In our Time Programme 29
AI Computers

Melvyn Bragg : Hello, it's 31 years since Stanley Kubrick and Arthur C. Clarke gave us HAL , the archetypal thinking computer of the film "2001-A Space Odyssey". But as we head towards the millennium, are we any nearer to achieving the thinking , feeling computer? Or is it just a dream? And should it remain one?
Igor Alexander is one of the leading figures in artificial intelligence. Professor at Imperial College London, he's credited as being the first person to design a computer which could recognise a human face. He's also the inventor of MAGNUS , a neural computer, which he says is an artificially conscious machine. He's currently writing a book called "Towards Conscious Machines". The Philosopher John Searle]. Currently Professor of Philosophy at the University of California, Berkeley, his publications are numerous, he's just published "Mind, Language and Society".

Igor Alexander, can you tell us the difference between machine consciousness and human consciousness?

Igor Alexander : Yes, it's every bit of difference in the world! The reason we talk about "machine consciousness" at all, is because it tells us something about real consciousness. What we do on a computer, is to basically simulate the biological end of the brain, because in 40 years of artificial intelligence, this has been sadly lacking. People have been writing programs which do smart things, and they might even beat Kasparov at Chess, but they haven't really handled the way in which biology creates consciousness. Now, what we do in artificially conscious systems, we basically simulate the brain, and allow some behaviours to emerge from that, and we then then study these behaviours and say "Oooh that's a bit like consciousness in a human being", or, "It isn't at all", and it's which structures of the brain give us more of a conscious type behaviour that are interesting and those that give us less, that gives us a bit of a handle on the biological basis of consciousness. But the computer and the machine almost get out of that equation.

Melvyn Bragg : So you're saying that a conscious machine is an oxymoron?

Igor Alexander : Well, it's an interesting oxymoron. I like to think that in fact it does make sense as long as you use.....you put the word artificial in enormous capital letters, and consciousness in tiny little nine point print. The object of the exercise is to give us a bit more understanding of ourselves. Of course, as soon as you say the words "artificially conscious machine", people think that you're building something that looks a bit like Swarzenegger, and it's going to rush out of the studio and kill everybody. That's a great big misunderstanding in what we're trying to do. We haven't quite got around to that yet!

Melvyn Bragg : Professor Searle, what's your view on conscious machines?

John Searle : Well, in a way I'm a more extreme person on the other side. I think the brain is a machine. So we are conscious machines. I don't have any problem with conscious machines, and I don't indeed, see any difficulty in principle, in building an artificial machine that was conscious. My debate with people on the other side, such as it is, has been about the relation of consciousness and computation, and if you define computation, as it is traditionally defined as a set of manipulations of zeroes and ones, then I think that it's obvious that, that by itself isn't sufficient to guarantee consciousness.

So I see the prospects of building artificial consciousness as not at all impossible, I mean we're a long way from being able to do it, because we don't know how our own brain does it. But the objection I have is that you could get that, just by producing the right kind of external behaviour. The point is you've got to find out what's going on inside, that produces the behaviour. So my pocket calculators better than I am at addition, but it's not conscious. Not even close.

Melvyn Bragg : So how would you go about building your artificial consciousness machine?

John Searle : Well if I was going to build an artificial consciousness machine, the first thing is to find out how our brain does it, and in fact that's the most exciting area of research in science today, is to try to figure out how that brain works. It's kind of a scandal that we really don't know how the damn thing works! And there's a lot of. . .

Melvyn Bragg : Why do you think we're....you say it's a scandal.....why is it a scandal?

John Searle : Well it's a scandal because the most....I mean it's the most important organ we've got. I mean it's the basis of our life, and we understand all kinds of things about our body. We understand DNA. We're starting to understand how the genotype produces the phenotype. We understand cell growth and division. We understand. . . we're even beginning got understand ageing, I'm sorry to say! We do not understand how this brain causes consciousness. We got what? 3 pounds, a kilogram and a half of this stuff in our skull....

Melvyn Bragg : Porridge, Jonathan Miller calls it.

John Searle : .... the texture of English oatmeal, the texture of English porridge, and it produces not just consciousness, but memory and love and emotion, creativity, poetry, everything. I'd like to know how it does it, how does it work?

Melvyn Bragg : Does this make you feel that you're being rather reticent and diffident, a bit of a failure there, Igor......

Igor Alexander : I know.

Melvyn Bragg : ... your claims seem to be very limited compared with what you're being offered by your opponent! (Igor laughs)

Igor Alexander : Well, as John was speaking, my head was bopping up and down, not because I was falling asleep, but because I agree with him so much. But an interesting question is *why* is it so hard to understand the brain.

Now one reason, is that it's very hard to do experiments on the brain. You can do psychological experiments, you can look at the structure of the brain. You can even do some super scanning of the brain, which is a little bit inaccurate, but it does tell you something. Now this is where I feel that we can make a contribution, by moving some of this behaviour into an artificial domain, where we can tear it apart, and we can have hypotheses, we can say "well maybe the brain works in the way it does", for example, in vision, the vision system is incredibly complex, 48 little boxes in our heads, which at the end of doing their work, do something about giving us visual awareness. Now there are hypotheses about how this might work, but they're hard to check. But when you've got a machine, and you can simulate that system, you can say, "well okay if you do it this way, it does appear to create visual awareness, if you do it another it doesn't". So I'm very happy to feel that I'm contributing a little bit to what John's calling for, in terms of getting a better understanding of the brain.

John Searle : Well, I'm sorry to say that Igor and I don't seem to be disagreeing about anything. In fact Igor defends a position I've called weak or cautious artificial intelligence, where you think of the simulation of cognitive capacities as like the simulation of anything else. I mean we use the computer to study digestion, or to do studies of weather patterns, and nobody thinks "Well if we do a computer simulation of a rainstorm, we are all going to get wet", and similarly a computer simulation of consciousness isn't thereby conscious, but it's a very useful device for *studying* consciousness.

So, I would hesitate to call it "artificial consciousness", but I like you other word, it's the "simulation" of consciousness, and that is immensely useful. So you can use the computer to study how the brain produces consciousness, as you use the computer to study how the digestive tract produces digestion.

The mistake that a lot of people make, and I'm pleased to see Igor doesn't make it, is to think that somehow or other, the simulation is the real thing, and what I've always argued is the simulation of consciousness stands to real consciousness, the way that that simulation of say digestion, stands to real digestion. You get a perfect computer program, to simulate digestion, that doesn't mean you can then stuff a pizza into the computer and it will digest it. No, it's just a picture, a model, and that's what I think we get with the brain.

Melvyn Bragg : D'you think......? Sorry after you Igor.

Igor Alexander : Well, I'd like to ask John a question, because you do say, John, very clearly that you feel that there could be some form of artificial object, I don't know, beer cans or whatever, that could, under certain conditions become conscious. I believe that, part of that track can be through virtual mechanisms, which in a virtual way, do become conscious. The reason for that is that if consciousness wasn't the subject of my simulations, if I couldn't tell "well this thing becomes more aware than that", it's artificial awareness, it's artificial. But at what point do we start calling that sort of stuff, "Oh yeah, it's some sort of artificial consciousness"? It's not *living* consciousness, and that's the mistake I think people make, that they confuse living things, with non-living things, that's the big distinction. But in the non-living domain, at what point would you accept that something's become conscious?

John Searle : Yeah, okay, now I think we may be getting a genuine disagreement, I'm pleased to see that! The point is, we know the brain does it. The brain is a biological mechanism, now if we're going to do the *same thing*, using a non-biological mechanism, we know it follows logically, we have to duplicate the brains causal power to get us over the threshold of consciousness. Just as to build an artificial heart you don't have to have muscle tissue, but you've got to duplicate the hearts power to pump blood. Now similarly if we're going to do it with a brain, you've got to find out how the brain does it, and duplicate, not merely simulate that.
So suppose that, as some current theories say, it's a specific set of biochemical and electrochemical mechanisms, using the architecture of thalamocortical system, that's a big hunk of tissue in the middle of the brain. Then what you do is find out what the electrochemical features of that are, and you then recreate those. You don't computer model it. Zeroes and ones aren't going to be enough to do it. You've got to get the actual.....you've got to get the actual causal power of the machinery, and that's what I would look for. Now **how would you test it?** Well, the best way to tell whether or not a system is conscious is to be that system, but it's hard to do that with any system other than yourself. But for example, if parts of your brain were decaying and they were replaced with this artificial mechanism, and you kept your same consciousness, that would be pretty damn conclusive, you'd be pretty confident that that mechanism was producing, was causing, and not just simulating consciousness.

Melvyn Bragg : D'you think that is at all practical or possible, what's just been proposed by John Searle?

Igor Alexander : Well, I'm not exactly clear what John's proposing. Because in some ways it comes through like "you have to understand the biological first, before you can do your simulations". I actually think that the two go together. That the sort of systematic character of what it is that causes consciousness is expressible in several ways. The biological way is obviously one, but at the end of the day it's a very complex system, and it's understanding the essence of that system that's important, and I think the way you do that is by doing it on both sides, by doing some simulation and doing some brain work. But if you're stuck with brain work then you're limited by experimental techniques and things. But because the subject of the artificial becomes consciousness, you have at some point got to say "okay here we do have something which gives me clues about consciousness, which is not biological", and it might give us some clues about how the biological works. So the only difference between us is the perspective from which we arrive at the problem, I think it's plausible and I think it's possible to do.

John Searle : I want to emphasise though, the word you used at the beginning, the notion of simulation, and I don't think you have to understand the brain in order to do a simulation. I think the simulation may well be, as you pointed out, a device for helping us to understand how the brain works. But Melvyn's question was "how would you produce the real thing?", not just the simulation of it, not just a picture, a model of it, but the real damn thing, and there it seem to me you've got to know how the brain does it, and the simulation will help you to understand how the brain does it, but ultimately, if we're talking about computation, a simulation is all zeroes and ones, its a Turing Machine . It's a series of algorithms.

Melvyn Bragg : But d'you think you are going to get to a stage where you can have a machine which has, let's take a very simple....well not a simple idea, but let's take something that people can associate with instantly, can have freewill?

Igor Alexander : Let's just take that and surround it by the other things it would have to have, besides freewill, because defining freewill, I think is very difficult, and I think sometimes what people call freewill is just arbitrary thinking. But there are a lot of other things which are needed for us to believe in freewill. If I have a machine that is aware of it's existing surroundings, knows it's own position within that world, is able to shut off it's senses and think about these surroundings, plan what it's going to do next. Use, in some sort of very language system that we haven't got yet, use the word "I" in a convincing kind of way. I think you'd be able to stand away from this thing, and say

"Well, let's try and get interested in how it's consciousness works". Now you ask this object "have you got freewill?", and it said "yeah I have do you want me to show you that I've got freewill? I'm thinking of a monkey in a tree, and I did that totally freely, nobody made me do that". So there are a lot of issues around the concept of freewill which can be dealt with in a very facile way, but there are a lot of other issues that aren't all that easy!

Melvyn Bragg : This is part of a larger question is when you've simulated or duplicated the brain, d'you have something which is the brain? Or have we created something other than?

John Searle : Well you could create something "other than" in the sense of using a different causal mechanism, as you can create a diesel engine that'll power your car as well as a petrol engine. But what we're interested in is not that it does just an imitation of it, but that it actually produces the same causal effects.
Now you're asking about freewill and that's a hard one, because we don't know how we get our own freewill and we don't know if freewill is maybe an illusion, because the brain looks like a biological mechanism, like any other. *Freewill* is a feature of a certain kind of human consciousness, and in order to produce that in a machine, we've got to first understand how we could produce consciousness at all, and then the hard part, would be, after you've done that, as if that isn't hard enough. The next task would be to produce the kind of consciousness that human beings have when they're making free decisions, when they are engaged in some sort of free decision making behaviour, and that's a tough one. We're a long way from being able to do that.

Melvyn Bragg : D'you think it is....d'you think that it will. . . you will sometime. . . at sometime or other, given a following wind and enough time and enough intelligence brought to bear on it and as Newton said "thinking on it continually", d'you think you will be able to discover in the brain, direct relationships between parts of our brain, things inside our skull, and say imagination? Do you think there's bound to be a direct connection there? Or do you think that at a certain stage, the mass that is the brain turns into something else which becomes the mind, and becomes consciousness which is so elusive and so varied....... ?

John Searle : I don't think it turns into something else. Now one way to hear you question is to ask well "will we be able to find localisations in the brain of specific capacities, like imagination?", and to some extent we already have localisation. I mean we understand the visual system, and we understand the auditory system, and we understand the lang....to some extent, the language system, tends to be on the left side of the brain in most people. So we will get some sorts of localisations, but whether or not we'll get precise localisations, we'll be able to say "right there is where you're imagining your seaside holiday", that seems to me unlikely, but still it's a factual point for investigation.

Melvyn Bragg : Well it's not just that's why you see your seaside holiday, it's more Francis Crick's astonishing hypothesis, that these particular interactions between. . . these firings between these particular specific neurones lead to that specific sort of imaginative thought, I mean that is the sort of, almost unimaginable detail. Is that at all on the agenda as being able.....?

John Searle : It seems to me it's got to work something like that, because we know that's the only machinery you've got to do it. You see what seems so bizarre about it is, you look at this plumbing, I mean here it is, is this plumbing apparatus, and you think, "that's what does the thinking?" and the answer is "yes, that's what does the thinking", and we want to know exactly how does it work?

Melvyn Bragg : These sort of intestinal coils of porridge?

John Searle : You've got.....I mean a point about it is, you pointed out how hard it is to study, the reason it's so hard to study is, you've got a hundred billion neurones, that's a lot of neurones, and then the real action that's taking place at the synapse, and there you've got zillions. I mean some neurones have as many as a hundred thousand synapses on one neurone, and there are days when I'm glad I'm not a neurobiologist! (Mel laughs) Because it's very hard to study that, and they have this quaint vocabulary, they say "we do not have techniques that are not invasive", that means you've got to destroy the brain to study it, or you've got to stab the poor animal, in order to get at the brain. Now we've got these imaging techniques and that's pretty good. The CAT scans, the PET scans, the MRI devices. But we're a long way from being able to get in there and figure out how it works in detail.

Melvyn Bragg : Given the size of the numbers, and then the further complexity of the interaction of these numbers, so you're getting into numbers which are (laughs) bigger than the size of the universe, as it were, what advances.....what can machines, such as those that you're working with achieve? How far along the road can they possibly get?

Igor Alexander : Numbers don't worry me all that much, it's principles. I think if you can discover principles with simulations, indeed, which are much smaller than the sort of numbers we have in our brains. Then you can stand back in awe of the brain if you can discover some interesting things that happen in smaller numbers. Now I think actually, we're further down the road than John is implying, in trying to understand that link, and I'm not talking the biological brain, and the big question always is, let me just touch the tip of my finger with these glasses, it's not television so people have to imagine that I'm doing that. Now all that happens in my brain is that something in my somatasensory cortex, which is somewhere where people wear their earphones, is firing away, the neurones are active. Now that is not the *only* thing that's going on in my brain. That's the localised thing that's going on, and the question is, "why don't I feel these neurones in my head, why does it feel like something's happening in my finger?". The reason is that there are a whole lot of other things that are going on in the brain which have to do with the things that my muscles are doing, the things that my eyes are doing, the things that I'm recognising with my eyes, and all that brain activity is coordinated into one item of consciousness which says that what's happening is out there and not in my head, and if you take that further to vision, and other ways of. . . other sensory effects, we're beginning to understand this very strange thing that we feel there's a coherent world out there, which is totally distorted in the way it's represented in our heads, but all these representations come together and tell us there's a world out there.

John Searle : Well, there's an evolutionary explanation, and that is of course, unless the organism has some way to get information about the world out there, it's not going to be able to cope, and not going to be able to survive. But the. . . and the marvellous thing about consciousness, the absolutely wonderful thing is, we have all this diverse stimuli, and we put all together in a single unified conscious field, so the perception of other people, the taste of the water that I'm drinking, the feeling of the clothes on my back, those are all part of a single unified conscious experience. Now the problem is, and we really don't know how to solve this problem yet, but there's a lot of good people working on it, how do you get from the neurone firings to the conscious feeling? How does the brain get over the hump? And a lot of people think "well we'll never understand that", but the point they forget is we know it happens. We know that the brain does it, and if we know the brain does it, we ought to be able to figure out "how does it do it?" and that's the most exciting problem in science today.

Melvyn Bragg : I flagged, at the introduction, this Chinese Room Theory, we've gone slightly away from that, but I know the number of calls and letters I'll get, if we don't answer this, because one thing the British audience want, you keep your promises on a programme like this, so we're going to for a few minutes talk about the Chinese Room Theory! (Mel & John snigger)

Will you explain your Chinese Room....

John Searle : I'd be happy to.

Melvyn Bragg : ....because it is very famous in the field of artificial intelligence.

John Searle : Well, it's a certain irony that you should ask me to do that, because I first did it on the airwaves, right in this building, when I did Reith Lectures , I did it for the BBC. It's a very simple argument, and it's only directed at people who that think that by designing a computer program, you thereby. . . the right computer program, you thereby guarantee a thinking machine.

Now what I imagine is, take some cognitive capacity that I don't have. I don't speak Chinese. So you just imagine that I am locked in a room full of boxes of Chinese symbols an I have a rule book called a computer program, and I get symbols in, in the form of questions, I look up what I'm supposed to do in the rule book, I follow the program and I give symbols back in the form of answers. Now, on the outside of the room, anybody would say, "well that guy understands Chinese, because look we ask him the questions and he gives the answers". But the point is, and I don't have any doubt about this, I don't understand a word of Chinese -and this is the point of the parable - since computer program, no computer implementing a program has anything that I don't have, then if I don't understand Chinese on the basis of implementing a program, neither does any other digital computer on that basis. So the idea that you can produce the understanding just by running a computer program, that that's enough, that's wrong.

Melvyn Bragg : Igor?

Igor Alexander : Well I think John did the world an enormous service by putting this argument together.

John Searle : Would you tell your colleagues in artificial intelligence that! (laughter)

Igor Alexander : Well you see they don't actually accept within the fraternity (laughter), so that's why I say what I say. But people are making outrageous claims about how a computer could understand a story, and it was quite obvious that if you asked it the right question, it would tell you that the question's invalid, and even now when you ring up a large corporation and you get this machine that tells you "If you want sales press 1, and if you want the director press 6" tells you exactly how far computers have got with language understanding. So you know, I'm very happy with what John did there, and in fact it did underpin some of the directions took after that, and that is actually to actually discover where this "aboutness", where this sort of rich representation that you need to get in a system, which is partly achieved by the evolution of the structure of the system, and partly achieved by how this thing organises its experience. That's where that kind of artificial intelligence came from, and it was due to John.

Melvyn Bragg : But it hasn't stopped people predicting and fantasising about what thinking machines will be doing in the quite near future. Do you think these are harmless fantasies, or d'you think that they're rather....they're a bit of a nuisa....even a worry?

John Searle : I think they do a lot of damage, and I'll tell you why. People get the illusion that all we need is better technology, and what....the beauty of the Chinese Room is it showed that technology doesn't matter. We're talking about the *definition* of computation, in terms of symbol manipulation. So the danger. . . the two dangers were, we'll get a technological fix for all our problems, and all that really matters is behaviour. If you've got the machine that *behaves* as if it understood Chinese, then that's all that really counts, and both of those are serious mistakes.

Igor Alexander : Yes I think technology has got to change a lot. It's not going to be the symbol manipulation technology of 50 years of AI that we've had. Technology that gets a little bit closer to, hopefully what we understand about the brain in the future is much more likely to lead us towards more helpful machines. Machines which are heavily artificially a tiny little conscious. But there will be more competent machines than the ones we have around at the moment.

Melvyn Bragg : Finally, and reasonably briefly, because we're coming to the end unfortunately, what do you think the use of artificial brain machines are?

John Searle : Well there are lots of uses right now. I mean what I call weak AI is good for oil exploration, medical diagnoses, weather prediction. So just using AI programs for practical purposes seems to me perfectly legitimate.

Melvyn Bragg : Igor?

Igor Alexander : Yes I think making driving safer, making aeroplanes safer, getting us better telephones systems when you phone large corporations! Those are the kind of things we might converge on in the future.

Melvyn Bragg : Why do you....finally, why do you think people are driven by this fantasy that they can be us. That we can create....

John Searle : Will of power. I mean what I discovered in AI, and Igor knows more about this than I do, a lot of these guys want to play God. They think....I mean as one famous AI person told me, " I am creating minds", you sit in front of your console, you sit in front of your monitor, and you type in programs and list and you're God. You're creating a mind. I think it's will to power, and it's exciting, I must say, I think it would be exciting.

Igor Alexander : Yes....(Mel laughs). . I don't have ambitions of divinity. . . but er.....

Melvyn Bragg : Such divine ambitions!

Igor Alexander : I'd quite like to be Walt Disney (Mel & John laugh), but in that process, understand the brain a little bit better.

John Searle :   He gets paid better than God!

Igor Alexander : Yeah!

Melvyn Bragg : Thank you very much Igor Alexander   and   John Searle, next week I'll be joined by Ian Stewart and Brian Butterworth to discuss the beauty of mathematics and thank you for listening.