

Science and Technology Committee

Oral evidence: [Robotics and artificial intelligence](#), HC 145

Tuesday 24 May 2016

Ordered by the House of Commons to be published on 24 May 2016.

Written evidence from witnesses:

- [RACE, UK Atomic Energy Authority](#)
- [AAAI and UKCRC](#)
- [The Royal Society](#)
- [Research Councils UK](#)
- [Article 36](#)
- [Future of Humanity Institute, Centre for the Study of Existential Risk, Global Priorities Project, and Future of Life Institute](#)

[Watch the meeting](#)

Members present: Nicola Blackwood (Chair); Victoria Borwick; Jim Dowd; Chris Green; Dr Tania Mathias; Carol Monaghan; Graham Stringer; Matt Warman

Questions 1-82

Witnesses: **Dr Rob Buckingham**, Director, RACE, **Professor Stephen Muggleton**, Association for the Advancement of Artificial Intelligence, **Professor Nick Jennings**, Royal Society Working Group on Machine Learning, and **Professor Philip Nelson**, Chair, Research Councils UK, gave evidence.

Q1 Chair: I welcome the panel to the first session of our new inquiry into AI and robotics. Thank you all for coming. We are very excited about this inquiry into one of the Government's eight great technologies. The subject summons up quite a lot of—I don't quite know how to say it—divergent views on exactly what the future is. We have the likes of Elon Musk saying that the development of AI is the same as “summoning the demon” and is probably the biggest threat facing the world, while others say that the idea of general artificial intelligence is still decades away. Professor Jennings, can I start by asking where exactly you think we are with the development of AI, and what do you think is a realistic expectation of its impact on our society?

Professor Jennings: You can answer that question at several different levels. At the simple level, artificial intelligence software is routinely used in our everyday environment. If you use Facebook, the algorithms that recommend friends to you are AI algorithms; if you use a smartphone, the voice recognition software uses AI algorithms and techniques. Specific AI, in very narrow niche areas, is already here and is routinely available. More general artificial intelligence, much like you see in the movies and as the subject of some of the quotes you mention, is an extremely long way away. To achieve that is an exceedingly difficult scientific and engineering challenge, so it is a really long time into the future. As we move towards it, machines are getting smarter and more effective and efficient. That is how I see the trajectory happening. There will be lots of useful applications that use AI technology from now into the future as we move towards ever smarter, ever more general technologies.

Q2 Chair: Professor Muggleton, do you think these technologies will replace people, or are they going to remain essentially tools for clever people to manipulate and increase productivity?

Professor Muggleton: A lot of the interest in artificial intelligence at the moment is related to spectacular breakthroughs that have happened recently in areas that have largely involved neural networks. The question whether or not human beings will be collaborating with such machines is related partly to advances we need to make beyond the limitations of some of the techniques in the public eye at present. In particular, greater efforts need to be put into systems that communicate effectively with human beings, so that you do not have a situation where the world's top Go player is beaten at his own game, and the only person who can explain what happened is the human being. There is no built-in ability to communicate or collaborate from the neural net side.

Q3 Chair: This is a point about general understanding of the development of AI and making sure that there is confidence in exactly what is going on with AI and overcoming potential fears and misunderstandings. Dr Buckingham, given that we do not know where this revolution is going, what do you think the Government can best do at this stage to support the sector?

Dr Buckingham: One word: test. We need to do things for real and find out what does and does not work and what we can learn from it. You have to do that in a safe place and make sure you have the ground rules completely established so that you make progress in an informed, managed and safe way. That was very much at the heart of the UK RAS strategy. We talked about the need to evaluate in detail. Take the issue seriously, but at the same time, do not miss the opportunities for new jobs and growth. I would answer the jobs question by saying this is great news because it will enable us to be more productive and do things that we are currently unable to do. I see it as a good thing for humanity and not something we should be worried about, but we should be taking it very seriously.

Q4 Chair: Professor Nelson, what are you doing in the research councils to make sure not only that there is wider understanding of the accurate situation with AI and growth but that the ethical issues surrounding the development of AI have been properly taken into account in research development?

Professor Nelson: There is a whole raft of things under way sponsored through the research councils, especially my own, EPSRC. We have between £30 million and £40 million current investment in robotics and autonomous systems in a range of institutions across the country. We are also investing in networks to make sure we gather together all the expertise. One of the features of the landscape is that we have great expertise in various specialist areas that are networked across the country. We have outstanding work in Southampton, Bristol, Oxford, London, Cranfield, Loughborough, Liverpool, Leeds, Sheffield and Heriot-Watt and Edinburgh. All those universities have excellent expertise in various aspects of robotics and autonomous systems, and we are trying to network all those people together on the back of various investments we have made. We have invested in about five centres for doctoral training in various of those universities. There has been a good capital injection from Government; back in 2013 about £25 million was invested to help those various centres to develop their work.

We also take the whole business of ethical concerns around these technologies very seriously. We convene cross-council workshops where we talk about principles of robotics, for example. Those are published on our website. You can see the early findings from those workshops that we are currently updating. All our work is carried out, we hope, in a framework of responsible innovation. We try to put down guidelines for researchers to think about the societal and economic impacts of the work they are undertaking, especially when it comes to ethical issues and so on. We appreciate that it is a complex area with many facets. I think we have a reasonably good spread of work going on across the UK. Of course we could do more—there is no question—and doubtless you will not be surprised to hear me say that.

Q5 Chair: Given the complexity and the potential for misunderstanding about the direction of the research, are there any areas where you have concluded that researchers should not be developing AI because of the potential for it to bring more disadvantages than benefits at this stage?

Professor Nelson: It goes back to the ethical frameworks. As the field develops, there will probably be a clear need for standards and regulations. Any new technology probably needs that sort of framework within which to work. I think it is a job for Government to help with that process and ensure that the public develop trust in these systems. For example, people are generally very comfortable about flying in aircraft, because they know there is a civil aviation authority that regulates the whole business of flying. It is a very low-risk activity because, as soon as there is an accident, there is a proper investigation into it, as we have seen just recently. All of that is made clear and it is a very robust process. As technology in this area develops, a need will probably arise for that sort of intervention, to check autonomous vehicles, for example, currently a big topic, to ensure that they are properly regulated and to build trust in the community. I am sure all of those are very important issues.

Q6 Chair: Dr Buckingham, do you have specific recommendations for the Government not just for building public trust but educating the public on sensible practice in engagement with AI, and on what is really going on?

Dr Buckingham: Absolutely. We recently won some funding that has the catchy name PAVE—people in autonomous vehicles in urban environments. The key piece is people. This is about people. It is not about robots; it is about human beings, and we need to make sure that the public, regulators, insurers and lawyers—all those sorts of people—are engaging in the technology. A lot of stuff talked about in the press is great, but the conversation needs to be much deeper than that. One of the things we want to do is engage lots of people around, for instance, the driverless car tests that are happening around the country. They are great. The UK has taken the lead in this area. These cars are the first robots. This is the first time that robots—cars with the sensor tech that makes them autonomous—are highly visible in public spaces. It will be interesting to see how people respond to them. Some people will love it and some will be concerned, but the fact that we are doing it in the UK is to be commended.

Q7 Chair: Professor Jennings, do you have specific recommendations for how we can address the specific behavioural, ethical dilemma?

Professor Jennings: There are a number of points, one of which is around data. Some of the key bits of artificial intelligence and machine learning rely on data. Data is the fuel for all the algorithms to do their stuff and make smart decisions and learn. There are a whole load of issues associated with appropriate management of data to make sure that it is ethically sourced and used under appropriate consent regimes. That is a really important piece. As part of the Royal Society working group, we engage with Ipsos MORI, which is holding a number of public consultations to learn what the public currently understand about this technology; what they think the key issues are and how far they are willing to delegate autonomous artificial decision making to systems. Understanding that is really important.

If you look at other controversial areas of government, for example, stem cell research, the way the Warnock Commission worked and how the corresponding human fertilisation authority came after that is a good model for something that works well. If you contrast that perhaps with the genetic modification of plants and the way that was handled, it did not really engage the public early and quickly enough. While I said at the beginning that I think this is a general technology for the distant future, engagement needs to start now so that people are aware of the facts when they are drawing up their opinions and they are given sensible views about what the future might be.

Q8 Victoria Borwick: The million-dollar question is what do you think will be the greatest potential benefits to society from artificial intelligence and robotics over the next 10 years or so?

Professor Nelson: They could be hugely wide-ranging. This technology is all pervasive and has potential applications—I am talking about AI and robotics as a catch-all—that will clearly impact industry. Aerospace, automotive and those sorts of industries are already hugely involved and benefit from the application of these technologies. There is all sorts of scope for doing great things in automated agriculture. In the service sector, many routine jobs may well be more readily done using some of these computational techniques. There is a whole range of things. In healthcare we have great expertise in this country in surgical robotics, for example. There is a lot of work in assistive robotics for the elderly

and so forth. It is a technology that is hugely broad and can be applied in many different spheres of human activity. That is why it is a very important technology, and certainly the research councils regard it as a key priority for them.

Q9 Victoria Borwick: This debate is rather timely in view of the fact that we have Invictus games winners in the House today. Does anybody else want to make any comments on where they see the advances and benefits for society?

Professor Jennings: I would like to pick up and amplify what Bill said. The joy of artificial intelligence is that it is a broad-based technology; in almost all areas where humans are involved in decision making and action you can see benefits to different degrees from this sort of technology. In particular, some of the medical stuff is a fantastic opportunity. The rate at which new medical knowledge is generated is, by some measures, doubling every 18 months to two years. Medical practitioners have little chance to stay on top of that. By the time you have graduated from medical school, the amount of knowledge has doubled and then doubled again. You want the machinery—computers—to help you understand that and bring it to bear when making decisions. It is not to replace humans, to come back to the original point, but to work in partnership with them and make suggestions such as, “Have you thought about this? It could be this or that.” That is the key future for AI.

Professor Muggleton: I very much agree with that. Great advances have been made in the UK using AI technologies in medical sciences. You have large amounts of data from genome projects and testing. Machines are able to go through millions of hypotheses and select the best out of a large space and then present it to scientists. That does not replace scientists; it amplifies what they can do, much in the same way as a telescope amplifies what astronomers could do. There are huge possible benefits.

Q10 Victoria Borwick: What do you think are the technical challenges that innovators will have to overcome?

Professor Muggleton: One of them relates to the technical problems I mentioned earlier. When you have complex hypotheses of the kind I have just described being brought forward by computers, it is vital that scientists understand what is being claimed to be the case so that they have the opportunity to go away to the literature and discuss it with their colleagues, and understand the implications of where the suggested techniques can be used.

Dr Buckingham: It is not just the technical stuff. There are lots of non-technical issues around these technologies, about how we use them and how we engage with a much wider part of society—regulation, insurance, the legal part, finance and investment. We must not focus so much or exclusively on the technology piece; we have to link it up with all the rest, because that is the way we will create jobs and growth. We need to move those great ideas out of universities and into companies. That is what we should be focusing on in the UK. It is no good just doing the research; we should be up there leading with the creation of companies and jobs.

Q11 Victoria Borwick: How do you think all those opportunities in robotics and artificial intelligence are being capitalised on by UK industry and Government?

Professor Nelson: In the research councils we always look for industry engagement in our investments, and we have some very good results in that regard. We launched an autonomous systems partnership, as I think it was called, back in 2012 where multiple companies were involved with about 16 universities. There were probably double the number of companies than universities involved. More recent investments have been very effective. For example, with Jaguar Land Rover we made an £11 million investment recently in five projects in the autonomous vehicle space. That was very much about human interaction with the vehicle. We work alongside industry constantly.

Our colleagues in Innovate UK are also very interested in this area. I have been having very constructive conversations with them about launching a national initiative in this area. In our submission we referred to a national flagship institute. I believe it is time we did something like that, connecting distributed excellence in science to probably a distributed network of innovation activity in some sense. Because of the complexity of the landscape it will take some careful thought, but, as we have found with previous initiatives, we open it to competition. One of the criteria is how many partners from industry you can bring with you. We find that works and we get an awful lot of industry alongside very quickly. Our universities are getting better and better at connecting with that landscape. There is a great opportunity, and I think Innovate UK agrees with me.

Professor Jennings: I might take a Government angle on that. Until recently I was a Government chief scientific adviser. Many Government Departments are clearly very interested in these sorts of technologies because they provide a way to process vast amounts of data. Government do not have lots of analysts and people to throw at a problem in terms of analysing ever more data. We get ever more data about everything from a variety of different sources, and we are not magically going to grow the capacity of the analyst cadre in Government. Given that, you need the machinery—the computers—to do very much more of the processing and analysis for you. Many Government Departments are starting to get a handle on that and understand that it is a key technology for them.

Dr Buckingham: That is the data side, and then there is the robotics side with the machines in our physical world. When I am older I want to remain mobile. That is the best way to keep down my healthcare costs. I want to be able to go places, see stuff and remain mentally and physically active. It may well be that autonomous vehicles, or driverless cars, are part of that. They are not by any means the whole solution, but they could well be part of it. There are much wider social implications. When we talk to permanent secretaries across Government all of them say that this affects military stuff, health, housing, nuclear decommissioning and DECC. All of these areas are impacted by the data and the robots in the physical world.

Q12 Victoria Borwick: Do we risk losing market share, or do you think we are operating in a particularly—

Professor Nelson: I think we do. You only have to look at the investments being made in other economies, particularly in Japan and South Korea. South Korea has been investing

\$100 million per year for the past 10 years or so. Those are very significant investments. Japan has just put \$350 million into a big programme of assistive robotics. I worry about keeping pace with these developments. In the USA, a lot of private sector investment is going in, especially in the area of autonomous vehicles. That is well broadcast. As the technology develops it is important that we keep pace, and we have a great starting point.

Q13 Jim Dowd: Given the apparent hostility to the idea of autonomous weapons systems, is there not a huge fear among the public, certainly among what we prefer to believe are civilised and democratic societies, that we are creating an entity in autonomous systems generally that we may not be able to control?

Professor Nelson: The fears are understandable. This goes back to the earlier comments I made about regulation and having societal control over these activities. I think the trials of driverless cars in California demonstrate lower accident rates than in the normal course of driven cars, so it goes back to being able to prove that this works, showing people that you can verify and validate—I think those are the technical words—these technologies. Provided that is properly managed, those fears should be put aside. There are obvious military applications for robotics and, depending on your point of view, very successful ones. One of our companies in the USA, QinetiQ, has produced very effective devices—you have probably seen them—to defuse devices that are set to explode and disable people. There are all sorts of very useful technologies associated with that work.

Q14 Jim Dowd: But it has to be more than the proposition that a not very good driver is better than a bad one; it has to be a much firmer base than that.

Dr Buckingham: The real test will be when your insurance costs go down when driving an autonomous vehicle. There will come a point when you will not be able to get insurance to drive your own car because it will be that much safer. We tolerate far too much death on our roads, so anything that moves us closer to a position where we can just get there safely and reliably and faster, with less congestion and fewer CO₂ emissions along the way, surely must be good.

Q15 Jim Dowd: Given that no human systems are infallible, how can we ensure, particularly if we are working in what is known as the learn and adapt environment for AI systems, that what they are learning and adapting to is what was intended?

Professor Muggleton: All technologies that are developed have inherent dangers if they are to be useful. We need to ensure that we can develop a methodology by which testing can be done and the systems can be retrained, if they are machine learning systems, by identifying precisely where the element of failure was. Without that probe—that ability to identify such a failure point—it does not have the strength of other engineering technologies, but it is there that it potentially has strength as well, because these systems can learn from errors.

Q16 Jim Dowd: Given the fact that there are problems with the verification and validation of static software systems—very few come out glitch-free, and everybody says, “Never buy a

dot zero release”—how can we adapt to ensure that we have adequate safeguards with active systems?

Professor Jennings: Your statement about software is generally true at some ends of the spectrum; at other ends of the spectrum, where integrity is really important and verification and validation of software is done using advanced state-of-the-art techniques, you can prove certain things about the properties of your software—that it will do this and it will not do that. Checks for safety and liveness properties for static systems are now largely doable if you are willing to invest in the tools and techniques for them. You are right to point out that when you move to more dynamic and adaptive systems it becomes more challenging. That is an area of science that we do not really know how to do in its general form, which is why basic research is needed in those areas. It is one of the familiar key issues; you want to be able to have the best guarantees about your software.

Professor Nelson: I concur with that. It is a fundamental problem in science. Some people call it the black box problem. It absolutely needs some work to help us through that, because the potential for these learning systems is so vast that bringing them to bear in a trustworthy way is surely of huge benefit if we can solve that problem.

Q17 Jim Dowd: Who should be responsible? Should there be standards for verification and validation procedures, and who should devise them?

Dr Buckingham: The answer is yes, there should be standards, and they will be developed through existing routes.

Q18 Jim Dowd: By whom? Would it be public bodies or practitioners?

Dr Buckingham: Generally, public bodies. These things develop in phases, don't they? They sometimes get developed by private organisations and are transmuted into public standards, because that is the way the market socialises all these good ideas and tests them out. In this area, we would also advocate some oversight. We have talked about there being a way of making sure we stay in touch with the development of these techniques more generally. We started by saying that the public need to be brought along, and made aware of what is going on as it is evolved.

Professor Nelson: I believe a British standard on robotics has recently been issued. I confess I am not intimately familiar with it, but I believe it is there now. The British Standards Institute has a long track record of engaging practitioners and developing appropriate standards through a well-tried and tested process, so this is the first step along that line, at least in this country.

Q19 Chair: Professor Nelson, when you started your answer to me you said that public trust in AI and robotics needed to be similar to that in aviation, where there is a civil aviation authority and an accident investigation branch, so there is confidence not only that there is a secure regulatory environment but that when something goes wrong it is robustly investigated. Is the proposition that there should be a similar kind of structure put in place for driverless cars and any other kind of specific applications that come along?

Professor Nelson: I would have thought that would be sensible. It really depends on the application and how quickly we might see driverless cars, for example, come into service. There will probably be a period of gradual uptake of these technologies as cars become more sophisticated. I am sure all our cars are now more sophisticated as a result of some of these technologies being deployed; the modern motor vehicle is far more sophisticated than it ever was before. It depends on when the time comes to regulate fully autonomous cars that we can basically walk into and sit down and off they go. I am thinking some years down the line, but the direction of travel can clearly be seen. We probably need to start thinking now about having something in place to keep track of those developments, which has oversight of them and enables thought to be put into how the regulatory environment might be put together. We need to start thinking about it now, even though it will not be next month when those cars are travelling on the roads.

Dr Buckingham: The current practice that has been brought out by DfT on the pathway to driverless cars—also known as the driveway to pathless cars—is really good because it enables testing to happen in the UK in controlled ways. One thing we must not do is put too much red tape around this at the wrong time and stop things developing. One of the key points is to make sure that we are doing that testing in the UK transparently and bringing the industry here so that we understand what is going on, and that we start to apply the regulation appropriately when we have more information about what the issues are. One of the risks is that, if we over-regulate, it is bad for making use of the technology. We are at an early stage. This technology will have an impact over decades, and it will only get more and more important as we go forward. You cannot imagine that computers are not going to get faster and software will not get better. This will not stand still; it will develop and develop. We are at the early stages, and that is why we should be involved.

Professor Nelson: I would like to reinforce that in relation to my comments about regulation. It needs to be a phased process so that the practitioners and experimenters are engaged the whole way along the line and we do not stifle innovation as a consequence.

Professor Jennings: It is also important not to view AI technologies as one thing. You very rightly drew out autonomous cars. That is a sub-area that might require very different forms of regulation from a hand-held device where we are using AI for speech understanding. There is not the same level of danger or risk associated with both of those, even though there may be common elements of the technology. Regulation is required in the application and use of the technology, not in the foundational principles of them. The foundational principles will be common to the very serious applications we have spoken about and to some of the more routine ones that are unlikely to cause harm to anyone.

Q20 Graham Stringer: In a way, I apologise for this question, but I cannot resist after Jim’s questions. It is the “Terminator” question, not about time travel but the other part of the film where the weapons system Skynet becomes self-aware and decides it does not want human beings to look after it any more. Do you believe there is any danger, even in the middle distance, of AI and robotics getting out of control as they did in that film, and, if not, why not?

Dr Buckingham: Shall I be brave? What do you mean by middle distance? The problem with these existential questions is that it would be unwise to think that over a very long time we will not be in a place where life has changed and moved on and we have invented

some amazing stuff and moved forwards. If we get drawn too much down the sci-fi route, we miss the really important thing: what is the impact now on jobs and growth, and what should we be doing to invest in those issues now so that we have a leadership role and can shape them happening? I am all in favour of great films and will watch lots of them. They ask all sorts of awesome questions around human consciousness, what it is to be human and all that great stuff, but we are an awfully long way away from “Terminator”.

Professor Muggleton: Absolutely.

Dr Buckingham: We all agree.

Professor Jennings: My assertion would be that no one in this room would be troubled by that during their lifetime. To take the leap from current state of the art—things we currently know how to do—to that is such a difficult thing to do. It is not going to happen any time soon, and we just do not know how to do it.

Q21 Graham Stringer: But it might some time in the more distant future.

Professor Nelson: The future is notoriously difficult to predict.

Q22 Graham Stringer: But you’re not ruling it out.

Professor Nelson: It is a pressing issue to make sure that we harness the technology, use it well and regulate it properly. If we do that, those scenarios are headed off right away. That is the key.

Q23 Graham Stringer: Professor Nelson, I think you said we needed to invest for reasons of productivity and GDP, essentially. There is little evidence that the robotics and autonomous systems strategy has led to any improvement in productivity. A recent Nobel prize-winner in economics said that, hard as he tried, he could not find in the productivity figures any benefit from computers. What you are saying seems to make sense, but do you have the evidence?

Professor Nelson: A number of studies have been done. The honest answer is that we probably are off the pace in this country in the deployment of robotics. I read a recent study by the Copenhagen Business School that looked at the capacities of different industrialised nations for adopting robotic systems. I think we came out as having the greatest potential for improvement, if I can put it that way. The plain fact is that we know we have a productivity puzzle in the country; we are not as productive as we might be or ought to be, and that may well be a contributing factor. Our robot density per number of people employed, or per population, is quite small compared with some other industrialised nations.

Q24 Graham Stringer: That is not true of computers, is it?

Professor Nelson: I was thinking more in terms of industrial robots.

Q25 Graham Stringer: You cannot move anywhere without seeing computers, yet they do not seem to appear in the productivity figures, which are stuck. I am just looking for an explanation.

Professor Nelson: My remarks are very definitely aimed at the deployment of industrial robots. I do not have an answer on the general use of computation. One fact we have learned from the internet revolution and the rapid deployment of computers is that worries about large-scale unemployment have largely evaporated because other jobs have been created. That is an important point to make. Sometimes people worry deeply about the deployment of AI and robotics because of the potential impact on employment. They are right, because lots of activities could be supplanted, but at the same time experience seems to show that you generate other industries and technologies and you have to upskill the workforce to deal with that. I understand the issue.

Professor Muggleton: Looking at the employment statistics for people with computer science degrees coming out of British universities, there is huge take-up. In our department, people are being approached when they have not even finished their degree in the hope that they will be drawn into a variety of different industries.

Dr Buckingham: Obviously, STEM subjects are absolutely vital. If we do not raise the issue of skills, we surely should. We must make sure that future generations are able to engage in this area, and an understanding of STEM subjects is core to that. Robotics is a great way to engage the next generation; they will just lap it up.

Professor Nelson: In UK Robotics Week, which our network is launching at the end of June, there will be public engagement. School children will be educated in 3D printing and visualisation, and robotics.

Q26 Graham Stringer: Just over a year ago the Government said they would establish an RAS leadership council. Do you have any idea when that is going to be established?

Dr Buckingham: It did not happen; there was a change of Government and a slight change of direction. As a community, I am sure we are willing to take up that challenge. We would like to do it. In the bids that went to Government the autonomous vehicle piece got funded and some of the other things did not, so the idea of now moving towards a RAS institute would go hand in hand with setting up a leadership council.

Q27 Graham Stringer: What would you expect from the leadership council? Would you expect it to be a driving force to move the robotics and AI sector forward, or would you expect it to develop governance systems and guide research? What would you expect from it?

Dr Buckingham: Both of those. I think the key word is leadership. It is leadership in this area; it is being responsible and setting guidance and developing standards—all the things we have said. It is being aware of all those things, but also co-ordinating across sectors. You have mentioned farming, space, nuclear, transport and on and on. If we are to maximise our impact in the UK, we want to be learning across the piece. Our community is very well networked, so we can learn what works in farming and what works in space, and then you start to learn faster. A major challenge for a leadership council would be to

say, “How does the petrochemical industry get impacted by what has just been developed in the farming industry and what is happening in space?” There is a co-ordination role as well as the leadership and standards role. That was all the stuff talked about in the UK RAS strategy. The RAS strategy was not paid for or requested by Government; it was a slightly external thing done under EPSRC, KTNs and Innovate UK and we need it to be formally adopted. A letter went from Greg Clark, which responded, and we now need to ram that home by backing it up with investment.¹

Q28 Chair: In the absence of a leadership council, are you confident that RAS 2020 will be driven forward and delivered, or not?

Professor Nelson: We are funding a RAS network, as I mentioned earlier. We put in about £500,000 just to network our key centres in working together. That network is responsible for the UK Robotics Week at the end of June. We are still trying to co-ordinate matters in the absence of a leadership council, but something like that to interact with, or be a governing body for, a national initiative in this area in some sense is going to be very important. It goes back to the whole involvement of Government and regulation as things develop. That pathway needs to be cleared so that people can understand what the next strategic move is to be across the UK.

Professor Jennings: There is a role for some clear, visible flagship endeavours in this area. Lots of excellent work is going on in bits and pieces around the country, and we need something that tries to bring those together and show leadership. As you have all been picking at, this is an important area that has lots of interesting and challenging ethical issues, and something that brings things together and acts as a national forum for doing that is required, and now is a good time to do it.

Q29 Matt Warman: Setting aside the Skynet problem, one of the things that concerns the public is the straightforward loss of jobs. When people talk about previous industrial revolutions there is the sense that long term they created jobs rather than destroyed them. There also seems to be an acceptance that there is a big adjustment hump to get over. Have any of your organisations done any work on what this looks like in terms of jobs? I do not mean just in the long term but in that short-term hump, and who is at risk in that period?

Professor Nelson: An ESRC-sponsored group at the London School of Economics recently published a report on that. I believe they looked at the impact of robotics over the past 15 years or so. They have shown that there is no detriment overall to employment, but there is a shift in the skills base, which is perhaps what you would expect. Some of the lower-skilled jobs decline, but higher up, the skills range actually increases. That is something the ESRC has done. It would be very valuable to have some forward looks on it. Some folks have been speculating about it, and potentially there are very large implications for the economy. If you reduce the number of people you need to employ to keep the country afloat, you have to rethink the economic model. There are some deep

¹ The witness later clarified that “The RAS Strategy was developed by the RAS Special Interest Group coordinated by the KTN with support from Innovate UK and EPSRC. The letter from the Rt Hon Greg Clark MP to Professors David Lane and Rob Buckingham (March 2015) provides a detailed response and makes a number of recommendations.”

questions to be answered that would be worthy of basic research in the economics and social sciences field.

Professor Jennings: There was what I thought was a very well balanced report from the World Economic Forum earlier this year that did an analysis of the jobs susceptible to these sorts of things and how that relationship worked. I commend that to you as a well-balanced analysis of the area.

Q30 Matt Warman: Are you aware of an equivalent UK version?

Professor Jennings: It goes into country-specific bits. It does it at global level and then goes into specific areas.

Q31 Matt Warman: But we have not done anything home-grown beyond what you mentioned.

Professor Nelson: There may well be more work out there, but that is certainly one recent study that I know has been very well done. It was published last year.

Q32 Matt Warman: It seems to me that is a key part of it. As someone who very much wants to get people behind this agenda—I was going to say this brave new world, but that is the wrong phrase to use—we need to win that argument more than anything else.

Professor Nelson: That is absolutely right. We keep talking about a national flagship institute of some sort. That will be a key part. The societal issues, economic issues and regulation of the ethics—all those matters—could be dealt with under that one umbrella, because in that regard it is a very rich landscape.

Professor Jennings: In the Ipsos MORI study of public engagement that I spoke about, jobs were one of the important issues. Even though people were not overly familiar with the technology, when it was explained to them in focus groups that was one of the things they latched on to.

Professor Muggleton: The natural priority in concern over jobs is the implications for education. The technology we are talking about is advanced computer science and engineering. We need to ensure that from the youngest ages people are getting more advanced ideas that draw them into computer science, engineering and so on, and that at the top end we have more people in this area at postgraduate level, not just undergraduates.

Dr Buckingham: In the UK RAS strategy we talked about the next generation of smarter tools. Going back to that point, it is about increasing our ability to do stuff in lots of different ways. The best way to explain that to the public is to show them—physical demonstrations, moving away from hypothetical discussions, which can go down all sorts of rabbit holes, and saying, “This is what’s happening. Here you go, this is the way it works,” and being open and transparent and building confidence in that way. You see exactly the same thing with the petrochemical industry, which would like to use robotics to inspect in very hazardous environments, for instance. The key there is to get the people

who are paying the money to trust and rely on those systems. It is the same question for industry or the public; it is about building trust. The way you do that is by being open and visible and making it understandable.

Q33 Matt Warman: I think you are right. One way of doing that is a central umbrella body, if you like. Do you have a view on whether that should cover all aspects that are potentially affected by AI, because that seems pretty large to me, and—possibly a simpler question—do you have a view on how that central body might be funded, set up and operated?

Professor Nelson: We think about this a lot in the research council world. We probably play to our strengths, frankly. I recited a number of universities that have real strength in these areas. We probably appraise the areas of robotics where we think we have real national standing, and put out a call for expressions of interest in having a hub, for example, allocated to one particular facet. You might have an autonomous cars hub, a manufacturing technologies hub or an environmental investigation hub—we are very successful at that under NERC sponsorship. You might have that collection of hubs. Surgical robots are another area where we excel. We have to make an appraisal of what strengths we have, and we have good data on all that. We know the landscape fairly well. Frankly, we will then have a competition and see what emerges and what other investment we can bring along. That is always one of our criteria, as I mentioned earlier. What industrial engagement will you get? Frankly, we try to join the innovation activities together with those sorts of hubs. We have done this with quantum technologies recently quite successfully. We have four quantum technology hubs where we have fantastic engagement with industry. I would use that as a starting model. Because of the distributed nature of what we have, it makes sense to build on what we have built already, and we could take a long, cool look at what has come out of the research community.

Q34 Matt Warman: Do you have in mind how many hubs you would like?

Professor Nelson: I guess half a dozen. It depends, frankly, on how much investment there is. We would need extra investment from Government for this. There is no question but that there would have to be an extra investment in both capital and resource to make sure we could fund significant critical mass activities over a range of these disciplines for five to 10 years.

Q35 Matt Warman: Have you got as far as thinking about how much that would cost?

Professor Nelson: It would be hundreds of millions. That is as far as I have got. We know that other similar investments on a national scale like that are in the hundreds of millions bracket. I would not want to commit myself any further. We need to think this through carefully. It also depends on the balance between research and innovation. In my discussions with Innovate we feel it is equally balanced. The need for investment to pull through technologies into industrial deployment is about the same as the need to solve some of the very fundamental issues we talked about earlier today.

Professor Muggleton: We have got to a point where artificial intelligence is really applicable now, but the United Kingdom has many claims to having started the whole area

of artificial intelligence. We should not forget that what we have developed so far can be applied, but we need to keep ahead of the game. We also need to continue the far-ranging research that has singled out the United Kingdom as one of the most innovative nations in this area of research.

Dr Buckingham: I would have a slightly nuanced view on both of those. We absolutely have to do the research; we need to do the innovation piece so we create companies that generate jobs and profit. That is the way you get a return on investment. We have to capitalise in this area by building an industry around it. Hundreds of millions is peanuts. This field is going to be huge. We talked about the fact that robotics and autonomous systems are going to affect everything that moves, not tomorrow or in five or 10 years, but that is where we will get to over a number of decades. Everything that moves—in hospitals, factories, on our roads, in space and under the water—will be impacted by this. Then there is the data piece on top of that. These are big issues, and we as a forward-looking country cannot afford to be reticent in this area. We should be taking a leadership role, and we can do so.

Q36 Chair: Can you explain to me why, if we are one of the most innovative and forward-thinking nations in the world in terms of research in AI and robotics, we have been ranked as having the greatest potential for improvement in uptake in the Copenhagen report? What is the gap?

Professor Nelson: Again, I need to be careful and qualify the remark.

Q37 Chair: Is it accurate? Can you give me the context?

Professor Nelson: This was the uptake.

Q38 Chair: I understand, but what is happening there?

Professor Nelson: It is something about industry not investing in innovation. It is probably beyond my skillset to diagnose the problem here and now. You have to be careful not to generalise, because some companies are absolutely up to speed and doing this very successfully, but others just do not have the impetus, for whatever reason, to invest in some of these technologies. That is what you read in the literature.

Q39 Chair: But the point of the proposition is that, if you invest in AI and robotics and it is taken up throughout the economy, it will lead to great productivity gains and benefits in all sorts of areas of society, so we need to understand the take-up bit as well.

Professor Nelson: One example where clearly investment has been made is in South Korea. I think you will find that their robot density is very high. There is probably a connection between investing in this technology and getting industrial uptake. I would like to think so.

Dr Buckingham: I was thinking about the leadership council piece. That must be a clear element of it. A leadership council would be made up of industrialists as well as

academics and interested parties from across the piece. Surely, that must be one of the key issues that will be addressed. What is the return on investment? How do we embed this technology in the UK and make sure that it does not just disperse? We are doing great research. How do we make sure that that sticks in the UK and how do we create the companies that come from it? I suppose one answer to the industrial take-up piece is that we have to learn from the stuff that has not gone quite right in the past. These are the early stages of a really exciting technology, and we should be in it.

Professor Muggleton: As an example related to your question about whether we are so advanced in this area, a book for the public about machine learning was written by an American last year. He laid out the five main areas of machine learning. If you look at the leaders in the five areas he identified, three of the five either came from the UK or are still working in the UK, so we are among the most advanced in the most advanced technologies.

Q40 Chris Green: Sticking with the theme of skills and education in the UK, pretty much every inquiry we have done so far has highlighted that there is a crisis to some degree in people learning STEM subjects. Has that had an impact on robotics and artificial intelligence in the UK?

Professor Muggleton: We are starting to see more people coming through with mathematics and science backgrounds, which is encouraging. We need to do more, because this is a technology that will be critical for the leading countries of the world, as we have all been saying. We are already doing well but we need to do better.

Q41 Chris Green: Has it had an impact on the advancement of robotics and AI so far, or is it perhaps more a concern for the future?

Professor Muggleton: It has had an impact on the kind of postdoctoral and doctoral work going on, because the people involved have the right skills to be able to do that and take it forward. It is also having an impact on industry within the country. For instance, in the last five years the majority of our graduates at Imperial College were going to the financial industry after they got their undergraduate degree. They are now increasingly staying on to do postgraduate work because the big take-up is in the tech industry. For instance, companies like DeepMind are hoovering up quality people and providing them with employment in London.

Q42 Chris Green: In that sense, there is quite a strong message at the moment that you do these subjects at school, go to university and there are fantastic opportunities and places in industry to have a great career afterwards.

Professor Jennings: Yes, absolutely, but still not enough people are doing the basic computer science required. That starts at school all the way up to university and so on. However, this technology is very pervasive, so we do not want just more computer scientists; we want many other elements of society with different backgrounds. We want medics to be able to understand what machine learning algorithms can do for them and what artificial intelligence is relevant to them. Some of it is about growing the core—the

army of computer scientists who will make technical advances in these areas—but the rest is disseminating broadly the potential impact of these technologies.

Q43 Chris Green: At the moment is there almost a silo? Many other areas will look at it and say, “This is not something we would do or something we would touch.”

Professor Jennings: It is not a routine part of a medical undergraduate course to learn about AI. It could easily be in the future.

Professor Muggleton: One of the routes into that is interdisciplinary research, which has been happening. Increasingly, AI is involved in medical research and other scientific areas of research because it is valuable in those research areas. Whenever there is big data, as there is in most science and engineering, AI techniques are often the technique of preference in those areas, and people learn about them through interdisciplinary research.

Q44 Chris Green: Is there much competition between industry and the innovation and research side in this field?

Dr Buckingham: What do you mean by competition?

Q45 Chris Green: Are there sufficient people with the right skills and qualities to man both areas?

Professor Nelson: Our centres for doctoral training in this field are regularly over-subscribed; I am told there are about 10 applicants for every place. That is for postgraduate training in this field in a variety of robotics and AI-related areas. There is real enthusiasm out there among the young to get involved in the field. Probably one needs very well-rounded engineers in many activities, who understand some of the algorithms in computer science and so on, as well as the electro-mechanical aspects of the devices they have to work with. There are calls yet again for more engineering education, and we have heard that many times before.

Dr Buckingham: Let’s not get this wrong: we need more people who are good at STEM. That is an absolute given. It is vital that we have those people.

Q46 Chris Green: You highlighted earlier the investment going into South Korea and Japan, and the leadership in the United States of America. Are we in any way at risk of losing overseas a lot of our people who are coming through? Are we going to have some kind of brain drain?

Professor Nelson: I confess I do not have any data on that to hand. The job market in these areas is very fluid. The tech companies in California attract a lot of our top people out there from graduation. There is no question but that that happens. I am not exactly sure of the magnitude of those effects, but clearly where you have exciting new technologies and industries you will attract people to work there. It has already been mentioned that London is doing remarkably well in attracting jobs.

Dr Buckingham: It's stick and stay. How do we get them to stick and stay in the UK? Gosh, that sounds terrible. Don't use it as an advert. We should be investing so that that happens. It is not just about the research but about innovation and jobs, and then you get investment from the companies. Companies are coming here, for instance, around driverless cars because they see tests happening here that are not happening in other parts of the world. You get virtuous activities that are positively reinforcing, and that is where we should be. We should take a positive view. It links with skills. You start becoming excellent in an area, and then you worry less about brain drain to South Korea, China or America, and you start to get people coming here. This is a time when we could get on the front foot.

Q47 Chris Green: It is a distinct area in terms of the skills that we need. We talked about people coming through, perhaps getting their first job. For people already in the workforce there is going to be a significant impact. Blue collar workers of the past have been affected by automation, and straightforward manual processes are being done by robots. With artificial intelligence perhaps there will be far more impact on the white collar sector. Do you have any particular concerns about retraining? Are we going to see large numbers of white collar workers who perhaps when they are 50 years old find there are no opportunities to retrain, and are left on the side and cannot re-engage in the job market any more?

Professor Jennings: The job market is changing, as we are probably all aware, in terms of the constant need for retraining in the light of these technologies. I see these technologies as an augments of many of the professional white collar activities, as you describe them. I think they will generate a whole load of new jobs around them in the way data science and data scientists created them. It has gone away from augmenting basic operations on data to letting people concentrate on higher-level things, and then a whole load of data scientists are concerned with curating and presenting that data and doing the analytical tools for it. Continual retraining is going to happen ever more. The view that you train at university when you are 18, have one career and that is it will disappear over time. It is important that universities and individuals are willing, and we have structures that let people retrain continually.

Q48 Chris Green: Would you go along with the idea that we can have fantastic innovation in robotics and artificial intelligence, but when it comes to the person who actually uses those ideas, not the person developing them and innovating, it is a kind of grey box? As long as the application and interface between you and that technology is done in the right way, for most people with a particular job it should not be too much of a challenge.

Professor Muggleton: That is the key point with white collar jobs, because they require a lot of interaction, verbal thought, abstract thinking and so on. We need to have technologies aimed at being able to interact at that level, not the knee-jerk level. Those are starting to be developed at the moment, and we need to keep on top of that. There are many indicators that some of the biggest potential applications are exactly what Professor Jennings said, essentially collaborations with additional power that can make people more productive by having an interactive system that can effectively communicate with them.

Q49 Chris Green: And, hopefully, deal with our productivity gap.

Professor Muggleton: Indeed.

Q50 Carol Monaghan: I am wondering about the current gender balance within the robotics and AI industry and where you see that being in perhaps five to 10 years' time. What has been done to improve it, if improvements are required?

Professor Nelson: I am not sure of the gender balance in that particular sphere. I know that, in engineering and physical sciences, in our research community sadly only 16% are female, which is too low. It is a terrible waste of talent. We are doing everything we can across the research councils. We have an action plan on diversity to do our utmost to rectify the gender imbalance wherever we can. On our advisory bodies and teams and on our council, we are getting good representation now. We are trying at least to do that. We can fix that; it is within our grasp to do it, but while A-level physics has only 20% female participation it will always be difficult to overcome the problem. It goes right back to school and getting more females coming through the A-level system with physics, maths and so on.

Professor Muggleton: That is right. Efforts are being made. There are programmes that support females going into engineering. There are departments that are doing fantastically and others that are not. I think our department is unique in that we have more female than male professors, and the head of department is a female. I do not think that is happening across the board, but the more it happens the better for the subject, particularly since this is not a technology that will be used just by men.

Q51 Carol Monaghan: From some of the things described today, I can see girls being really interested in getting their teeth into it, if they fully understand what opportunities are offered by robotics. Is that message getting to them? It is an ongoing problem, as Professor Muggleton said.

Professor Jennings: I think you are right. AI and robotics open up and touch on many things that girls are interested in: creativity, working in teams and those sorts of activities, which are more appealing, rather than maybe just traditional computer science, which is stereotyped by males programming on their own. AI is a lot more inclusive and team-based, and that should play well to gender.

Q52 Jim Dowd: In response to Graham's apocalyptic vision of "Terminator" and the rest, I believe it was also a theme in an infinitely better science fiction series—the "Star Trek" movies; in "The Wrath of Khan" the premise was exactly the same. Prior to both, a chap called Isaac Asimov wrote the original theory of how humans interface with robotics. Is that simply fiction, or could something along those lines be codified and engineered to prevent the apocalyptic vision of Mr Stringer?

Dr Buckingham: Asimov's three laws are awesome. How long do we have? The stories that followed, exploring how those laws break down, are fascinating. They are brilliant. It is a commentary about people, not about robots. These things are often not about robots but about human behaviour and interaction. An awful lot goes on inside our heads that we do not yet quite understand, and we should not underestimate how amazing we are. I said

earlier that this is about people, not about robots and how we use them. The science fiction pieces—I am starting to ramble.

Q53 Jim Dowd: Essentially, the premise of it was hierarchical programming. There is an overall or complete prohibition of any action or inaction that causes harm to human beings, blah, blah, blah. Everything else was secondary to that. Is it not possible to engineer that?

Professor Nelson: I referred earlier to the EPSRC principles of robotics, where we got some of the thinkers in this space in a workshop with AHRC, as it happens. They replaced the Asimov laws by five other principles. I am reluctant to try to recite them now, but you can find them on the EPSRC website. That references Asimov's three laws, so they have been rethought, if I can put it that way.

Chair: Thank you very much. If I can draw us back from the apocalyptic abyss for a moment, I thank all of you for the time you have taken to give evidence. This is a complex area with lots of different component parts. We are going to be conducting a reasonably long inquiry into it, so I suspect we will have more questions to ask you in writing. I hope you will respond to us so that we can produce a report that is rigorous and engages with the topic. We believe that is important to raise public awareness and to make sure that Members of this House have an appropriate level of engagement. Thank you all for your time today.

Examination of Witnesses

Witnesses: **Richard Moyes**, Managing Partner, Article 36, and **Dr Owen Cotton-Barratt**, Research Fellow, Future of Humanity Institute, gave evidence.

Q54 Chair: I welcome both of you to the table. You sat through our first panel and heard the questions asked, some of which were earthbound and some not. Can I start by asking both of you the same question I put to the previous panel? How close are we to seeing intelligent general purpose machines that can handle complex problems, and what would be the implications of that for society?

Dr Cotton-Barratt: That is a great question. It is extremely hard to have confidence on this question. It seems that over the next five years we will see significant advances—I agree with the previous panel—that will very likely bring large benefits to society. Over the long term there is the hypothesis that we may at some point see artificial intelligence get up to human level in almost all domains. It is hard to say when that will happen. There have been a number of surveys of people who work on AI to try to get their estimates on this. The estimates they give are surprisingly close. The average expert, across a number of surveys, tends to give the date by which they think it is as likely as not as between 2040 and 2050, and says there may be a 10% chance of seeing it within the 2020s. Maybe there are elements of optimism because they are the people choosing to work in this field, but they know the technology better than outsiders and we should take their judgment seriously.

Q55 Chair: Richard Moyes, the potential for artificial general intelligence in the 2020s is 10%. Stephen Hawking and other scientists wrote an article in which they said that AI was potentially the best or worst thing that could ever happen to humanity. That is quite a polarised position to take. Do you think it is a helpful way for the debate to continue? If not, how should we be thinking about it?

Richard Moyes: For a start, I have no idea of the future of general intelligence development; it is not my area of expertise. It is very important that people are debating and considering that slightly longer-term trajectory of where the technology is taking us, but in our involvements in these issues we are focused much more on the role of specific artificial intelligence and algorithmic systems within weapons systems as we see them developing today. We are focused much more on how computers, sensors and algorithms are being incorporated into weapons systems more and more, and the implications that has for the development of weapons in the comparatively near term. In many respects, that is quite a separate issue from the implications of broader generalised intelligence in society.

Q56 Chair: The principles that stem from it are the same, whether we are talking about artificial general intelligence and its implications for society or about specific applications that have ethical concerns. Is there enough serious research going on into the ethical implications and how legislators and regulators should be responding?

Richard Moyes: I do not know about the general framework. Our work is focused substantially on the UN convention on conventional weapons, which meets in Geneva under the framework of a treaty that basically has the power to regulate and prohibit certain weapons. Under that framework, delegations from different countries come to the table and bring their perspectives to bear on what controls, or lack of them, they think should be enacted in relation to weapons systems with developing autonomous capacities.

Looking specifically at the UK's engagement in this area, there would be value in seeing whether the specific types of policy position that the UK has been taking in that context with respect to weapons could in some way be integrated more, perhaps with broader consideration around how these capacities—computer autonomy, AI and algorithms—are going to be managed in other areas. From our perspective, we have some concerns that we might take a rather more permissive attitude in the context of discussions on weapons than we would want to take in other areas of society where perhaps we imagine the risks bear more directly upon populations to which we are accountable. From what we see in relation to weapons, there is certainly scope for more integration of thinking and perhaps broader policy oversight than we are seeing in relation to the Ministry of Defence on its own.

Dr Cotton-Barratt: On the question of more general and powerful artificial intelligence and its impacts on society, I agree with the previous panel that in decades to come we may see that everywhere in life. If it happens it could be very transformative, and that is the thought behind Stephen Hawking saying this could be the best or the worst thing ever to happen to humanity. It could also be somewhere in the middle, but if it is to transform the way our society works, it is extremely important. Even if it is still some decades off, it deserves attention and our trying to understand what could be good and what could be bad about it. Are there any ways to make it more likely that we get the good rather than the bad outcomes?

There is a small but growing research community looking into these questions. I think the UK is world-leading in this at the moment. The Future of Humanity Institute, where I work, and more recently the Cambridge Centre for the Study of Existential Risk are looking at these questions. There is also the Leverhulme-funded Centre for the Future of Intelligence, which I think will involve people from Oxford, Cambridge, Imperial College and Berkeley. I do not know of any other academic centres elsewhere in the world looking at this. Right now we have the intellectual leadership, and I think that over the coming years and decades that will more and more become an extremely important area, and it will have to tie up with the practitioners of research into artificial intelligence.

Q57 Chair: We are talking now about research-level ethical questions and long-term questions, but in practical terms what are the immediate questions for legislators and regulators that you think we should be thinking about?

Dr Cotton-Barratt: It is too early to have a definitive answer for many of those long-term questions, and as legislators you should not be trying to create one. However, when we are making rules and setting up systems today to deal with the incoming improving technologies that we are going to see over the next few years, we should try to make sure that they remain robust to possibly quite radical improvements in capabilities in the systems. We should not just look for a solution that will work now, but try to imagine whether, if we have much more powerful systems, that is the way we want it to work in our society. There can be a bit of stickiness in the way things are set up. If we have good direction when we are setting out on that, it might help to make sure we remain well placed in the future, and may also serve as a lead that other countries can follow.

Q58 Dr Mathias: The Ministry of Defence says that the operation of UK weapons systems would always be under “human control”, and Article 36 says that this should be “meaningful human control”. Can you explain what that is, and why you think you need to make that clarification?

Richard Moyes: Thank you for the question. The word “meaningful” is there to indicate that we need to debate and define the form of human control we are talking about. When we talk about complicated technological systems, we cannot just take human control as being self-evident in some way. There are a number of areas where we can imagine different examples. If a weapons system is operating in a situation where a human defines the specific object it is going to strike against, and if it is activated and strikes that object, we can be fairly clear that there is human control. If there is a range of objects of a similar type and a human stipulates a category of object types within a certain area, and the weapon is activated and strikes at objects of that type within that area, there is a form of human control but it is perhaps less specific than in the first example. Similarly, you can step further back and say that the weapons system has a broad category of possible target types programmed into it; it is set out across a broad area to operate over a longer period of time. Now, all of a sudden, where the actual physical force is going to be applied in that situation is rather unpredictable in the context of the commander who has initiated the attack. In that situation we are seeing a different form of control from that in the first example. All those examples are relatively proximate to certain types of existing systems.

We see different levels of control being enacted, so within the international community the issue of autonomous weapons is being debated in Geneva. We are trying to build a structure to that conversation that is focused on thinking about the human elements of control and human judgment in control that we want protected in the development of these technologies. The term “meaningful human control” is a political policy device to try to get a number of actors on the same page and build up a conversation about where the threshold of control should be.

We have written some materials on the types of factors that bear upon control, some of which are in the technology itself in terms of its reliability, predictability of operation and perhaps transparency—whether you can see how it works—and some of which are information based and are about what the commander knows about the area where the system is operating. That bears on predictability and constraints in the area of operation, and the systems of accountability within which those uses of force are taking place. We think you could lay out a fairly clear framework of key elements of human control. We would like to see the international community doing that together as a collaborative process of policy development. We would like to see the UK encouraging that and participating in it, but as yet it has been a little reluctant in that respect. We are hoping that next year, when perhaps a more formalised group of governmental experts discuss the issues, we will get the conversation focused more on laying out what we think are the elements of human control that we want to protect as technology develops in the future.

Q59 Dr Mathias: The UK Government also say that humanitarian law is sufficient for the legality of weapons systems and we do not need separate legislation. Why do you think a separate ban is needed?

Richard Moyes: Separate but perhaps still within the framework of international humanitarian law. We are certainly not looking to reject the utility of the humanitarian law framework. Of course, technically it is applicable only in situations of armed conflict, and we need to be aware of that as a legal specificity. Our particular concerns are that, as it stands at present, international humanitarian law regulates the use of force in the framework of attacks, where attacks are particular applications of force for military objectives. It may be that more than one weapon is used; it is not a single bullet fired at a single person, but it is a relatively contained application of force. Military commanders have legal obligations around attacks to make certain judgments as to whether or not an attack is proportionate, and other such rules.

One of the things we are concerned about is that, as these technologies are able to operate over wider areas and perhaps longer periods of time, the concept of what is an attack gets bigger and bigger. Coming back to the objects on the table example, we might understand that attacking a relatively small group of vehicles close together is conceptually a relatively coherent matter. If we then imagine vehicles spread over a much larger area of a country, or across towns and cities, the use of a weapons system can progressively move from one to the other, striking against them. That seems to be a much bigger version of an attack. We are concerned that a military commander in that context does not have the contextual understanding to make informed judgments about where specifically force will be applied and what the likely implications of that are going to be. By stretching out the concept of attack we are stretching the structural fabric of the law itself, and that is

something we should be very wary of. We should be pulling more towards regular human/legal judgments about the application of force in conflict situations, essentially to ensure that we are applying that human judgment on a more rigorous and regular basis.

Q60 Dr Mathias: It implies a lack of confidence in the military.

Richard Moyes: Not so much in the military. Unless we develop common understandings of how these things are to be interpreted, the whole international community has the capacity to interpret the legal frameworks differently. Already under international humanitarian law you get quite different interpretations from one person to another and from one country to another. It is quite an open, flexible framework. We feel that developing specific rules in relation to the role of autonomous weapons systems would allow us to ensure that all countries are approaching this from more or less the same perspective. There are still likely to be some fuzzy lines in these areas, but we think that a legal framework would get states more clearly on the same footing. New legal rules have a communicative function, which will help to guide where research, technological developments and commercial interests go, as well as bringing a wider sense of what the public and others consider acceptable behaviour in conflict situations—a sort of stigmatising standard-setting function. We think that all those things can come from new law rather than just relying on the existing legal framework.

Q61 Dr Mathias: Acceptability is not going to change, whether it is autonomous or non-autonomous.

Richard Moyes: When you start to think about machines making decisions about who should be killed or where force should be applied, quite a lot of people feel slightly instinctive revulsion or rejection of that notion. There are all sorts of issues in considering what are decisions and the like, but we often find a general sense that having machines making decisions about who gets killed and where force is applied in context makes people queasy about the direction we are taking in relation to technology and society. When I talk about acceptability around those sorts of issues, I mean that we want to strengthen, from our perspective, the sense that having machines making decisions about life and death matters for people is not something we want to encourage; we want to try as much as possible to keep that as essentially a human function.

Q62 Dr Mathias: Do you see any benefits from an autonomous weapons system? If we talk about friendly fire, collateral damage and more targeted offence, do you see that as an advantage?

Richard Moyes: There are potentially significant advantages in autonomy and sensors in computers in relation to weapons systems. I do not generally think of myself as a proponent of weapons development and the like, but the key issues we are concerned about are the identification, selection and application of force to targets and machines making that kind of choice, but, at the same time, we can see in weapons systems a capacity for more accurate application of force more narrowly to specific objects. As long as a human is in the background applying it, potentially it has very positive impacts for the surrounding population.

Dr Cotton-Barratt: This might be a place where you try to take a long-term perspective as well and decide how we want our society to be organised around this. The question of whether or not we allow machines to make decisions about taking lives may be a natural bright line; it could be a natural stopping point and we should think carefully before crossing it, rather than saying that in the short term it looks as if it might be beneficial, because from the longer-term perspective it might be the better place to stop. It is a complicated issue. I am not sure.

Q63 Graham Stringer: When you are dealing with systems that can adapt, learn, plan and change what was intended, how can you verify that those robots are doing what was originally intended?

Dr Cotton-Barratt: That is a great question. There should be more research on exactly that question. This is one of the potential challenges to the future deployment of these systems. At the moment, the systems tend to be reasonably simple, relative to what they might be in the future. We can understand the underlying algorithms but not how they are making decisions. One of the previous panellists said we do not really know how the machine was better than the best human Go player. At a high level we can say that is how it learns, but we do not know how it thinks about the game of Go. If we can build better tools for understanding that as it proceeds, it is more likely that we will be happy deploying even very advanced systems in the world and giving them autonomy.

Q64 Graham Stringer: I am not sure I completely understand either the questions or the answers. Can the workings be transparent if it is not possible for humans to determine why the system is reaching that decision? How can you have accountability and control of a system if you do not know why and how it is coming to its conclusions?

Dr Cotton-Barratt: It is a real obstacle. There are degrees of transparency in terms of who it is transparent to. Is it transparent to the people who wrote the computer program? Is it transparent to the people who are using it? Is it very publicly transparent so that anybody who wants to understand the system can look it up? There are also degrees of resolution and how fine-grained is the understanding of its thinking. To the extent that we can improve and push up our ability to understand that at a fine-grained level, we can have more confidence in these systems. That was the direction Professor Muggleton said was good, and one we should be exploring further. I absolutely agree with that.

Richard Moyes: From our perspective, when we are pushing back a little on technology developing in a certain way with respect to weapons, it is sometimes presented to us as if technology is on a relentless track in a particular direction and we have no power to move it either way. Of course, it is important not to buy into that approach, but it makes me wonder a little bit whether some of these factors may produce something of a natural break or a natural steering in certain respects. If within a military framework we are going to ensure that somebody is accountable for the use of a system, how they are accountable—the way accountability bears on them if the system is not transparent—may become a problem for them as an individual. The key issue is that, as long as we enforce accountability, maybe people themselves will say they do not feel comfortable being held accountable for a system when they cannot quite understand its functioning and cannot be completely sure what it is going to do. This links a bit to issues about building trust in

technologies. Military actors like to trust their equipment; they are in situations where they depend on it, and systems that are perhaps too complicated to understand and are slightly unpredictable in their behaviour may not engender trust within those institutional frameworks. I am not suggesting in any way that we should be relying on those forces to put in place the best policy responses, but there are some factors that will perhaps make these systems problematic in an operational context.

Q65 Graham Stringer: If somebody sets off a weapons system and it kills a lot of people and it was the weapon that did it, not the person, that is a very clear example of the real problem in terms of the law and accountability. Do you think there is a similar problem when it comes to intelligent learning machines carrying out surgery or medical procedures? The machine may make a decision, but the patient cannot give their permission for that decision because they cannot have an interaction with the machine. They can have an interaction with the person who sets off the machine, but if there is no transparency do you think it will limit the practical applications?

Dr Cotton-Barratt: It could.

Q66 Graham Stringer: That is a political answer.

Dr Cotton-Barratt: People may be unhappy with it in some cases. People may be willing to make trade-offs if the statistics show that it is safer to have opaque systems that are automated than not to have that automation, but there is a cost. People may be unwilling to take it, and we should perhaps be unhappy with them taking it if they do. This supports a push towards developing meaningful transparency of the decision-making processes, as well as just developing better ones.

Q67 Chair: It is not just about transparency. Presumably, not many of us will understand all the algorithmic processes and decision-making functions, but we might learn to trust an active verification process that stands between us and a machine learning system that gives us confidence that the system is going to do what it says it will do and it has not learned to do something different. How far off realistically are we from having anything like that, and who is working on it? Are we world leaders in that as well as in ethics?

Dr Cotton-Barratt: In my mind, that is another form of transparency. It seems like a useful kind of transparency tool that does the interpretation and turns an opaque system into one you can understand. Maybe you do not understand all the details but you understand some of the salient high-level ones, and that it will do what it says it is going to do. There are a few people researching this at a broad level. One can try to study the question in two ways. One can look at what we understand about automation as a whole and what properties are going to be needed. One can also look at what machine learning systems people are building today, and how we could implement this on top of that. I do not know enough about how much research of the latter type is being done. Of the former type, I think there is a little but not much and there could be scope for a lot more.

Q68 Chair: Is it too early, or is it just that not enough people are doing it?

Dr Cotton-Barratt: It is mostly that not enough people are doing it. It has an odd property. For most technologies I think it would be too early. The closer you get to having the technology, the easier it is to do the research, because you can work with things that have been built and understand how they work. It is only because of the potentially extreme transformational nature of this technology that it seems important enough to try to design ahead for how to interact with more powerful systems.

Q69 Chris Green: What levels of autonomy should be permitted in robotics and AI systems? I think we have covered bits and pieces, but what level of autonomy should be permitted?

Richard Moyes: Characterising and structuring it in terms of levels could probably be done in various ways. I am not sure I am in a position to lay that out. From our perspective, the key issue of concern is the selection of targets against which force is applied. On autonomy of movement of a system to a target, I feel relatively comfortable. I am sure some issues can arise from that, but the specific focus of our concern is the identification and application of force to the target being in the hands of the weapons system. I am not quite sure how that is characterised as a level.

Q70 Chris Green: If you are going to deploy a weapon, should there always be a human in the loop?

Richard Moyes: Yes. From our perspective, we think a human should be specifying the target against which force is to be applied. To give a bit of breadth, that target may be a group of objects; it may not have to be a completely unitary object. We accept that in the use of force you can apply weapons systems against targets comprised of various objects. There is some potential for breadth, and within that process a machine may apply itself most efficiently to those objects, but there have to be some boundaries to the concept of a target.

Q71 Chris Green: Would we be more comfortable saying that artificial intelligence could restrict what we do—a version of the dead man's handle in a train? If we can apply that, would we be far more comfortable about it, or does it have the same ethical or moral considerations?

Richard Moyes: Those are interesting questions. It is possible to feel more comfortable with that. From a personal internal sense, if a weapons system was being applied to a target, and on seeing a red cross on the roof of a vehicle in the immediate vicinity of its detonation it deactivated itself, I would struggle in moral terms to see that as a problematic aspect.

Q72 Chris Green: In that kind of instantaneous decision, an automated system would be able to identify it and switch off faster than a human could react, so you give control in that sense.

Richard Moyes: But there are slippery slope issues in these areas. You would need to frame it carefully, but there is potential.

Dr Cotton-Barratt: More broadly, looking beyond weapons systems, when considering what level of autonomy should be required, this is definitely not a question we are going to be able to settle now. I see two main forces that affect the level that we think is appropriate. One is what is safe with the current levels of technology. We are coming to the point where it seems that we can have autonomous cars, that it is safe and it is a reasonable thing to allow. Ten years ago that would not have been a reasonable thing to allow, because the technology was not there and it would have been damaging. There is also a kind of limit. If we imagine much more capable systems, do we want full autonomy? There might be a reasonable principle that, whenever an important enough decision was being made, the human should be meaningfully in the loop. You can make autonomous decisions for small things, but for larger plans you should meaningfully consult a human.

Q73 Chris Green: I do not know how a car would in any way have the morality of a human, but in years to come we would want a value alignment in some way between the system and our moral codes. How would you go about formulating and designing that?

Dr Cotton-Barratt: That is another big question. A lot of the people who to some extent are concerned about the longer-term impacts of artificial intelligence focus on researching exactly that key question. When systems are powerful enough that we cannot necessarily understand and monitor everything, and they may have large impacts on the world, we still want to make sure that the actions they take are aligned with our interests. People are currently at the stage of building research agendas to try to address that question. There are a number of different ways to try to break it up and say, “If we answer this question and this question, it looks like we’ll be in a better place.” I think it is still early days for that area.

Q74 Chris Green: We need to make sure that computer scientists and engineers and social scientists increasingly work together so that as these technologies are being developed the moral perspective is there, but also that the moral characteristics of artificial intelligence are developed at the same pace. We do not want what the social sciences can bring to be one, two or three steps behind.

Dr Cotton-Barratt: I agree.

Q75 Chris Green: There was reference earlier to “Terminator”, so I will throw in “Demolition Man”. There you had artificial intelligence, which was activated by the Government of the day and had certain constraints. If people swore or behaved in a certain way it could restrict their behaviour; they could be fined or punished in one way or another. If you constrain people’s activity with artificial intelligence, monitoring what they are saying and doing, at some stage you get to 1984 and limit what people say and, therefore, what they think. We are a long way from anything like that, but it must be one of the questions you are considering.

Dr Cotton-Barratt: It is a real concern. We have a robust democracy in the UK and I am not too concerned about that happening here, but in some countries if leaders had access to technology that would enable them to do that maybe they would use it in that way.

Q76 Chris Green: Surely, you can see such an application at the moment. Not wanting to go into the Regulation of Investigatory Powers Act and that kind of thing, you can see how a Government would want to be concerned about certain words and activities, and for the state to clamp down on them and constrain them. Once it has started in one area and it can deal with that group, what about other groups? It is easy to see that slippery slope.

Dr Cotton-Barratt: It is a big question. It is a political question and one on which I do not have that much expertise. There may be technical tools people can develop that allow the monitoring of things while maintaining privacy in a meaningful way.

Q77 Chris Green: We always have to be sceptical of politicians and ensure that the Government do not have too much power in this regard.

Richard Moyes: Linked to that, again in relation to weapons stuff, one of the things you see in those systems is necessarily an encoding of objects in the outside world into categories of target or not target. Obviously, it seems less problematic if we are talking about armoured fighting vehicles, but if we start to move to a situation where we are encoding people with certain appearances or behavioural patterns, all of those things represent quite a pernicious process of reducing people to data points and then bureaucratically administering force around them. We are certainly not there yet, but that underpins some of the anxiety about how we start to encode ourselves and society in particular fixed ways. We have seen in the past that, where bureaucratic structures begin to encode people in those kinds of ways, it is potentially highly problematic.

Q78 Matt Warman: This question is related, but different. With the recent headlines in the *Daily Mail* about Google getting access to the Royal Free's data, there is a sudden sense that AI is a mainstream concern but perhaps not as well understood as it should be. Do you think that AI in principle poses new challenges to things like patient confidentiality that we have previously taken for granted, or is it the same rules but just with a different set of tools?

Dr Cotton-Barratt: It may increase challenges. I am not sure whether in the case of this particular data there is any AI involved, but in the future there may be challenges coming from it. There may also be large benefits. If it can automate the processes and increase consistency in judgments and reduce the workload for doctors, it could improve health outcomes. To the extent that there are challenges, essentially it means there is less privacy from the same amount of shared data, in that people can get more information out of a limited amount of data. We will want to find ways to handle those challenges, and that may be about making sure that the data is held in the right places and is properly handled and controlled. I do not know the details. My understanding from the outside is that they are very much aware of it and want to make sure that the data is held securely.

Q79 Matt Warman: Is it in a sense less worrying if only artificial intelligence has looked at, say, patient data rather than human intelligence?

Dr Cotton-Barratt: It may be. It depends on how the artificial intelligence has been set up and what it will do with the data at the end of it, but if it is just kept internal and never reports to any humans the confidential data that is useful for the algorithms, there is no scope for anybody to abuse the data.

Q80 Matt Warman: Who should be responsible for drawing up the rules on whether it should be kept private?

Dr Cotton-Barratt: I do not know enough about good ways to reach a robust conclusion on governance. You will want to get a set of rules that industry and the NHS have signed up to and that the public are happy with, but I do not know what the best process is for producing such a set of rules.

Q81 Matt Warman: The Government are going to set up a council for data science ethics. Do you think in that sense that we need a council for AI ethics to try to work out those sorts of problems?

Dr Cotton-Barratt: Data science is a broad category, but it involves a lot of machine-learning algorithms that are the same kind of algorithms as today's AI algorithms, so the ethics for data science in this case may be about the right level of foundation for it now. Maybe in a few years, if AI becomes more sophisticated, we will need to revisit that question.

Q82 Chair: The message we have received quite clearly from both panels is that we are at quite an early stage of development with a lot of these technologies and we do not know where this revolution is going. If I can update some of our cultural references in the Committee, there is obviously quite a lot of misunderstanding among the public and low levels of public trust in AI and robotics, some stemming, for example, from "Captain America: Civil War"—just to show that I am up to date. What do you think the Government, or other bodies, can do to try to help prepare society for the changes that are very likely to be coming as a result of these technologies?

Dr Cotton-Barratt: When establishing guidance or legislation for the nearer term, the Government can try to look out to the longer-term impacts and improved capabilities and bear them in mind, and design with them in mind. The Government can try to support more work in understanding the eventual impacts. There is quite a lot of uncertainty. Maybe making machines act in a way that is aligned with human interests will turn out to be easy; maybe it will turn out to be hard. Some reasonable people think both, but I do not think we should just assume that, therefore, it is easy. The Government could do that by supporting the academic research community or perhaps by internal studies. The academic community is going to include social scientists as well as AI scientists, and there will be legal questions as well. There are a lot of questions. The more we can resolve them in advance of having to have answers, the more we can check that the solutions we have are robust.

Richard Moyes: In very concrete terms, we would like to see the Government over a longer period of time supporting more formal discussions in the convention on conventional weapons and lethal autonomous weapons systems. We would like to see the Government using that as an opportunity to articulate and develop thinking around the forms of human control that should be considered necessary in the use of force. It would need a little bit of preparatory work at national level before we were in a position to do that, but there is an opportunity for the UK in a diplomatic landscape to have an influential position on how we orientate the role of computers in life and death decisions. We should take it, rather than watch the debate drift away from us.

Chair: I thank both of you for the time you have taken to give evidence today. This is a complex and challenging area, and we are grateful for the time you have taken to take us through it. We may well come back to you with questions to clarify different points. I hope you will respond to us. Our next evidence session will take place during UK Robotics Week, very appropriately, but for now I draw this session to a close.