather than recidivism rates. Is that proxy bias? You could argue that perhaps it is, and that could be reflected in the system.

A second very interesting thing came out after several computer scientists began analysing the system. They pointed out that there were inherent trade-offs mathematically in the types of fairness you could get, given the structure of the data. It would be possible to come up with

criteria for fairness all of which feel very intuitive and are things you would want, but it was mathematically impossible to get all of them at once. This illustrates some of the nuance in setting out principles for what we want these systems to do.

A third thing is that there has been follow-up work by Jon Kleinberg and some collaborators. They looked very carefully at the dataset and compared human judges making similar decisions. This is a very tricky area, which requires some very clever statistical analysis, and we have to regard this as preliminary; but in their paper, they make the very interesting argument that, if you include algorithms in the process, there is the potential to get better outcomes. Of course, human judges also have their own biases in various ways, including societal biases. There is some interesting research that says parole boards make harsher decisions when they are hungry, so humans too are fallible. There is interesting potential, as pointed out by that paper, for algorithms to help humans work together to create a better outcome.

Q97 **Chair:** Your reference to the criminal justice system raises another issue. There is a lot of debate about whether there should be regulation or whether there should just be self-regulation, and whether it should be by sector or by an overarching regulator of some sort. What are your views on those two issues? Should there be regulation or not, and should it be by sector or overarching? Charles, you have already alluded to sector regulation.

Charles Butterworth: As a company, we are regulated by the FCA, and clearly that provides principles, rules and frameworks that are well understood by a number of parties in the sector.

More broadly, there is a huge proliferation of available data and datasets becoming available. How they are interpreted and how people use those datasets for conducting business or providing outcomes to individuals is ultimately a question for codes of conduct. Then the question is whether it moves into regulation. For financial services, there is a well established regulatory path, so from my perspective we work within that framework and are happy to do so.

Q98 **Chair:** Are there other contributions?

Dr Nguyen: AI is already being regulated, because it is applied in different sectors—for example, in the healthcare sector—and there are existing financial regulations in terms of non-discriminatory behaviour. There are also privacy regulations on the use of data, so all of those existing regulations apply.

With respect to regulation on some of these issues, because it is indeed a new technology, we believe that at this point it is right to look at best practices for the principles that need to be put in place on how data can be accessed or used. There is also room for principles and self-regulation. However, we believe that specific regulation needs to be sector-based to

follow the existing trend of sectoral regulation, and, as usually put, we believe regulation should be technology neutral, as opposed to technology specific.

Martin Wattenberg: My colleagues have made the point very well that a lot of laws already apply, and of course we comply with them. Sector-by-sector regulation makes much more sense in general because algorithms are such a broad concept. It is as broad as a business process. Because they apply in so many different situations and at so many different scales, with so many different stakes, the right thing to do will always be extremely contextual, and having a framework that recognises that context is critical.

Q99 **Chair:** Nick, do you have anything to add?

Nick Pickles: I have two points. First, businesses can do more to be transparent. One of the questions we think a lot about is how we better inform our users. I can share this with the Committee afterwards in writing. It is not strictly an algorithmic change, but we went through a huge amount of data-driven testing before we made the change to 280 characters. We published the data on that testing so that we could inform our users, "This is why we made a change." In this space, businesses being able to explain to users, "This is how your data is being used and this is how you can access it"—that transparency—is an important part of our responsibilities.

Secondly, it is obviously something that will be affected by GDPR. Algorithms rely on the data being fed into them and those data are regulated, but the Government have also identified that this is an area where wider ethical and policy discussion is definitely needed, and investing in world-leading thinking on these challenges is something that I am sure all our companies are very keen to be a part of.

Q100 **Chair:** Social media have played a big role in elections and referendums in recent times. The issue has been very much in the news. Do you share the concerns some have raised about whether algorithms that filter adverts and news stories could undermine democracy in any way?

Nick Pickles: This is something we have prioritised as a company and have taken incredibly seriously. Earlier this year, we established a dedicated information quality team to look at these sorts of issues, and we are co-operating with the Electoral Commission and the DCMS Select Committee, as well as ongoing investigations in the US. We responded to the Electoral Commission on Friday with some results of our findings.

Different platforms work in different ways. In Twitter, we are not using an algorithm to show you content; you are seeing content because someone you have chosen to follow has retweeted that content, or posted it themselves. We have heard before the reference to the societal way this works. People tend to follow accounts they agree with, in the same way as we tend to read a newspaper we agree with, or socialise with people

who do not hold differing views. I suspect my Christmas table will not be the only one where certain topics are not for discussion over Christmas dinner, to keep the food on the plates.

Our big question as a company is, how do we make sure people see as broad a perspective as we can? What has become very intrinsic in Twitter is the hashtag. You might click on the hashtag Budget 2017. That is not taking you to accounts that you have chosen to follow; it is taking you to everyone sharing an opinion on the Budget, so it could be journalists, different political supporters, different politicians, official Government messages and analysis by independent experts who focus on the detail. The way we see hashtags being used is very important. As a company, we think a lot about how we can show people information around topics, rather than necessarily saying, "You're just following these accounts and we're going to show you more like them." That is absolutely critical in how we maintain robust democracies and protect them from outside interference, and to protect civic debate from becoming polarised. That is very challenging in a political environment where it is polarising, but we are spending a huge amount of time investigating it and working with researchers to understand what good solutions look like.

Q101 **Stephen Metcalfe:** How many Twitter users are not what they might seem—not real people, but bots retweeting other people's stuff?

Nick Pickles: We estimate that less than 5% of our users are bots; these are not included in the numbers we declare to the stock market. It is also important to say that there are bots that, for example, monitor Wikipedia edits, or post readings from air quality sensors, so the perception that any automated account is per se a bad account is more nuanced than that. It is a very small proportion of the traffic we see around issues such as elections or big public events.

Q102 **Stephen Metcalfe:** Of those retweeting about important social issues only 5% are bots.

Nick Pickles: Less than 5%. I am happy to share more specific numbers on that.

Q103 **Stephen Metcalfe:** It is an interesting area. I want to go back to discrimination and unwanted bias, the potential causes of it and, therefore, how one might eliminate that or take steps to stop it. How big a problem do you think unwanted bias and discrimination is, and what is the cause in your view? I would welcome brief comments from all of you.

Nick Pickles: From our point of view, we need a more diverse workforce. In particular, as a company we have made a very clear commitment to publishing data on how diverse our workforce are, because an algorithm is made by people and the bias will be transferred, so for us wide perspectives around people working on these issues are critical.

Martin Wattenberg: The sources we have heard raised here are training data and so forth. That is very important, and it gives us clues for ways

to fix these things. One thing my group has done is release an open source tool called Facets that helps people inspect their training data to look for potential problems.

Dr Nguyen: We try to address it in two ways. First, when we look at data being used for training and for evaluating algorithms, we try to see how complete it is; we try to look at it from multiple perspectives and see whether it is representative and whether there are different sources of data involved, because sometimes using multiple sets of data will balance the sources.

Our researchers are engaged in developing the kind of tools Martin spoke about. For example, there are certain words associated with gender. You can measure whether the use of certain words will lead to discriminatory outcomes; "receptionist" may be more associated with women and "doctor" more with men, so if we can identify and measure the association, we can put in tools, or at least raise these as potential issues. As I mentioned before, as a global company it is very difficult to say what bias is across the board, so the best we can do is try to give measurements for these kinds of issues.

Charles Butterworth: Alongside codes of conduct and best practice, the introduction of GDPR has a number of pieces in it that try to address the avoidance of discrimination. We then have all the testing. You have heard several comments on how data are prepared and managed. On top of that, I would add focus on outcomes, rather than specifically the technologies, because there is a raft of different technologies that can be used. If we are thinking about consumer outcomes, we can test to make sure that the outcomes that are being created are really good ones. That is a very good way of making sure.

Q104 **Stephen Metcalfe:** How do you do that in your business?

Charles Butterworth: Ultimately, within the regulatory framework, one of the key tenets of the FCA and the way it works is looking at consumer outcomes. An example would be that, whenever we are creating a new product, we go through a process of looking at what outcomes the product would have and testing them against different variables, so we ensure that what we are building does not have adverse outcomes, and that detriment is not caused in the process. That is a requirement; we have to be able to evidence it. If the FCA came to have a look, we would have the processes to show that we had been through all of that.

Q105 **Stephen Metcalfe:** That applies to all the products you provide.

Charles Butterworth: Yes.

Q106 **Stephen Metcalfe:** One of the things you do is verify people's identity for use on gov.uk websites. When there is a failure to verify people's identity, what process do you have in place to assist them through that? When the computer says no, how do they challenge you? When they say, "Actually, I am the person I say I am," what is the process?

Charles Butterworth: We have call centres. Whenever consumers have issues with our products and services, they can call in and challenge that process. Under GDPR, there is strengthening of the right of individuals to be able to challenge what data we and other businesses hold on them and how it is being used, so data accuracy is a key tenet of that. Obviously, one has to be careful; we have to be able to weed out fraudulent challenges from real challenges, but that is part of the process we go through. Effectively, it needs to be through online education, which we provide on our website, and at call centres where people get answers to issues they may find.

Q107 **Stephen Metcalfe:** Do you rerun the process without the use of an algorithm so that you go back to a human-based system, or do you just say, "That's the outcome of the algorithm. We're sorry, but we've failed to verify your identity"?

Charles Butterworth: We do not publish the algorithms themselves. When we work with businesses and build algorithms for scoring for a lender, we provide detail to the lender on the algorithm itself. For consumers, we do not publish the algorithms. We think, and it is well established in the industry, that that could lead to misuse and fraudulent use, so our emphasis is very much on education about the key drivers of the algorithm, how it is being used, what things can impact, whether it is identity or your credit score and how you can improve it, and so forth. That is the focus of the industry.

Q108 **Stephen Metcalfe:** I had a constituent who failed to get their identity verified and found it almost impossible to find out why, other than that they failed to have one piece of documentation, which happened to be a passport. They could not get an answer from you about why that was and what the problem was in verifying their identity. They applied for a passport, reran the process and got their identity verified. The first problem is about challenging the outcome generally. Secondly, if there are gaps in data, how do you address them? How do you tell people, or find alternatives to data gaps? I suppose that applies to all of you.

Charles Butterworth: First, I am sorry your constituent had an issue, but for anybody who is holding data that are used for these kinds of things—we are a company like that—it is incumbent on them to make sure they constantly check data quality. As described, there need to be good processes that allow challenge. In our industry, that challenge was in existence prior to GDPR. GDPR emphasises several pieces of it, but effectively within the credit bureau you could challenge the accuracy of the data. We follow a regulatory process within a given timeframe, when we investigate the question and provide an answer, and, if it is an honest mistake, it can be rectified. We have requirements to follow that process.

Q109 **Stephen Metcalfe:** Does anyone want to make further comments on data gaps?

Nick Pickles: From our point of view, often the data gap is context rather than data, so one of the things is recognising that often you cannot machine-read intent just from text. In our case, the data gap is asking our users, "Please explain why you are reporting this." They can explain the context around it. That is a human process, not an algorithmic one. It is important to recognise the limitations of algorithms. When analysing intent, it is very difficult to build in sarcasm, for example; it would certainly challenge my own Twitter account. It is about recognising where you need more human data and less reliance on algorithms.

Q110 **Stephen Metcalfe:** Some forms of discrimination and bias are illegal and defined in law: gender bias, racial bias and so on. Are we taking the issue seriously enough? This is a developing and emerging technology, so do we need to do more at this early stage to ensure that everyone is taking it seriously?

Martin Wattenberg: We are taking it very seriously. At Google, we apply a comprehensive approach to find ways to address it. One is that we are putting significant resources into research to discover new ways to address the problem. About a year ago, there was a very interesting paper from a member of PAIR on a way of attacking discrimination in machine learning; it would let you take a black-box system, or something there was very little information on, and repair the system in certain ways and enforce constraints on bias. There is another very recent piece of research out of Google in which you can take two machine-learning programmes and have them engage in a little bit of a struggle, where one is trying to show that the other is biased in some way, and the other is trying to avoid that. I think that holds tremendous promise. That is an important theoretical aspect.

At the same time, we are doing practical testing. We are also trying very hard to educate around this. We have published web pages with interactive diagrams that help make clear the scope of the issue and the kinds of things involved, so we take it very seriously.

Chair: Can we keep answers brief, please?

Dr Nguyen: Absolutely. Just to reinforce that, we are very serious about it from the perspective of both research and practical engineering guidelines. For example, our researchers, with other academics, have established and are engaged in a fairness, accountability and transparency in machine-learning conference. Along with other of our colleagues, we have established an organisation called Partnership on AI, with the idea of sharing best practice on how to address this. Internally, in terms of engineering guidelines, we have formed an internal senior executive-led organisation called AI and Ethics in Engineering and Research—Aether—so that we can discuss and be proactive in the development of our systems as well as in our consultation, and work with customers around these kinds of issues.

Q111 **Martin Whitfield:** We have talked about bias and, probably as a result of discussions about bias, transparency in relation to the algorithm rather than possibly the dataset in the first instance. I would like to explore that. We seem to have one side that says, "Open up the algorithm so everyone can see all of it," all the way to the other end, which is, "I won't explain anything," or, "I will explain what's happening in the black box." Could each of you say where you stand on that rainbow of opportunity?

Nick Pickles: From our point of view, which is probably slightly different from more consumer-facing products and decisions on things such as healthcare, where there could be real consumer detriment, there are two main things. The first is being transparent with users about the outcomes of the algorithms, as we have heard, making a change so that users can understand exactly why they are seeing certain advertising. To the Chair's question on elections, we have announced a transparency advertising centre where we are going to disclose more data on the types of adverts that you have seen and why you might have seen them. One of the things is, hopefully, informing people about how their data are used, which helps them make better choices.

Q112 **Martin Whitfield:** Maybe I can draw you in. Rather than the actual dataset, I am talking about the actual algorithm itself and your views on opening that to all.

Nick Pickles: I am sure colleagues can speak more to the intellectual property side of things. One of our big concerns is that we use algorithms in a defensive way to protect our users, and the risk is how you stop bad actors gaming your algorithms. In our case, we put a huge amount of resource into stopping the kind of automated accounts that game our trending topics. But, by making those defences public, you inform the bad actors, so that is a big concern for us.

Martin Wattenberg: There are a couple of issues around transparency. It is a fascinating area. One is simply asking what you want in terms of an explanation. With some of these new algorithms it can be very hard to know the right level of detail. Right now I am picking up this glass of water. You could ask why I did that. Absolute full transparency would be to give you the exact chemical and electrical state of every neuron in my brain. That is not really useful. Instead, you would want an explanation such as that I plan to take a drink. The same is true of many of the algorithms we are dealing with. With some of the new technology known as deep learning, the form of an algorithm could be many matrices each with millions of numbers, and it is hard to know what you can make out of that.

Dealing with bad actors is an important issue, and there are certainly intellectual property issues as well. It is very important that we have some view as to what is happening. I will give examples that we have for our users. We have a My Account page that people can visit to understand how their data are being used. It does not sit there unused; we had more than 1.6 billion visits to that page in the past year. In

general, there is a balance in the level of transparency. Full transparency may not be possible; it may conflict with issues around privacy and security. At the same time, you might want some view as to what is going on. Instead of thinking about algorithmic transparency, I would promote the idea of algorithmic translucency. It is like frosted glass in a bathroom; it lets in light but not certain details.

Dr Nguyen: In terms of transparency of the algorithms, we believe that the primary purpose in this particular case is to enable understanding, and that is not done by code-sharing of data. It takes a lot of data scientists to understand exactly what is going. Even then, it is impossible, so we very much believe in the principle of explainability. How is the algorithm trained? What was used to test it? What were the criteria for validating the algorithm? How will it be maintained once it is deployed, and what are the feedback and monitoring mechanisms? Having that kind of information available will tell us a little more about how the algorithm was derived and developed and how it will be used.

Another part of transparency of the algorithm is what the probability is of certainty with respect to the recommendation, and what went into that recommendation. For example, if an algorithm says a particular loan is being rejected, what are some of the reasons? If it is 80% certainty, what made it 80% certainty, and what is the remaining 20%? The notion is that AI and machine learning are tools that need to be implemented and used by people with full information, so it is very much about how they are used, and the outcome of the algorithm.

Chair: Can you keep it tight? I am conscious of time.

Dr Nguyen: Sure. I am sorry. Another part is that from a technical perspective an algorithm can be very good. What that means is that it can predict the outcome in training datasets very well, but it may be completely opaque, so are there other technologies we can use to develop parallel algorithms that will get close to what the dataset says but are more understandable? Those are the different ways we look at transparency of algorithms.

Charles Butterworth: Under GDPR, there is an enhanced requirement for data controllers to be able to demonstrate compliance with GDPR provisions, of which transparency has elevated itself. We think that is pretty healthy in overall terms.

As a business in the provision of algorithms to other businesses—for example, to lenders—we have an audit trail. Either third parties can come and audit or we provide all the underlying algorithmic data and variables to those customers, so that would not be an issue.

To pick up what I mentioned before, there is an issue about exposing the underlying algorithm for the end consumer with a credit score, because it allows gamification of the system and that would be to the detriment of the credit industry. That has to be the point at which we look to

explainability and transparency to help people understand the drivers of the algorithms and how they can impact positively, in this case on the credit score.

Q113 **Martin Whitfield:** In essence, to understand it, as long as the user knows enough about the algorithm and what is being input and the data source, you are satisfied—I use that word carefully—about the actual writing of the algorithm. Is that a business decision, because there is obviously innate value in the algorithms you create, or is there some other reason why you want to protect the coding within the algorithm? It is interesting that a number of witnesses have said to us, "If someone knows about the coding, they can game it," but similarly we have heard evidence that these things are so complex that no single person understands them anyway. I know there are different types of algorithms, but is there a single answer for algorithms, or is it dependent on the situation? Is it business protection or something more, or less?

Martin Wattenberg: It is many things; it is very contextual. You have put your finger on one of the issues, which is that different algorithms have different properties. Knowing in a clear way how certain algorithms work can help you game them. An interesting example from Google's history is the PageRank algorithm. The initial algorithm that helped to launch the company was an interesting way of looking at web pages and saying that they are authoritative if other authoritative pages link to them. It sounds circular, but there is a way of breaking that circular logic, which is essentially the insight that started the company. The founders were academics; they wrote a paper about that. However, knowing the algorithm and understanding how it worked let people like spammers create fake sites—networks of synthetic sites that linked to each other in such a way that it allowed the system to be gamed.

Since then we have learned ways to deal with that, but we keep those very close to the vest because we remember the history. At the same time, there can certainly be other situations where different measures are appropriate. This is a great example of just how contextual these questions are.

Q114 **Martin Whitfield:** Many answers for many problems. The patent industry obviously seeks to protect algorithms. Do you think that it is fair to consumers to rely on that protection by way of safeguarding your commercial interests but guaranteeing a sort of opening into it, if it was needed?

Martin Wattenberg: Again, the PageRank algorithm is a great example. It was patented but that did not help, so multiple protections may be needed.

Q115 **Martin Whitfield:** I know we are pressed for time, so I will push on to the last part of my question. Is the most important thing a right to an explanation of a decision, thus bringing together the dataset, the algorithm and the lending decision or journey, or whatever? Is the right

to an explanation, which has been raised in the Data Protection Bill, where the answer lies?

Charles Butterworth: The GDPR provisions, as presented, go down that route, which is providing the data subject access to much more transparency about how their data is being held, how it is being processed and whether it is being used in profiling; and the right to erasure. That really talks to looking at it from the perspective of, "How do we protect consumers?" rather than going after the notion, "Let's look at whether the algorithm is bad or good per se," because the real issue is how it is set up. Looking at outcomes, and being outcome-based on that GDPR set of provisions, is a very healthy way for us to go forward in this space.

Q116 **Chair:** Are you able to give an explanation if the algorithm has given the answer?

Charles Butterworth: To the extent that we do not publish it but explain what factors are influencing individuals and how, in our case, they could improve their credit score, we can explain to people how that works. Under the GDPR there are a number of different provisions, including in our industry the fact that you do not have the right to erasure. For example, in the credit bureau, you cannot just say, "I don't like my credit history; I am going to have it erased." What they are trying to do is build a transparency arena where people have much greater access to understanding where their data is and what is happening to it in any of the businesses. Transparency is probably the most significant raised bar, versus other things, because quite a few of the provisions were in existence for a number of industries. That transparency thing is what we are looking for.

Q117 **Bill Grant:** This is about accountability. As a lay person, I sense that algorithm use is, or will be, global and it will not recognise borders unless it receives precise instructions to do so. In that context, how do you think accountability for the use of algorithms should be exercised, if you feel it should be? How would you achieve that, and to whom would you be accountable for the use of algorithms, if that makes sense?

Charles Butterworth: In the context of Experian, we have two. We are accountable to the FCA and are regulated by it. Hence, from that perspective, as a supervisory body that is precisely where I would look. I think the GDPR is looking at the inputs: is the data accurate and is it being processed in a fair manner? I think the ICO and the FCA, working in tandem as they do, form a very good regulatory and supervisory view of the use of algorithms in our industry.

Q118 **Bill Grant:** That is with us at present; it is already there.

Charles Butterworth: That is with us at present. The GDPR regulation is enhancing a number of pieces. It basically updates the Data Protection Act of many years ago. It comes into force, assuming all of it is passed,

at the beginning of next year. That enhancement is coming. The FCA has been regulating since 2014, so that is absolutely there today.

Q119 **Chair:** Is the ICO in any way capable of considering discrimination issues? Does it have the skills to do it?

Charles Butterworth: That would be a question for the ICO.

Q120 **Chair:** You are saying that the system is good and effective.

Charles Butterworth: I am saying that the provisions of GDPR are trying to avoid the kind of negative consumer detriment outcomes we are discussing today. How that is done in practice is a different matter. The GDPR will put additional volume of activity, information and requirements on the whole system.

Q121 **Chair:** Are there any other brief responses to Bill's question?

Dr Nguyen: This goes back to the earlier question about sectoral regulations.

Q122 **Bill Grant:** Carolyn, earlier you touched on sectoral-based and technology-neutral regulation. Could you expand your thoughts on sectoral-based and technology-neutral regulation?

Dr Nguyen: You mean examples of existing sectoral regulations.

Bill Grant: Yes.

Dr Nguyen: If I may use US examples, in the US within the health sector there is HIPAA, which regulates privacy on the use of healthcare information. That is an example of sectoral regulation that would apply to how algorithms were used. In the financial sector, there are regulations that prevent discriminatory behaviour in loan processing, for example. That would be another way in which an existing piece of regulation would apply to the application of outcomes of algorithms in machine learning. Within the automotive sector there are clearly regulations with respect to safety. That is another example of sectoral regulation in the application of algorithms.

Q123 **Bill Grant:** Is there an element of self-regulation rather than a governing body?

Dr Nguyen: Yes. I think it would have to be a combination of the two. For many of the harms you are bringing up here with respect to fairness, bias and behaviour, there are existing sectoral regulations, but for some of the potential principles that should be applied in the development of the technology, specifically how the technology is applied, we should consider best practice and self-regulation—for example, fairness, safety, reliability, transparency and accountability, and inclusiveness, which has not been brought up here—so that we see those working together in combination.

Q124 **Bill Grant:** Are there any other contributions?

Nick Pickles: We and various others in the industry signed a letter last week recommending and urging the Government to make sure that the ICO remains part of the European Data Protection Board post Brexit. As you said, this is a very international issue. The ICO is part of that European framework, and discussion is very important.

Q125 **Bill Grant:** The recent industrial strategy White Paper promised investment of £9 million in a centre for data ethics and innovation. Do you see that complementing the private sector or competing with other sectors? Indeed, if it goes ahead, what should be its main objectives or aims once it is established?

Martin Wattenberg: I was glad to see that announcement. It is excellent to have a societal conversation around all these issues. I have two points about what I would like to see. It would be good for it to be as broad-based as possible. It is important to bring not just Government, industry and academics into the conversation, but non-profits and artists who are using machine learning. There are artists who use machine learning. Bring in all of society. It is very important that every stakeholder has a voice.

Q126 **Bill Grant:** It would be inclusive.

Martin Wattenberg: Exactly. Secondly, it can be a force for helping innovation. We see today that in the machine-learning world the sphere of knowledge is expanding very fast. I have been involved in many different academic disciplines in my life. How fast the frontier of knowledge is advancing today is wonderful, and I would love to see that continue, or perhaps even accelerate.

Bill Grant: Thank you for your enthusiasm.

Q127 **Vicky Ford:** I am sorry for being late for the session. Nick, could you repeat what you said about the ICO staying part of the European network? Who has called for that? Are there any other third countries in that network? What other international bodies do you think the ICO should be working with?

Nick Pickles: I am happy to share the letter. I think it was co-ordinated by the Advertising Association, so it was mainly businesses in the advertising space that signed it. I will arrange for a copy to be sent to the Clerks afterwards.

The international conversation follows very nicely from the question of how the Government can bring in other people, and the approach of different countries and different cultures to data and privacy, which we certainly see around people using their own name for example. Different countries have taken different approaches on a huge range of issues. The more international this framework can be, to bring in different international perspectives, the more it makes the UK a stronger place to base international businesses.

Q128 **Vicky Ford:** I understand that. I just asked you, which other third countries are on that board, and are there other international bodies we should be on?

Nick Pickles: I think the European Data Protection Board replaces the article 29 working group, so it is a direct crossover from the existing GDPR framework. I think it is intended to be for European Union countries, so it will be for the UK to join and retain its position. The ICO has been the deputy vice-chair in recent years; the previous commissioner was. It would be good to have the US involved, given the importance of the US to the digital economy.

Q129 **Chair:** The Government have not given any view on the purpose of your letter yet. Is that right?

Nick Pickles: It was only submitted last week.

Q130 **Carol Monaghan:** Carolyn, you mentioned best practice a number of times and the importance of ethics. Regardless of that, there is still suspicion among the general public about the bias that algorithms can have. Could audits of the way algorithms work help with accountability and bias?

Dr Nguyen: When we think about audit, there is a spectrum from self-audit or certification all the way to third-party audit. Along that spectrum, part of the question is what criteria would be used in third-party certification. The criteria for validation and testing would have to be very clear. Given where the technology is at this point in time, it is not clear that those criteria are well understood. This is a case where, given the current state of development of algorithms and AI, it would be better to look at best practices, in combination with the kind of things that we are trying to do internally, which are to put in place engineering guidelines, engineering processes and internal understanding, and sharing what we are doing with others in industry who are doing exactly the same thing.

Q131 **Carol Monaghan:** Would anyone else like to comment on that? One of the pieces of evidence we had was from the Alan Turing Institute, which supports certification mechanisms to test algorithms. In order to give people reassurance about the use of algorithms, is there a place for certification of the algorithms?

Martin Wattenberg: Again, it would be sector by sector.

Q132 **Chair:** Taking your example of the criminal justice system, is that an area where you think some sort of certification or auditing might be appropriate?

Martin Wattenberg: Yes, absolutely. That is a very interesting example. If you have an algorithm in the criminal justice system, it is very important that a defence attorney is able to challenge it appropriately. There should be a very high level of scrutiny. In some ways, the analogy

is whether it is appropriate to certify machinery. For some kind of machinery, yes; maybe not for others.

I would tie that to the question about auditing in general. On the possible spectrum of auditing mechanisms, there are ones that are very open where you use existing APIs or publicly available products that are less intrusive and very easy. There is a whole other end that involves very intrusive, and perhaps difficult to implement, mechanisms that allow in third parties. That needs to be thought through very carefully, because there are implications for privacy and security in letting in people who are not part of an organisation. It is important to think through in a global environment how those mechanisms might be used in an authoritarian Government, for example. A whole set of consequences needs to be looked at very carefully, and it is highly contextual.

Dr Nguyen: I want to go back to the earlier conversation about explainability in terms of what information is already available, or should be made available, with respect to how the systems work. It also comes back to what should be certified. For example, there are conversations with many of the standards organisations about whether you can certify the actual development process of the system. That is a very different conversation, which could be addressed through voluntary standards organisations. There is also, potentially, certification of the outcome. That is a very challenging way to define what should be certified. Can the outcome be exactly specified, because it is contextual? It comes back to what should be certified and what issues we are trying to address.

Q133 **Carol Monaghan:** One of the issues is how personal data is used by the algorithm; people need to know that it is being used in an unbiased way. Martin, can I take you back to the criminal justice system? We could take instead the healthcare system, insurance sector, or whatever sector. Who would do the audit of a particular sector?

Martin Wattenberg: That is an excellent question. Let us take criminal justice and what you might ask for if something is being used to make consequential decisions about a person's sentencing. There may be some sectors where we would require that a human be heavily involved in a decision. I do not think we would like a situation where criminal sentencing happened completely automatically. If an algorithm was being used to help in those situations, it could be a case where perhaps the precise code should be transparent. This is a policy question that needs to be worked out in a larger forum, but in something like the criminal justice system it is important to explore all options.

Q134 **Carol Monaghan:** Off the top of my head, I am thinking of healthcare. If I want to get healthcare insurance and the algorithm is tying into my purchases in the supermarket and knows that I buy red wine every week, is that going to affect my insurance? That is why I am asking who should audit the particular algorithms for a sector.

Martin Wattenberg: That is an important issue. I am not an expert on the insurance industry. It is now highly regulated. Perhaps some of those regulations could apply as is, but, not being an expert, I do not know.

Charles Butterworth: There are existing audit rights in several sectors. Financial services is an example where there are rights to audit, but I am sure that is not the case across all sectors.

Carol Monaghan: I'll buy my wine with cash.

Q135 **Clive Lewis:** It is time for everyone's favourite: big data. It follows on a little bit from what Carol said. What development is each of your organisations pursuing in algorithms and big data, and what challenges do you think you will face on consent? For example, Google DeepMind had a problem with its kidney analysis app recently and the NHS. Perhaps you could talk me through where you are and what you are doing on that, the challenges you are facing and how you will overcome them.

Martin Wattenberg: Before I start, one thing I want to make clear is that I work for Google. DeepMind is a separate unit within Alphabet, and it is run autonomously.

Q136 **Chair:** You deny all responsibility.

Martin Wattenberg: I cannot speak in detail as to what happened there.

Q137 **Clive Lewis:** You can plead the fifth.

Martin Wattenberg: This is the framework in which I would think about questions like this: as the NHS is deciding on partnerships, their data is very sensitive and a lot of careful balances have to be struck very thoughtfully. However, it is important to weigh the possible risks against the possible benefits. One of the things in the health space right now is that there have been really impressive advances in machine learning. They have come from academics and from industry. It is really dramatic. We have seen cases in which, for example, a machine-learning programme can rival human pathologists, even if the pathologists have been given an unlimited amount of time to look at a slide. There is real potential benefit. I am not talking about monetary benefit, but saved lives. That is the trade-off that needs to be made.

Q138 **Clive Lewis:** There is a monetary benefit, maybe not in this country potentially, given the nature of the NHS; but for some organisations there is a monetary benefit, isn't there?

Martin Wattenberg: Yes, and I am sure that is something they need to take into account. In general, these issues need to be thought through carefully. There are some potential benefits to partnerships. If one organisation can partner with another and perhaps speed up the delivery to patients of technology that can save lives, that is one big factor to take into account.

Q139 **Chair:** Are there other brief responses? We are tight on time.

Charles Butterworth: We think about big data in respect of the new big datasets that come into the arena and what we do with them. An example at the moment is open banking. The push the Government have made on that will make available a whole bunch more data than was previously available, which we and others are trying to work with to answer some of the questions that have been posed—for example, around affordability. If you ask where the industry is pushing towards, it is the use of data and data analytics on new datasets that become available as a result of Government policy, and looking at how we can push forward and innovate in that space to provide affordability-type assessments that then go back into the industry.

Dr Nguyen: I put three things on the table. One of the issues we are looking at is to try to promote anonymisation of data. There are purists who say that anonymisation will never be able to prevent de-identification. However, we believe very strongly that it is akin to using a lock on a front door. A lock will not stop all break-ins, but the use of a lock plus a law against break-in will take us a long way, so we are very clear on the use of anonymisation techniques.

The second point is about the contextual use of data—for example, a piece of personal data like gender. That is very sensitive, but not in all applications. That is another part to think about.

Thirdly, in terms of text data mining, it is very important to make sure that data can be available. That is getting into some of the copyright issues that currently exist, where the UK is not quite up to par yet against other countries such as the US, Canada and so on.

Nick Pickles: We are in a slightly different business. When our algorithms mis-perform, one of the implications is that we interfere with someone's free expression, and we may remove their speech from our platform. False positives are things we have not really discussed, but they are important. When you run an algorithm, mistakes will be made. It is about recognising that some decisions should always be made by a human, not by the algorithm, but the algorithm might surface something for a human to review. We are taking action on about 10 times as many accounts for safety violations by using algorithms to surface quickly more accounts to our agents. It is a matter of acknowledging that there will be a false positive rate, and building in from the beginning of the process how to deal with those false positives, and how we are transparent in accepting when they happen so that people understand why they happen. For us, text analysis is still a very early field in this area.

Q140 **Clive Lewis:** You have touched on my last question. Is the law on consent sufficient in this country? The Data Protection Bill is going through Parliament at the moment. Is there scope for changes to it? Could it be more robust than it currently is to deal with big data and consent?

Charles Butterworth: The provisions being contemplated have six ways of processing data. The ICO and the Commission have been very clear that consent is not the only way in which data can be processed fairly. There are legitimate interests. There are six different ways of processing data. We think it is a robust piece of legislation, and therefore we support it. We are not calling for any addendums to be put in because we think we are missing a whole chunk of the ecosystem.

Q141 **Stephen Metcalfe:** Nick, I want to pick up a point you made about the 5% of Twitter accounts being bots. You have just talked about protecting people's freedom of expression, but a bot is not a person with freedom of expression. How many thousands or millions of accounts does that 5% represent?

Nick Pickles: It is 5% of 330 million. This is where maths is demonstrated not to be my strongest skill.

Q142 **Stephen Metcalfe:** Is it 16 and a half?

Nick Pickles: To put it in context, we remove about 3.2 million accounts every week for violating our API rules, our automation rules.

Q143 **Chair:** It is a massive issue, isn't it?

Nick Pickles: It is a very important issue; it is something that hostile actors seek to game, which is why one of our challenges is that some third-party research does not see our defences, because if we made them visible hostile actors would be able to understand how they work and game them. We catch about 450,000 accounts every day that have suspicious log-ins. I can give you the exact data. In a three-month period this year, we removed about 100,000 applications that were using Twitter's API and were collectively responsible for 1.8 billion engagements. We have removed a huge amount of content to try to limit these actors, and we are seeing the results of our work.

Q144 **Stephen Metcalfe:** Because those bots are more prolific, are they?

Nick Pickles: Generally speaking, bots are used to post high-volume content and that is where people have identified them. It is not always the case. Last week, an individual in Scotland was accused of being a bot by somebody and then volunteered to the *Daily Record* that he was actually an upstanding citizen who was tweeting a lot. Some people are very enthusiastic about issues and tweet a lot themselves.

Q145 **Stephen Metcalfe:** You talked about using algorithms to detect safety violations as well. What constitutes a safety violation?

Nick Pickles: There are different ways of doing it. One is that we receive reports from users all over the world. The order in which we look at those reports is prioritised by use of different signals; some of it, which you may have seen, is by putting in warnings. We take a reply to a tweet, lower it in the rankings and say that the tweet may be offensive or

sensitive, and we hide it. We have a huge range of policies in place to determine what activity is allowed on our platform.

Q146 **Stephen Metcalfe:** You will be aware that today there is a story running on the front page of a national paper about individuals sharing what probably most people would describe as inappropriate thoughts about sexualising children—paedophilia. Is that not a safety violation? Should you not be taking a more active role in searching that out and closing down those accounts? They may not be acting on those thoughts, but most of us would find it deeply disturbing that this is freely available. It may be their freedom of expression, but I am not sure we want to read or see it.

Nick Pickles: Absolutely. We have a zero-tolerance approach to child sexual exploitation, and these accounts were not engaging in that. We did seek advice from law enforcement around this activity. We have a very close relationship with law enforcement, and we regularly provide data to bring to justice people who have engaged in offending. We continue to keep our policies under review as behaviour changes, but with regard to that specific story, I think some of the details were not on Twitter; they were on another service, and there has been a degree of conflation between the two.

Q147 **Chair:** You mentioned that bots accounted for less than 5% of the traffic. One in 20 is quite a lot, isn't it? Particularly with regard to democracy and the conduct of elections, do you think there is a need to reconsider the regulation around campaigning through social media, given the scale of the problem?

Nick Pickles: There are different things. Platforms have a responsibility to take down actors who are using bots to try to make a topic trend, and to remove malicious automated accounts. We are investing very heavily in detecting those. Interestingly, academics say that behaviour is already changing as a result of that investment. In the UK, there is a proposal in the House of Lords; in the US there is the Honest Ads Act. We have already committed to the idea of an advertising transparency centre so that people can see the content that is advertised by political parties. Interestingly, I was at an event along this corridor where it was highlighted that political advertising was exempt from advertising regulation by choice, so there is a whole range of very important issues we need to work on in collaboration.

Q148 **Chair:** They perhaps merit consideration.

Nick Pickles: Absolutely.

Chair: Thank you all very much indeed. We appreciate the time you have spent with us.

Examination of witnesses

Witnesses: Sheena Urwin and Professor Bowers.

Q149 **Chair:** Good morning, both of you. Thank you very much indeed for coming. Would you mind briefly introducing yourselves and saying where you are from?

Sheena Urwin: I am Sheena Urwin, head of criminal justice in Durham constabulary.

Professor Bowers: I am Professor Kate Bowers from the Jill Dando Institute of Crime Science at UCL.

Q150 **Chair:** Can you both try to make sure that you keep your answers as succinct as possible so that we can get through everything we need to do? The first question for both of you is: how are algorithms being used in this country by the police and criminal justice system generally? Sheena, you obviously have direct experience from Durham.

Sheena Urwin: I do not have experience of the entire policing of England and Wales and how different forces are using algorithms. We have a specific example in Durham constabulary in the form of the harm assessment risk tool, which is part of a research programme we have undertaken in collaboration with the University of Cambridge. That tool assesses the risk of serious and non-serious reoffending, as well as no reoffending at all, over a two-year period. The research programme is called Checkpoint.

Despite the narrative that appears to be out there around bail or custody, the actual purpose of the model is to support decision-making in relation to intervention around rehabilitation and reducing reoffending, trying to improve life chances. Can we encourage some suspects and offenders who have more chaotic lives away from a life of crime by identifying their risk of reoffending, coupled with other information that the custody officer needs to consider? That is the purpose and aim of it.

It builds on research that has already taken place in West Midlands police with an experiment, again linked to the University of Cambridge, called Turning Point. Turning Point and Checkpoint are very similar, and offer deferred prosecution to encourage offenders away from a life of crime. The difference between the two is that Turning Point is about first-time entrants in the criminal justice system, and Checkpoint in Durham constabulary is about looking at whether we can stop the revolving door of criminal justice for some offenders who go into the criminal justice system and out the other side. Is there something different we can do? We look at deferred prosecution and the best decision support we can give our custody officers to look at who is eligible for it.

Professor Bowers: I have been more involved in the kind of research that looks at the use of geographical methods and algorithms in policing. That particular type of work is often termed predictive policing.

Q151 **Chair:** In which areas to focus resources.

Professor Bowers: Yes. One of the reasons I can say something about this is that for many years I and colleagues at UCL have been developing thinking about methods of mapping crime and crime-related problems. One of the things we have been doing over time is to produce our own algorithms about where we think areas of high risk might be. They have been based very much on theoretical ideas about the way in which offending occurs. We were really interested in the spatial and temporal patterning of criminal behaviour. We started to see trends that perhaps we could use to say, "This is where crime happened in the past. Can we use that information about an offender, or broader information about what is happening in a particular location based on an event they have just undertaken, to start thinking about where crime is going to happen in the future?" It is a very theoretically driven approach.

We found that by adding that sort of contagion of risk element, in the form of an algorithm, to our mapping tools, we could modestly improve the predictive accuracy of where crime was going to happen in the future, so we have been interested in developing those systems in a policing context.

Q152 **Chair:** You have to make sure that you recognise the dynamic nature of policing and crime, and presumably there is a danger of reinforcing past patterns and not taking account of what is changing.

Professor Bowers: Do you mean that the data generate more activity in areas?

Q153 **Chair:** You end up intensively policing an area because of past patterns of behaviour, which might distort policing priorities in terms of what is happening in the future.

Professor Bowers: Of course, there is a risk of that. One of the interesting things about predictive policing is that you need to think about the data you are entering in the first place. If it was arrest data, it could well be the case that extra patrolling might lead to extra arrests. However, predictive policing algorithms are often based on recorded crime data, which are very different. When the public have interaction with a crime event, they nearly always go to the police directly and say, "We need resources in these areas," so quite a lot of the predictive algorithms we produce have been based on recorded crime data, which in a way is a demand from the public for police services.

Q154 **Chair:** Going back to my original question, how widespread is its use now in the UK?

Professor Bowers: We have undertaken our own pilot evaluations with a number of different police forces. We have tried it in the West Midlands police. We had a great relationship with Greater Manchester police where we implemented these types of algorithm, and with the Met. However, as you have probably read, there are other systems, such as PredPol, which are used widely, particularly by Kent police in this country.

Q155 **Chair:** What is the one being used in Kent?

Professor Bowers: PredPol is a commercially developed system based on theoretical ideas about the way crime spreads, and Kent in particular has taken it up.

Q156 **Chair:** Given that it is spreading and there are pilots around the place—a commercial product is being used in Kent—has there been enough discussion and public debate about the application of algorithms in the criminal justice system? Is it satisfactory that this picture is emerging, and developments happening, without much public debate around it?

Professor Bowers: It depends on the type of algorithm. For example, the geographical algorithms I am more familiar with are very different from those Sheena has been dealing with. We focus on areas, whereas the kinds of mechanisms Sheena uses look at individuals. Therefore, there is a different set of risks and issues from the point of view of the degree to which they expose individuals. Predictive policing algorithms have perhaps been more freely applied in this country in the past, given that they are a step on from other mapping techniques that have been used by crime managers for a longer time.

Sheena Urwin: Choosing to go into the public domain and explain what our harm assessment risk tool is and does was a deliberate decision by the organisation. We want to encourage public debate on it. We think we were the first in the country to have this kind of model, and the public need a voice in what we are using it for. Making a submission to this Committee, and sitting here and having a conversation about it now, all contribute to that public debate, and we hope that going forward we will be able to have some kind of national guidance or regulatory body around it.

Q157 **Chair:** You think there is a need for national guidance.

Sheena Urwin: Yes, I do.

Q158 **Chair:** Can you describe the Hart system a bit more, and how you use it?

Sheena Urwin: The harm assessment risk tool is supervised machine learning, and the method of choice is random forest technology. It is important to make that point because there are myriad different methodologies around machine learning, which is another reason why I think there needs to be an independent regulatory body. Very often, the questions you are asked demonstrate that people perhaps know more about a different kind of machine learning from the one we are using. It is random forest technology that was built using five years' worth of custody event data. It uses 34 predictor variables based largely on prior offending history.

The model itself has been independently validated. We have put quite a lot of information about that into the public domain already, certainly more information than has been in the public domain about some of the

American models. These kinds of models are more prolific there. The model supports decision-making for the custody officer. I do not know whether any of you have been in the custody environment. The custody officer's biggest priority is detainee welfare. However, they also make decisions on a number of different matters. Part of that is bail and part of it is custody, charging and out-of-court disposals. That is probably where the current narrative around custody and bail has arisen, but the decision support is really about out-of-court disposals. The custody officer looks at whether or not the offender is eligible for the Checkpoint intervention programme and takes a number of factors into account, one of which is the risk of reoffending.

Q159 **Chair:** It is to help to decide whether a non-criminal justice approach, or an approach other than charging the individual, may be appropriate. Presumably, because you are being cautious with it, it is fairly risk averse. Is that right?

Sheena Urwin: Yes. On the deferred prosecution side, we are in collaboration with the University of Cambridge, as I said, and we have put in place quite a rigorous scientific experiment; it is a randomised control trial. In the deferred prosecution model—the Turning Point experiment—the first-time entrants produced some quite positive results. As with any experiment, we cannot predict the outcome other than to describe it as green shoots. There are some positive indicators.

Q160 **Chair:** When people go through this alternative approach, you are not finding that errors have been made and they turn out to be very risky, and reoffend and so forth. Is that right?

Sheena Urwin: I cannot give you the definitive outcomes from the randomised control trial. It is ongoing, although we are coming to the end of it, and then we will need to await the outcomes. My understanding is that it is certainly showing positive results in reducing reoffending and de-escalating offending, and the Lammy review has recommended, on the back of Turning Point and Checkpoint, that deferred prosecutions be rolled out across the country.

Q161 **Chair:** If it is risk averse and you want to be very sure that people are suitable for this alternative programme, is there a risk that you are missing out on a wider group of people who may be suitable for the alternative programme and who then unnecessarily go through the criminal justice system?

Sheena Urwin: That is true while we are in the experimental phase, because we are trying to understand what works and what the evidence base is. Outside those criteria and outside the experiment when it is finished, all offenders will be able to be considered by the custody sergeant as eligible.

Q162 **Chair:** Within the pilot, does the custody sergeant ever override the guidance from the algorithm? In other words, if the algorithm concludes that an individual is not suitable for the alternative approach, does the

custody sergeant ever exercise his or her discretion to put them on to the alternative approach?

Sheena Urwin: Yes, they do. I cannot give you the figures, but they go against the algorithmic forecast because that is not the be-all and end-all; it is decision support. There are a number of other factors that a custody sergeant needs to consider, and is statutorily obliged to consider.

Q163 **Chair:** At the end of the trial, is it your intention to disclose all the detail?

Sheena Urwin: Yes. The outcomes of the Checkpoint trial will be published. We have already put into the public domain the validation of the forecasting model; the predictor variables are already out there, and the way importance scores contribute to the forecast is out there. Currently, an article is being written by Dr Geoffrey Barnes, who built the model in collaboration with us. It is about how he built the model, so importance plots around the different predictors will be out there. In the last month, we have separately published an article about an ethical framework that we think would support policing.

Q164 **Chair:** Do you think it has the potential to end up with more effective use of resources than just leaving it to humans?

Sheena Urwin: It is part of the decision-making process. In that environment, I do not think we could have an algorithm that makes the decision.

Q165 **Chair:** But presumably you are arguing that potentially it leads to better outcomes overall for society than not using it.

Sheena Urwin: Yes, it is decision support. The forecasting model is decision support that helps the custody sergeant make a decision about the out-of-court disposal. We are checking whether that out-of-court disposal, the Checkpoint programme, results in better outcomes. That is all we are looking at. We want more consistency in decision support. There is a different custody sergeant on every shift, for however many shifts a day and days of the week, so you get that human bias in any event. The hope is that this decision support tool will help with consistency. We have made a basic comparison, which again is in the public domain and is part of the validation, and which we will probably publish it in 2019 when the results are complete, of the custody officer's judgment of risk and the algorithm's judgment of risk.

Q166 **Chair:** Kate Bowers, in your submission you say that the PredPol system, which identifies crime hotspots, is only one part of the puzzle. Is there an opportunity for algorithms to advise on the action to take once a hotspot is found?

Professor Bowers: Yes. PredPol is based on scientific principles, and one of the issues we have with the use of some of these systems is whether or not, when you apply these things in the field, they lead to what we call meaningful implementation. In other words, at the end of the day, does the intervention speak to the problem that the geographical

algorithm has found? It is all very well to say that an area is at high risk according to the predictive algorithm, but deciding what to do in reality to reduce crime in that area is one of the things we are interested in trying to address.

There are all sorts of issues about what causes problems with implementation in areas of high predictive levels of risk. For example, problems can include the fact that predictive algorithms forecast very large areas that police could never get to within a particular shift; they might forecast areas that are quite far away from each other and the geography of the hotspots might be very fragmented, so the idea that a patrol can go through all of them is perhaps a bit unrealistic.

Another problem is that, quite often when these areas are found, operational police are just asked to patrol them without any understanding of why they are doing so. Systematic evidence on high-visibility policing is that it works in crime prevention. However, if you find predictively an area of high risk that has been generated through a process of understanding of the way offenders learn and behave there, you can use those ideas to try to come up with a good mechanism for reducing crime in those areas. For example, crime does spread. Research we have undertaken shows that for a short space of time—two weeks, for example—the risk of being burgled is considerably heightened if your neighbour within a radius of about 400 metres has been burgled. That shows that offenders go back to work the same areas, and offending spreads.

A great way of dealing with that in practice is to do something known as cocooning. The idea is that you protect the house that has been burgled from repeat victimisation, but you also move to the neighbouring houses that have a heightened risk of being victimised. That sort of approach speaks to what the predictive algorithm says in the first place is an area of high risk, whereas just putting in high-visibility policing might not be as effective in helping to police the patterns we have found.

Q167 **Chair:** In the US, algorithms are being used in some states to make sentencing and parole decisions. Do you see that coming to this country? What are the risks and opportunities? Should it be excluded from our planning in this country?

Sheena Urwin: My understanding from the reading I have undertaken from doing research around the harm assessment risk tool is that those used in the sentencing and parole process in the States are decision support. Certainly, in the Loomis case the finding of the Supreme Court was that it was decision support. I think the findings were about being able to explain to the individual that that kind of process has been used and is part of the decision-making process. If it were to be used in this country, it would need to be decision support.

Q168 **Chair:** We have covered this previously, but isn't there a danger that with decision support you end up with not much discretion being used?

Psychologically, the chances of the outcome of the algorithm being applied are pretty high.

Sheena Urwin: Does someone blindly follow the algorithm because "It is bound to be more accurate than I am"—that kind of scenario?

Chair: Yes.

Sheena Urwin: I hear what you say. Some of that is about making sure that the people making those decisions understand the tool itself. For example, in our organisation we have had awareness sessions with custody officers and taken them through the algorithm. One of the things you have to accept, and it was highlighted in the earlier session, is that algorithms are not perfect; they get things wrong the same as human beings. We are just more accepting when a human being makes a mistake than when an algorithm does. The question is, who makes a better decision? Some of that is about the regulatory environment for these tools.

Q169 **Chair:** That is what you are saying is necessary in the criminal justice system.

Sheena Urwin: Yes.

Q170 **Chair:** Kate, do you have any view about the application of the approach from the States in this country?

Professor Bowers: It is not necessarily my area of expertise, but I have been following the work Sheena has been doing in Durham. When it comes to decisions about individuals, the Hart system provides extra resources on the basis of a decision, so in many ways that is a very positive thing; it says that these are high-risk people and providing them with extra support is going to help to reduce harm to the community. When it comes to thinking about sentencing and parole decisions, that is a different thing. There are lots of other factors we need to consider.

Q171 **Chair:** What is your view?

Professor Bowers: I do not think that, without very heavy regulation, we should go forward with those sorts of approaches in this country, for a number of different reasons. Sheena's group has been incredibly good at evaluation work. They have the ALGO-CARE system, which shows that they are really thinking about how algorithms ought to be applied in the assessment of risks. It is a fantastic model. What worries me is the rolling out of these models to places where perhaps there is not the same level of expertise or care.

Q172 **Chair:** Again, it calls for a national framework and clear rules.

Professor Bowers: Yes.

Sheena Urwin: The other thing to bear in mind is that the custody environment is in any event very heavily regulated. With our model, we need some regulation and oversight for algorithmic tools, but the

environment itself in terms of human decision-making—this is decision support—is heavily regulated by legislation in the form of the Police and Criminal Evidence Act and the Policing and Crime Act 2017. The law we have to comply with is already there.

Q173 **Chair:** Professor Louise Amoore from Durham University told us: "you might wish to say that there is no place for inference or correlation in the criminal justice system." Do you agree with that or do you take a different view?

Sheena Urwin: From my point of view, and taking Kate's point in her submission about these things having a theoretical base, if you look at life-course criminology we already know certain things about patterns of offending. That helps us to validate models on an individual basis. We know, for example, that male offending and female offending are very different; we know that male offending escalates in a more violent way than female offending does, so we need to make sure that we use that evidence base and have that theoretical background to validate the tools and models.

Q174 **Bill Grant:** I will touch on several areas. One submission the Committee received advised that algorithms should not be allowed to consider data that we would not normally permit if it was wholly a human decision-making process—for example, if an algorithm was tasked with capturing and applying information such as a parent's criminal history. Would you agree that algorithms should be made to match the normal human process, or should there be something different? I like your phrase "support mechanism, not a final decision-maker". That is very important.

Sheena Urwin: It is almost a Pandora's box. How do you know what is in somebody's mind when they are making a decision? How do you know what information is available to that individual so that you can limit what goes into an algorithm? We can say exactly what goes into an algorithm. You cannot say what is already in a custody officer's mind. Parental criminal history is not in the harm assessment risk tool. However, it could be in the custody officer's mind because in their 25 years' service they may have arrested the father. They could have that in their mind, or from other police systems available. The algorithm does not have that.

Q175 **Bill Grant:** I am warming to that. You are saying that there could be a number of custody officers over a period of time and they could vary in gender and in bias; there could be four or five individuals, yet the common denominator would be the algorithm, which you suggest would engineer out the bias.

Sheena Urwin: Parental criminal history is not in there. That is not something Durham is considering going forward, but, hypothesising, we know there are households that impact across agencies, if we think of the national troubled families programme and high impact households. If we wanted to look at an intervention to assist and support in dealing with intergenerational offending, with an algorithmic tool to support that, it

might be quite right to include things like parental criminal history. With these models everything is context dependent; you have to understand their purpose and context.

Professor Bowers: I absolutely agree with the idea that there needs to be a human decision-making process. One of the problems with the GDPR idea of right of explanation is that what human decision-making means is still quite vague. Work has been done by my colleagues at UCL to show that we need to pay attention to what that means. With these sorts of things, there are risks that people could just pay lip service to the fact that there is a human decision at the end of the day. We need to make sure that somehow we safeguard against that in the way we move forward.

From the point of view of the geographical stuff, which is my area of expertise, when you point a police officer in the right direction for patrolling, sometimes the officer gets there and says, "I don't think this is the right place to go. I'm going to go right, not left as the algorithm tells me." In some cases, that will be the right decision. However, it is very different from the kind of decision-making that is done at individual level.

In my field, some of the research we have done shows that, although police officers are good at knowing where long-term risks are, or where the very time-stable hotspots are, they are not as good at working out where flare-ups in criminal activity are. We asked officers to say where they thought crime was going to happen and we compared it with where the short-term hotspots were. They were about 20% accurate.

Q176 **Bill Grant:** In the US, an algorithm used for sentencing called COMPAS picked out a black person on double the number of occasions it picked out a white person; it got it wrong, even though the algorithm was not programmed to identify that. Do you think there will be a continual risk of such failings, if we use algorithms to assess people, on race, religion, unemployment and so on? How do you engineer that out?

Sheena Urwin: There is without doubt a danger of algorithms perpetuating bias, and that is why there need to be robust mechanisms for audit and validation, and a regulatory body that understands this kind of technology that needs to be separate from the others. It could be a branch of the ICO; I am not precious about that. We already know there is a lack of skills and available resources generally in this area. We need to get the best people in the right places, who understand the different kinds of machine learning, to make sure that an audit mechanism is in place so that there is oversight right at the beginning, at the planning stage for these algorithms. We need knowledgeable review with a knowledgeable review board, so that there is a review mechanism when something has gone wrong and we can look at lessons learned.

Q177 **Bill Grant:** I am wondering whether in some cases we engineer out bias but we also engineer in a bias. I welcome your comment about the variation in custody officers who already have a bias, but when you

engineer that out with an algorithm, the algorithm may introduce a new dimension of bias—if that makes sense.

Sheena Urwin: Yes. To take race as an example, race is not in our model, but gender is. You could ask why we have gender in the model. Is that not discriminatory? Are we not making it biased by having it there? As I highlighted earlier, the offending patterns for males and females are very different, so arguably is it not fairer to have gender in? Otherwise, we will be making inaccurate predictions about male and female offending. We could potentially underestimate male offending and overestimate female offending. The question is about fairness. You could take out each individual variable that goes into the model; anything can be taken out, but you will take a hit on accuracy. There is a balance between ethical decisions around variables that go into the model and accuracy, because ultimately, if the model accurately predicts the risk of reoffending, there are victims at the end of it who need consideration too.

Q178 **Bill Grant:** Kate, do you have a contribution?

Professor Bowers: In my mind, it comes back to evaluation and monitoring of these things. One of the great things about algorithmic methods is that you can evaluate many of them in terms of how good they are at making decisions. As somebody who has been working in this country on the What Works crime reduction toolkit, for example, the idea of moving towards evidence-based approaches, where we try out these things and gain information about how good they are at doing things, would be a useful way forward. We need to set it very much within a What Works framework if we start doing these things more generally in this country.

Q179 **Bill Grant:** That has neatly answered my next question. How do we evaluate the algorithm? Do we compare it with the previous systems? You rightly say that we need something overarching to evaluate just how effective they are. You are suggesting trial periods that will keep those that work.

Professor Bowers: Absolutely. We are getting better at the evaluation of crime reduction initiatives in this country, and we have lots of skills in doing it. The College of Policing could be very much involved as a partner in some of those ventures. One way of proceeding with caution, as I think we should, is to make sure that we have the resources properly to evaluate the algorithms. What you compare them with is an interesting question. For example, I said that police officers had 20% accuracy and comparing it with the percentage for algorithms would be interesting. However, comparing algorithms with what humans do is different from comparing with what you might say would be 100% accuracy, or going for the golden bullet of being exactly right, which is impossible.

Q180 **Bill Grant:** How receptive are police forces to this new type of policing? You are obviously bringing in innovation. In a previous life in the fire service, I was tasked to bring change and move the service on. I found it

quite challenging to convert people to modern ways. Do you find police forces quite receptive to the adoption of new technology to assist in modern policing?

Professor Bowers: You start and I'll follow.

Bill Grant: You don't have to go into it too deeply.

Sheena Urwin: I would liken it to going back to school. You have a class ready for PE; some are better at PE than others. It is very similar.

Q181 **Chair:** It is a variable picture.

Sheena Urwin: Yes, absolutely.

Professor Bowers: It depends on the motivation to go with these things; it depends on understanding. If a PCSO is sent to an area because they are told they have to patrol there, it is not very inspiring for the officer; but if you can give an explanation such as, "We know that offending happens this way, and, therefore, if you go to these places this might be what you are helping to prevent," it is a much more motivating way of doing it, as is going to places where they feel they have achieved something, instead of saying that you will try to patrol 10% of the area in however many days, patrol 0.5%, or the maximum achievable, so that a real difference is made with the resource you put in. That should inspire them.

Sheena Urwin: It is cultural as well as organisational. Durham constabulary is steeped in problem-solving and evidence-based policing, so a lot of our officers are very open to change. It is a variable picture, but we are steeped in evidence-based policing and problem-solving, and that is what this research is all about.

Q182 **Clive Lewis:** Kate, you were talking about evaluation of the algorithms. Rather than policing, I want to look at sentencing, which is also something algorithms are used for. One of the algorithms used for sentencing in the US incorrectly shows judges that black defendants are high risk twice as often as white defendants. Evaluation is fine, but I might be supping on porridge for five years while the evaluation is going on and then it finds that the algorithm has put me in prison, or perhaps increased my sentence, wrongly. Talk me through that. For policing, it might work, but there are other situations where algorithms and the evaluation process—looking at how they work and the sentencing outcomes—might not be appropriate.

Professor Bowers: That is the key to the problem. Given that we have trusted people like Sheena doing these kinds of things in this country, we have a chance to think about developing evidence-based approaches before we roll out these things. You are pointing out the danger of doing something before we know, because of people's urgency to start putting these new technologies into place. We should use this opportunity to

make sure we have regulation. If we have to wait for it, we have to wait for it, but that would be a much more sensible approach.

Q183 **Clive Lewis:** To follow that up—you may want to come back to this in your answer—in the US in many of the private sector companies that are producing these algorithms there is no transparency in the nuts and bolts of them. You talked about scrutiny of the algorithms, but that is not there because of commercial sensitivity. Does that apply in the UK? Would that be a problem if there was a regulator? Could there be a blockage to a regulator if commercial companies said, "We're not sharing our data because it is sensitive and we can't do that"? Could that happen in the UK?

Sheena Urwin: I cannot speak for private companies, other than to say that there is an argument that, if we build these algorithms with open source software, part of a regulatory body or code of conduct could say that open source software is used, for example. If you look at some of them in the States, or in the Loomis case, we still do not know what the methodology is; we do not know whether or not it is machine learning, whereas for ours we have said it is open source software; we have given the website and the library where the random forest algorithm is. We have given out the variables. We have said that the methodology is random forest. We have been quite open, recognising the need for transparency because with transparency comes accountability. The danger with private companies is that we get into commercial sensitivity arguments, which can mean there is not as much transparency.

To pick up Kate's point on transparency, we have to ask what that is and what it means. In our scenario, is the individual more concerned about how they managed to be high risk or medium risk? What information did the model have available about that individual, rather than giving 4.2 million decision points in a random forest to the individual? That tells them nothing, so you have not been transparent at all; you have just given them an awful lot of confusing material. You reach a point where you say, "What is transparent here?" We have been very open and honest about some of the workings of the model, even putting out the independent validation, but there comes a point where, if transparency becomes too unwieldy and the individual is unable to understand how the algorithm arrived at the outcome, you have told them nothing. Is it more that they want to know what information the algorithm had to arrive at that outcome?

Q184 **Clive Lewis:** If there was a regulator of some kind, which you seem to be in favour of, especially in terms of the private sector and transparency, obviously it would be specialists, not the man or woman on the Clapham omnibus; they would be people with specialist knowledge who could look at it in detail and say that the data being used are perhaps not up to par, not up to scratch, and are delivering outcomes we do not want. That is the private sector.

As I understand it, Liberty is saying that general data protection laws

mean that the Government—the state—is exempt from any kind of scrutiny or transparency. Is that something that needs looking at as well, because potentially the state could be producing and using algorithms and we, the public, would have no idea what was going into them. That sounds quite frightening, doesn't it?

Sheena Urwin: Part 2 of the Data Protection Bill covers GDPR, and part 3 covers the European law enforcement directive. That gives us an exemption, because it recognises the sensitivities around policing, which is why we need a regulatory body. We welcome the introduction of mandatory data protection impact assessments and we would not be exempt from that. For example, we can already go to the Information Commissioner, so there is a regulator to which we can show information and explain why we will not put something in the public domain. That process already exists, but currently the impact assessments are discretionary. The Data Protection Bill makes that mandatory for law enforcement as well, so that gives a level of reassurance for our communities, and for us, that the methods we are using are ethical and accurate.

Q185 **Chair:** Do you think there is a degree of urgency about getting this regulatory body in place? I am conscious that the landscape is developing, and individual police forces could take it further in all sorts of different directions. Do you think the Government need to get on with this?

Sheena Urwin: Yes, I do.

Professor Bowers: I do.

Q186 **Carol Monaghan:** Kate, could I take you back to something you said a few minutes ago? You talked about human decision-making and whether it got the hotspots right. You mentioned that 20% of the time they got it right. How accurate are the algorithms in getting it right?

Professor Bowers: In that particular piece of work, we used crime mapping software to say where the hotspots were, so they found the reality of the hotspots. We have not yet done a piece of work that has compared that with the predictive systems, but we found it quite interesting that the police were good at finding long-term bits of risk but not very good at picking up short-term patterns of risk, the flare-ups, which algorithms are good at capturing. They can say, "This is a place we often go to because we know these are particular problem estates," which is great. They have a fantastic amount of experience in knowing where to go, but they might be more likely to miss the flare-up that, if they do not do something about it now, could lead to quite a splattering of offences in an area for a space of time before policing can get there, which is one of the reasons why we developed predictive geographical algorithms.

Q187 **Carol Monaghan:** But you do not have a figure for that.

Professor Bowers: We do not have a figure for that yet; we have not done that piece of research.

Q188 **Carol Monaghan:** That would be useful. Is there a role for Her Majesty's Inspector of Constabulary in the regulation of the use of algorithms?

Professor Bowers: My colleague at UCL, Michael Veale, has been thinking about some of the things you could do. One of the ideas is to have a watchdog organisation or agency that could include academics and other NGOs, but would also have people who were experts in computational law, ethics and those kinds of things. It is broader than just one agency that says, "Here's your problem; you've now got to sort out the algorithms and ethics." We need to draw expertise from lots of different places to get the right board of people to act on that.

Q189 **Carol Monaghan:** You have just said that you are evaluating the use of algorithms to see how well they perform. Should algorithms that are used in policing be subject to an audit? I think what you are talking about is good practice rather than a requirement. Is there a requirement?

Professor Bowers: The impact assessment Sheena talked about is certainly one way forward. Do not let it happen until equity and data protection have been looked at in terms of impact. Our colleagues in the earlier session talked about certification as one way of going forward with these kinds of things. There need to be systems in place that say that one approach seems to be reasonable, whereas we are questioning another approach more.

Sheena Urwin: HMICFRS, as it is now called, can come into the organisation at any point and look. They can audit, but, to go back to my earlier point, we need to make sure that they have the skills and abilities to do so, because these are complex tools. We need to make sure that the people who audit them and look at them to see whether we are doing the right thing have the necessary skills and knowledge to do so.

Professor Bowers: I think police forces have ethics boards, which is another point.

Sheena Urwin: Yes, external ethics boards.

Professor Bowers: Just as universities have ethics boards when they are thinking about new research projects, a policing ethics board might be a good place to put this through before development occurs.

Q190 **Carol Monaghan:** That is interesting. Sheena, you mentioned that decision-making always had to have a human in the loop; you said there was a statutory requirement. I believe article 22 of the GDPR talks about the rights of individuals to be exempt from fully automated decisions that have legal effects. Can you guarantee that Durham is not taking the human out of the loop entirely?

Sheena Urwin: Understanding policing and knowing the custody environment, and that all decisions are challengeable, we have not changed that. When making these decisions, the custody sergeant has a number of factors to consider. Risk of reoffending by the offender is but one. Some of the things they have to consider are victim vulnerability, age, means and accommodation; proximity of the victim to the suspect; and victim vulnerability due to the impact of the offence, physical and mental.

Q191 **Carol Monaghan:** Would the algorithm not deal with all of that?

Sheena Urwin: No. The algorithm has 34 predictor variables within it, and the majority of them are based on prior offending history, so it assesses only risk of reoffending. The other factors that custody sergeants must take into account are those I mentioned around the victim. There are factors around the suspect, not just the risk of reoffending, because it is built only on Durham constabulary data. It can only be built on that, because those are the only data we have available to us. It is not built on police national computer data or the police national database. If somebody came from a different force area, we would still need to check those national systems to see what is in their history, because we need to consider the suspect when making those decisions.

You have to consider separately the offence itself for which the individual is standing in front of the custody sergeant; the nature of the offence and the impact on the public and the prevalence of that kind of offence in the area. You also need to look at the need to protect the public. What is the location of the offence? Are we talking about the offence having an impact in a capital city, or are we talking about a rural village? Those are the factors that need to be taken into account by the custody officer in assessing necessity and proportionality, which are covered under the Police and Criminal Evidence Act 1984 and the Bail Act 1976. All those factors have to be taken into account, and only one of them is risk of reoffending. Custody officers can be challenged through the court process on that kind of decision-making and its rationale. That has not changed.

Q192 **Carol Monaghan:** I am going to ask a question to which possibly you do not know the answer. Do you have any figures for the percentage of officers who overturn the decision of the algorithm?

Sheena Urwin: I do not have the figures; I can write to the Committee.

Q193 **Carol Monaghan:** It would be useful to have them.

Sheena Urwin: I can write to the Committee and give you what we have, but it is very new so there is not much data to show you.

Q194 **Carol Monaghan:** It would just give people confidence that they knew that it was possible to challenge the algorithm, and people were following the guidelines you are talking about.

Sheena Urwin: Yes. Bear in mind that it is not a decision; it is decision support, so the comparison is with the risk of reoffending and the judgment of the officer of that risk, not all those other factors. That is the difference. We can write and let you have it; it is a small amount of data.

Chair: I understand.

Professor Bowers: It is also a lot to do with data quality. We have to think about the limitations of the data that go into these algorithms in the first place. Often, people who doubt the quality of a particular decision might pick up on the fact that there has been some sort of data error. That is one of the things we say is useful about theoretical approaches, because if something comes out that does not look right, it is useful for someone to say, "Why do we have a cluster of crime where we weren't expecting to find it?" It could be that an earlier data error meant that lots of things were geographically located in the wrong place.

Chair: I understand the point.

Q195 **Bill Grant:** Kate, in your submission you tell us that the algorithms need to be informed by theory and failing to understand the black-box aspect could lead to inaccurate results. How do you combine the two to get accurate results?

Professor Bowers: It comes back to the fact that there are lots of different algorithms in the first place. The types of algorithms we have developed are not the kind of machine-learning aspects you might find in other places. We came up with an idea and used that to produce a set of rules to look at risk. A theoretically driven approach went through the kind of mapping we were doing, but the advantage of doing that is that later on you can say, "This is the reason why we have these areas coming up, or why we've got this particular result." It should be based on how we understand offending happens and all the research we have. Academic research has also informed Sheena's models. It means that, at the end of the day, when we come up with some sort of decision, we have a reason not just for believing that a particular risk is there, but why it is there.

Q196 **Bill Grant:** Following on from that, predictive policing algorithms, PredPol in its shortened form, appear to rely entirely on past events. Things are dynamic; they change. There are social media things—the snap thing or the crowd thing. Is there a weakness that it will fail to grasp the future or have an element of prediction of the future?

Professor Bowers: PredPol is based on a system similar to the one we produced, which I talked about, in that it imagines a contagion, which means that it is using a past event to predict where a future event might be on the basis of how it is spreading or happening. Although it uses past data, it does that to predict where areas that have not been subject to offending will be subject to it. One of the criticisms of these kinds of approaches is that a lot of the geographical data they use is past data. Other kinds of algorithms that have been developed use environmental risk factors. One of those is risk terrain modelling, which uses

environmental factors that we know are likely to cause crime or trouble. Those are places where you might start predicting risk.

Probably the most interesting developments we have seen in this field are those that take a combined approach, using environmental and demographic factors along with where previous crime has happened in areas. You can combine those two approaches. An example of a piece of software that is moving towards combining the two approaches is HunchLab.

Q197 **Bill Grant:** One of the things in the future is the use of social media. There is a name for it, but I am a generation away from it. You can pull a crowd. If your parents are away for the day you can have a party planned for 20 but 300 turn up. If there is a sunny day in a coastal area, loads of people turn up via social media. Can that be put into PredPol? Can that be predicted?

Professor Bowers: We have research going on at our university and other universities on the uses of Twitter and social media data more generally in looking at crime risks. Twitter data are a very useful population denominator. Population denominators from the census are often very static, so Twitter enables us to say that this kind of activity was going on in that area at that time, so we are looking into that.

Sheena Urwin: On the point about past data from the point of view of an individual risk assessment tool, there are different kinds of machine learning, which is why it is important that the right people review these kinds of models in terms of regulatory bodies. Our random forest model is supervised machine learning, which means it does not learn and alter its structure as it goes along. We have to rebuild the model with more up-to-date data, so we can check to make sure it is working correctly. When you say machine learning, some people assume neural networks, on which I do not pretend to be an expert, but I understand that they learn as they go along. Our model does not. That is why it is important to have the right people who know these things.

Q198 **Bill Grant:** Has any work been done to reflect on traditional policing compared with where PredPol has been applied? Is there any evidence suggesting that PredPol may not solve the crime but may relocate it? If you put in extra resources, those who peddle goods on the street may relocate their criminal activities.

Sheena Urwin: Hotspot policing is probably the most well-researched area of policing. I point to Kate for the academic view.

Q199 **Chair:** Can you give a brief answer?

Professor Bowers: We have done a systematic review of whether or not geographically focused policing initiatives displace crime to other areas. The systematic evidence from a large number of different evaluations is that it does not. On balance, you get some displacement, but you get as much of what we call diffusion of benefit, which is where those areas and

nearby areas of geographical risk are avoided, so on balance these things do not displace crime to other areas.

Q200 **Bill Grant:** A proportion of it is displaced but not in its entirety.

Professor Bowers: Yes. It more or less equals out in terms of the amount of displacement and diffusion of benefit that you get. It is generally quite an optimistic picture, so overall you would not just push crime to other places.

Q201 **Chair:** Finally, there are two brief questions from me. Perhaps you would give brief answers, if possible. One assumes that the more data we have, the more it can aid the application of the algorithm. Do police forces share data effectively in this regard yet, or is it still very much down to isolated pilots around the country? We have talked about West Midlands, Manchester and so forth.

Professor Bowers: One of the things we have done at UCL, which is a reasonably recent innovation, is to establish our secure data lab. It is an academic secure data lab that has been assured to PASF standards, so it is like a police station. We did that on purpose because we wanted much more enriched encounters between policing and law enforcement agencies and academics, but we need to do that in a place where we have regulations for data protection. What we are hoping to do from that lab and initiative is allow a mixing of different datasets in a very secure place. That will enable us to combine social and economic data with crime data, and perhaps data from the PNC.

Q202 **Chair:** It is a resource that could be used across the piece.

Professor Bowers: Yes.

Q203 **Chair:** Kate, a report by HM inspectorate of constabulary earlier in the year highlighted that "forces have not yet made effective use of predictive crime mapping tools." In assessing predictive policing, Kent police found that only 25% of target areas generated by the software were visited by police officers. Why has take-up been so slow? To what extent is there a problem with officers believing the results of the algorithms? Is there any view on that?

Professor Bowers: I believe the PredPol algorithm was being used in Kent. In combining different sensitivities of data, we can start looking at things like where policing resources go. For example, GPS tracking can help us look at where officers choose to go. The picture is right. It is correct. A lot of the areas designated as places that ought to be patrolled are not patrolled for a number of reasons. Sometimes it is because of the algorithmic sensitivities. With these things, you can say, "Tell me where the top 10% of areas are," which is very different from saying, "Tell me where the top 1% is." It needs to be balanced against the available resource.

Q204 **Chair:** There are just not enough police officers to go round.

Professor Bowers: Exactly. I have seen too many cases where the mapping system has said to operational police, "Go everywhere." We need to be going much more towards what we know to be the very high-risk areas. That is one of the reasons why that does not happen.

Q205 **Chair:** When both of you came along today was there one clear message above everything else that you wanted to get across to us? I have picked up very much that you see the importance of Government getting on with establishing a regulatory framework for the use of algorithms in policing and criminal justice in general. Is the take-away from your evidence today that there is an urgency about this? Is there any other supplementary point you want us to take on board?

Sheena Urwin: First, the point about the regulator is key. Secondly, from Durham constabulary's point of view, the narrative out there at the minute about this tool is that artificial intelligence in Durham decides if you are in custody or on bail. That is not what the algorithm does. It is about encouraging offenders away from a life of crime; it is about improving life chances, and it is part of a research programme. The reality is that policing decisions often are not black and white. A decision made today in certain circumstances will be different from a decision made tomorrow.

The cornerstone of policing around the protection of life and property is based on one fundamental thing: a police officer's discretion. There are various things that contribute to that: the law; the National Police Chiefs Council guidance; College of Policing guidance; and regulatory bodies such as the ICO and HMIC. We owe it to our officers to make sure that we have the best decision support tools available. This research programme is about giving them the right tools to do the job, so that our communities can be reassured.

Q206 **Chair:** Thank you. Kate, is there any take-away message you want to leave?

Professor Bowers: I agree. I really believe that we need a good system for evaluation of these things in this country.

Q207 **Chair:** Evaluating before authorising.

Professor Bowers: It should be throughout. It needs to be on the cycle. First, you need to evaluate whether or not from an impact assessment point of view it is a good idea; you need to evaluate whether or not it has been implemented properly, which is process evaluation; and then you need to evaluate the outcomes and whether or not they are effective in doing the job they have been set up to do.

Chair: You have been incredibly helpful. Thank you both very much indeed. We appreciate your time.