



Debate and Perspectives

'Computer models of the mind are invalid'

Ray Tallis¹, Igor Aleksander²

¹5 Valley Road, Bramhall, Stockport, Cheshire, UK;

²Imperial College of Science, Tech. and Medicine, Department of Electrical and Electronic Engineering, London, UK

Correspondence:

R Tallis, 5 Valley Road, Bramhall, Stockport, Cheshire SK7 2NH, UK.

Tel: +44 7801 834230;

E-Mail: raymond@rtallis.wanadoo.co.uk

Journal of Information Technology (2008) 23, 55–62. doi:10.1057/palgrave.jit.2000128

Proposer: Professor Ray Tallis

It is a great pleasure to have this joust with Igor who is not only a brilliant thinker about the mind, but also a great intellectual sparring partner. Igor has expressed his dissent from the view that I'm going to advance in support of the motion which is that 'computer models of the mind are invalid',¹ in his book: *The World in My Mind, My Mind In The World: Key Mechanisms of Consciousness in Humans, Animals and Machines*. He actually devotes five pages of the final chapter to what he calls 'Tallis's complaint' which I am now going to make public. The point of issue is whether computer models of the mind are valid. I am going to argue that they are not because the computational theory of mind is invalid. Igor may go on to argue that even if the computational theory of mind is invalid it may still be useful in the sense of being fruitful. I will leave him to prove that and simply suggest to you that while barking up the wrong tree may give a sense of progress it is illusory.

So what is the computational theory of mind? It is that the mind is to the brain as the software is to the hardware, or to put it slightly differently the mind is a set of computer programs implemented in the wetware of the brain. It is difficult to recall that this notion was once revolutionary and counterintuitive, because it now seems to be hardwired into the thinking of most people who have views on the nature of consciousness. And so in arguing against the conventional unwisdom one has not only to demonstrate that minds are not computers, but also to explain why such a daft idea should seem halfway plausible. Arguments against the computer theory of mind therefore, should include an explanation of the aetiology and persistence of the delusion. My case has four strands which are addressed in Igor's book, but I will restate them naively, as it were, to bring everyone up to speed on our dispute.

The first is the unremarkable claim that computers are not conscious. The second is the equally unremarkable claim that consciousness is not computational. The third will assert that minds are conscious and if consciousness is not computational then neither are minds. And finally, if there is time, I will examine the language by which those who believe in the computational theory of mind are self-deceived into thinking that the gap between body and mind, or between neural activity and consciousness, is crossed.

Let me begin with a something rather uncontentious – computers are not conscious. Very few people, even

proponents of the computational theory of mind believe that present-day computers are conscious, in the sense of being aware of themselves and of the world around them, or of being capable of happiness or despair. Super Crays with gigabytes of RAM are as zomboid as pocket calculators.

Most computational theorists are hesitant in accusing current computers of being conscious. The commoner claim is that while we do not at present have conscious computers, we soon shall have, or eventually, develop them. We should be suspicious of such claims. I am old enough to remember the predictions of people like Marvin Minsky, who claimed that by 1990, computers would be so sophisticated that they would be able to manage without us and would treat us rather in the way that we treat household pets.

We should be sceptical for another reason. Those who claim that one day we shall have conscious computers are not able to specify what additional features conscious computers of the future must have. There is much hand-waving: conscious computers will be more 'complex' or have a different kind of architecture – for example, a parallel architecture, based on so-called neural networks. I have yet to see a definition of complexity that will make consciousness seem inevitable, or even more likely, in the artefact possessing it. And there is no evidence that currently available, massively parallel, computers are more self-aware, are more prone to suffering or joy, or able to experience the sound of music or the smell of grass than their serial counterparts.

The intuitions guiding those who imagine the conscious computers of the future are often a little naïve. They include the notion that feedback loops or 're-entry' will awaken circuits to self-consciousness. The smartest artefacts, with the most subtle feedback mechanisms and self-monitoring, are unaware of their smartness, or of being a self or of their monitoring. The most striking testimony to the fact that very few people really believe that computers will one day be conscious is the way that the goalposts recede as computing power advances. Computers a million times more powerful than those available to Alan Turing when he put forward his famous test for machine intelligence are still not credited with the ability to truly think.

Turing argued that the question of whether a machine could think should be replaced by the question, 'are there imaginable digital computers which would do well in the imitation game?' In the imitation game, the computer

behind the screen responded to questions posed to it by a human subject. If the computer could fool someone into believing that it was a human being, good enough – it was truly a thinking machine. This outmoded behaviourist definition of thinking is still invoked as a test for thoughtful computers, though many computer games could pass the Turing test with flying colours, without anyone, not even computational theorists of the mind, being prepared to grant them the power of conscious thought. And this reminds us that *simulating* the behaviour of a thinking conscious being is not the same as *being* a thinking conscious being. What is more, simulation would count as simulation only to a conscious being who may or may not be deceived. The notion of simulation, in other words, presupposes judges whose consciousness is not simulated. And that is why the concept of the zombie would not arise in a world populated only by zombies.

Let me now come to the second strand of my argument against the computational theory of mind; consciousness is not computational. It is surprising how rarely those who advocate the computational theory of mind think about what actually goes on in a computer – the passage of small electrical currents through vast numbers of microscopic circuits. Now these events count as information processing, as symbol handling, as behavioural control, etc., only when the computer is being used to support and extend the capabilities of conscious human beings. The latter have to provide the consciousness. The computers are no more information handlers in their own right than a walking stick is something that walks on its own, or a clock by itself is something that tells the time. This point was made by John Searle in one of the most cited and argued over papers in philosophy in the last 30 years in which he put forward the famous 'Chinese Room' thought experiment. He imagined someone sitting in a room, receiving an input of Chinese symbols. The individual was totally ignorant of Chinese and understood nothing of the symbols. He was, however, in possession of the rules for processing these symbols and as a result of which he produced an appropriate output to a certain input.

Let us suppose, said Searle, that the input symbols amounted to questions and the output symbols amounted to answers. It would appear that the person in the room was answering the questions. The individual however, did not understand a word of what was going on. Searle used this as a compelling analogy of what goes on in a computer that links inputs with appropriate outputs. The events in computers do not amount to genuine understanding. Indeed, given that symbols *are* symbols only to someone who understands them, the events in computers, considered in isolation from human beings do not really process symbols. It is merely the passage of minute electric currents along circuits that may or may not cause other unconscious events to happen, such as the lighting up of screen in a certain pattern. This becomes symbol processing, becomes conscious understanding, only when the computer is serving a conscious user.

Digital computers primarily compute. For example, they add up two and two to make four. If consciousness were made up of a huge mass of such calculations, which aspect of the calculations would be conscious? The input? The output? The process? The addition sign? You may be

uncomfortable with my example, a single short calculation, but there is no reason why a long calculation should be more conscious than a shorter one. Or a multitude of calculations should be more conscious than one on its own. The same objections may be directed at those such as Patricia Churchland, who in her computational theory of mind, asserts that consciousness is not so much number crunching, as logic crunching, or mathematical operations such as vector-to-vector transformation. Patricia Churchland's claim that the mind is a kind of logic machine operating on sentences, begs so many questions, leaving aside the difficulty of imputing logic and sentences to electrical impulses in isolated computers, or come to that, isolated brains. Only magic thinking could enable one to believe that logic and sentences could exist despite being unrealised in, and unrealised by, conscious human beings. What is more, sentences and logic are highly abstract and hardly correspond to contents of consciousness, such as a feeling of warmth.

It is much the same with calculations. In one of the classic accounts of computational theory of mind, Philip Johnson-Laird argued that sense experience actually amounted to doing sums. Vision, he said, is rather like hundreds of problems of finding the values of x in an equation like $5 = x + y$. This is silly, but at least honest. The notion that something as basic as vision, enjoyed by beasts even less sophisticated than gawping humans, is made up of something that looks rather sophisticated, namely calculation, exemplifies one of the most consistent and odd features of computational theories of mind. It is that of inverting the commonsense hierarchy, in which we place qualia – uncategorised sensations – at the bottom and clever operations such as calculating at the top.

The computational theory of mind cannot deal with the globality of ordinary awareness, our openness to an unrestricted domain of events. Nor can it model the unity of the moment-to-moment field of consciousness, permeated by the extended future and past, whose numerous components nonetheless retain their distinctiveness. Every attempt to explain integration – 'merging without mashing' – has so far failed. One cannot make millions of computational events add up to a unified totality while at some level keeping them separate. You cannot have your rain shower as separate drops and as a pool. You can seem to be able to do, so by imagining several viewpoints within the computer or the nervous system, but if the viewpoints are themselves made up of activity within the computer, they are not in a position to view anything, never mind view separately the totality and the individuality of components, as we do. A swarm of nerve impulses, however disciplined by biological wiring, cannot generate a view upon itself.

For some philosophers and neuroscientists, my earlier argument that computers are not, or could not be, conscious is irrelevant to the computational theory of mind, because, so they argue, consciousness is one of the least important aspects of the mind. Many cite Lashley's argument put forward in his famous 50-year-old paper *Cerebral Organisation and Behaviour*. Lashley says there, 'no activity of the mind is ever conscious'. Actually that is to misrepresent Lashley. He went on to explain that he meant, 'experience gives no clue as to the means by which consciousness is organised'. This latter, is of course not



surprising. If I had to work out, from scratch how to make words sound in my head and how to organise those words to form a sentence, I could not 'think to myself'. Every conscious activity is, of course, underpinned by unconscious mechanisms, but this does not mean that conscious activity *is* unconscious mechanisms. The emphasis on mind as mechanism, as a coalition of unconscious machines, has enabled many philosophers to focus on the causal links between inputs and outputs and marginalise, or squeeze out, intervening conscious contents. This is the essence, as many of you will know, of functionalism, the theory that we have to understand what we hitherto thought of as mental contents, as contents of consciousness, simply as being the many links ensuring the right causal connections between input and activity. Functionalists, for example, most of whom subscribe to the computational theory of mind, see the mind simply as a complex of causal way-stations, between experience and behaviour. Functionalism, which is now largely defunct – though it has a posthumous life in the some of the writings of Daniel Dennett – was the last gasp of behaviourism.

Pure functionalism has become even more unfashionable since David Chalmers made the obvious point (overlooked only by certain philosophers in the grip of theories) that the conscious mind has phenomenal features as well as causal features. At the phenomenal level, the mind is characterised by the way it feels. At the causal level (which Chalmers calls the 'psychological' level), the mind is characterised by what it does, and feelings are irrelevant. Even so functionalists such as Dennett have mocked the emphasis on qualia, the actual experiences of things, and on propositional attitudes such as beliefs, as mere hangovers from a pre-scientific mode of thought. This makes it easier to assimilate minds to devices such as the computers.

Pretending that consciousness is unimportant is not an option for neuro-philosophers who are serious about their business. As has often been pointed out, you cannot be mistaken that the contents of your consciousness are in themselves real. If those contents are conceded to be real, but causally ineffective, being merely epi-phenomena after the computational fact, this makes them very difficult to account for. Why do they arise at all? Besides, even if they were unable to bring anything about, this would not make them any less central to mind or any less important to our lives. The difference between the sensations of an orgasm and a toothache would still be central to human mind-ness even if, implausibly, it had no role in behaviour.

And so I come to the fourth strand of my argument. This takes me into deep waters because this is where I look at the apparent attractiveness of the computational theory. Connected with this is the massive fudge on which most neural theories of mind, of which the computational theory is one, depend. I want to talk briefly about the language of the cognitive sciences, because it is the language which makes computational theory so difficult to escape from or to think outside of. This language has enabled philosophers and others, to ascribe to computers a stand-alone ability to do things that are actually performed by human beings with their help. We talk about computers 'doing calculations', 'guiding missiles' and 'detecting signals'. While much that is useful may go on in a computer, in the absence of conscious human beings, the electronic events count as

useful, indeed meaningful, only when conscious humans return to the scene.

Similar talk pervades the discussion of brain activity. Nerves are credited with 'doing calculations', 'detecting signals', etc. Indeed, much current neuroscience would be unthinkable without using terms borrowed from anthropomorphising computer talk. Attributes are transferred from humans to artefacts such as computers and then they are transferred to the brain or the mind in order to justify a computerised account of either or both. Neural circuits, according to the computational theory of mind, are haunted by homunculi which are made respectable by their time spent in computers. Use of the term 'information', in describing the activity of computers, brains and minds, has played a central role in supporting the computational theory of mind.

Many neuroscientists describe what bombards the nervous system as 'information', for example light, as visual information and what happens in the nervous system as 'information processing'. Since 'information processing' sounds mind-like, the mind/brain barrier is effortlessly crossed and no work has to be done. In fact, what we have is an elaborate play on two quite different senses of the word 'information'. The technical sense, used by information engineers, in which it is a measure of the load carried by devices such as telephones and which has nothing to do with meaning, significance, or consciousness, and the ordinary sense which refers to something that passes between one human being and another, with or without the mediation of the artefacts and which has everything to do with meaning and significant consciousness.

This, in summary, is my argument. First computers are not conscious and the fact that the entry criteria for the category of conscious being gets more stringent as computers get more sophisticated shows that they are unlikely to become so in the near future. Secondly, there is little or nothing of the conscious mind that is computational in any clear sense of the word. Thirdly, belief in the computational theory of mind is often based upon a self-refuting denial of the centrality of consciousness in the mind. Fourthly, the little plausibility that the computational theory of mind has depends on the sloppy use of words, with terms passing promiscuously between people, computers, minds and bits of brain. I put it to you that, if we cannot think of the mind as a significant computational entity, then modelling the mind on computers is invalid.

Opposer: Professor Igor Aleksander

My great problem with this discussion is that I so enjoy listening to Raymond because I agree with many things that he says. In fact, he talks of computational models of the mind in terms of the last 50 years of artificial intelligence and the various kinds of discussions that this has led to. I'm going to spend the next 20 min standing back from that approach a little and suggest perhaps, that this is not the whole picture. There are elements of computational discussions that have fed directly into the philosophical contributions to discourse about the mind and indeed have been such as to be not seen as invalid, but that they have already had an effect and might have an effect in the future. Certainly no one would disagree with the notion that

computers are not conscious. But there is a problem with words and the problem with words lies in thinking about computers. I'm going to present three ways in which one might think about computers which is a very short list from a much longer possible list.

The first thing that comes to mind when we think of a computer is that thing that sits on our desks. Because I am a little old I think of the computers that used to take up a whole room, with an additional room needed to ventilate the room that the computer sat in. Such machines would take about an hour to add up columns of figures for financial transactions. That is one kind of model of a computer, which does seem to come through sometimes when we argue that it is not possible to imagine that such a thing could be conscious. I agree totally. It is *impossible* to think that such a thing could be conscious, but there are other ways in which computers have influenced our thinking. The second approach that I will be talking about has to do with virtual machines and that is the situation where one simulates things on a computer and the structure of the computer, whatever the computer does, disappears and in fact it is desirable for it not to influence what it is that is being computed.

The third facet is the development of computational theory which is so different from physics chemistry and biology, that it has been taken on by those who want to talk about mind. So those are the three points that I would now like to expand.

There is one fact that the clumsy computer on your desk teaches us that is very much to do with Ray's fourth point. He has not actually used the word which is so beautifully used in his books, which is the word 'neuro-mythology'. I think that is what his fourth point was referring to and that is the strange idea that by looking at the function of neurons, and knowing everything you need to know about the function of neurons, you are going to be able to infer the presence of a mental state of a mind in the object. Now exactly the same thing happens in computers and we could call it 'transistor mythology'. No self-respecting computer scientist would attempt to look at the function of the transistors – such people know all about transistors or what they are doing at one particular point in the operation of a computer – to infer whether that computer is calculating the age of the universe or just downloading some pornography. It would be impossible to tell the difference between two. So that is not something that computer scientists try to do. What does this teach us? It tells us something about levels of description of complex machinery. The brain is the most complex machine on earth. Computers are nowhere near this. We are totally convinced of that, but folk working with computers find them complex enough to say that you cannot describe what goes on inside a computer just by looking, say at what happens at the transistor level. So they develop a way of talking about what happens at the transistor level, what sort of low-level programs control that, and where higher level programmes come in and so on. So the concept of a descriptonal level, an appropriate level of description, comes into our thinking through the existence of computers.

At this point I am reminded of a rather excellent paper that was written by Aaron Sloman, a philosopher and computer scientist at Birmingham University, and Ron

Chrisley who is in the audience. They suggested that in order to have some interesting discussions about mind, one can create a model which not only has levels, one on top of the other, which go from the lowest operating levels to management of that process, but also have layers going from input to behaviour, with the important part in the middle. That was published in the *Journal of Consciousness Studies* in 2003 in a special issue edited by Owen Holland who is also listening. Something we have learnt from the complexity of computers is the question of how we describe complex systems, of which the brain is one, and how it creates mind. The second point is that of a virtual machine. Let me give you a simple example. Most of you who work with computers can press a button on the computer and a calculator comes up on screen. When that happens you can forget about whether it is an IBM computer or whatever, because you can use that calculator as a calculator. It is a machine that is running on a host machine, the operation of which is so general that you do not know what it does.

It is possible to apply this idea of virtuality perhaps to running virtual examples of hypotheses that one has about how the brain works on this unimportant substrate. What is important is the simulation; it is the machine that one is trying to manipulate to see if it does 'that' when you do 'this'. Is 'that' like the brain and so on? Many friends and colleagues have told me that is a totally invalid process because you can simulate a hurricane on a computer and it doesn't destroy your computer or your office. So it is very different from a real hurricane, but that hurricane that is simulated on the computer is 15 miles down the road at the moment and the computer is going to tell me whether I have to run or not. I think that the same principle can be employed when running virtual machines that somehow ask questions of the brain and check hypotheses. The hypotheses may be incorrect, but here we have something: a material on which we can operate and do both science and possibly even philosophy, without worrying too much about what the computer does or how it is made. In other words, virtual machines have a reality which is usefully disconnected from physical makeup which allows a computational 'consciousness' to be discussed without falling into the trap of physicalism.

Perhaps I could give you some more examples of how one uses this virtual machine idea. A while ago I was listening to Max Velmans who has written a beautiful book entitled *Understanding Consciousness*. He is very much of the opinion that machine modelling of the mind is pretty much invalid with respect to the really powerful arguments that one can develop through just thinking and philosophy. He is a psychologist-philosopher and one of the things that he is arguing about is that people who model the mind try to find ways in which the internal representations are somehow to do with the external reality. He points out that what we sense is sometimes very different from the reality of the world out there. One of the examples that he uses is a Necker cube, which is a wire-frame drawing of a cube in which the parallel edges of the cube are drawn as parallel lines on the paper. When two lines cross, the picture does not show which is in front and which is behind. This makes the picture ambiguous; it can be interpreted two different ways. When a person stares at the picture, it will often seem to flip back and forth between the two valid interpretations.



By means of a computer model I have tried to show that there is some process in the brain which concerns the interaction between the dorsal path of vision, which is unconscious and is known to be unconscious, and the ventral path which is known to be the standard visual pathway, where the two interact and the first actually makes you grab things and the second makes to see things. The mechanism that is trying to grab something is trying to grab an impossible object and that has feedback on to the visual system causing the reversal which seems so strange. That hypothesis may be completely wrong, but at least it gives us an entry into a discussion about what may be going on by showing that it can actually happen in a virtual model.

The third point is the theory. I have just been looking at a wonderful book called *Automata Studies*, edited by Claude Shannon and John McCarthy that came out in 1956. That was early days in the history of computers, but they started by saying that the theory, unleashed by us having to think about how computers work, would be absolutely essential in future discussions about how the mind related to the body. They didn't actually prove that hypothesis, but there were three parts of the theory; the first had to do with neural networks, which was that neurons in various configurations have behaviours that were quite revealing in those days. They are not obvious in neurology because neurology tends to describe structures in the way that they look physically, where as these two people started to say, 'when something looks like that what does it do?' 'What are its internal states like? What is its internal state structure like?'

The second element that they addressed was to do with computation. They asked what could be computed by programming and what could not. This followed on the work of Turing, not his imitation game, but his work on assessing fundamental computational limits.

The third element, which I found absolutely fascinating, was that it was the first theory that dealt with objects with internal states. Up to that time everybody tended to think there was an input and output and maybe something happened in between, but it was not very important. It was the first time that people started writing down mathematical equations as to how the internal part of these systems could relate to what was happening on the input and how they might act on the world. I think that while this has been enormously influential, it is not addressed by those who attack the computational method.

So to close, I would suggest that, yes, we have had some daft theories out of computer science, but I think there is enough there for me to feel that the advent of the computer has changed the way that we think about complex mechanisms which include the brain. Therefore I think it is false to say that our computational models of mind are invalid. They have stimulated both computer science and some interesting discussions in philosophy.

Discussion

Richard: I think the question of whether a computer can be conscious is a very serious one. If you could create a virtual machine that was effectively a reproduction of the functioning of the brain with all its neurons and so on, you would think that it was really in with a chance of being conscious. I am not saying that it would be, but it would be

in with a chance. So it is a serious question, but I would like to suggest that it is the wrong question. To say, 'here is something in front of me', which happens to be a human being or a computer or a dog and then ask, 'is it conscious?' may actually be the wrong question. The right question could be, 'is it to me conscious?' If I ask the question 'am I, to me, conscious?' The answer would be, 'yes'. 'Are the other people in this room conscious?' 'Yes' and that is not because I have any direct window into their brains, but because we are all the same species and at a more practical level we interact socially so we know each other in that way. But if I ask, 'Is a computer to me conscious?' Well, probably not, because it is not me and I can't say, 'it is like me so it is probably conscious, because it is *not* like me and I don't socially interact with it, unless I'm some kind of ultra geek'. So we shift the question from 'is it conscious?' to 'is it to me conscious?' make it a first person perspective question and maybe that makes the debate looked different.

Ray: You raise lots of points which are embedded in the question, but it seems to me that the idea that if you simulate brain activity you will get some kind of conscious entity is predicated on the notion that the stand-alone brain supports consciousness. I have to tell you that a stand-alone brain without a body is a pretty dismal thing, a body without an environment is even more dismal and an environment without a culture and society is yet more dismal. So, however much you might replicate the states of the brain I think you would be unlikely to get anything close to the consciousness you describe.

Igor: It is sometimes worth asking what an artefact, that is claimed to be conscious, is conscious of.

James: When you say that the mind is one of a series of complex systems, I think it moves the argument towards asserting that if the mind is one of a series of complex systems, therefore it is a bit like one. I think anyone who has read books like *Flatland*, would agree there can be a similarity, yet there is a quantum leap between having a mind state and a computational complex system. My question is, 'what do you think can be achieved by the present efforts, such as those going on at The University of Manchester, which investigate what can be modelled by computational models, such as distributed processing, to understand the ability of the brain to recover language after a stroke'. I think Igor has partially answered that, but is that a use for this kind of computational model or is it something we should be shying away from?

Igor: To take the last point. It is enormously useful to be able to do that. It would be very difficult to do it by writing down equations on a piece of paper or building a model out of toothpicks. This is what computers are for. However, I ought to point out that those sorts of developments are based on a neurological hypothesis as to what might be causing the trouble in the first place. Things do get complex and beyond the reach of simple science. That is where the computer is very useful, because it can simulate a very complex system and allow one to explore it by asking the questions, 'What happens if I do this?' or 'what happens if I do that?'

Ray: One thing I should like to pick up from what Igor has said, and there are a lot of things that I agree with, is that computational models allow what has been described as reverse engineering, which is a way of realising what is required in order to deliver what the human body normally



delivers. I think that is extraordinarily illuminating and why in many ways I find computational models very interesting.

Ian: Did anybody specify the denotation of 'consciousness' at the beginning of the talk? Because if not I would ask what we are talking about or trying to simulate.

Ray: I think if you don't know what consciousness is then you are in trouble.

Ian: That is probably why philosophy has not addressed the question in any sort of effective manner, because no one has bothered to define what is being talked about.

Igor: I sometimes give talks to scientists who get very worried that the first thing I say is not a definition of consciousness. I have to say that the definition of consciousness is the problem and suggest that the modelling work is an attempt to solve this problem.

Ray: I am not sure that I entirely agree with that. If Ian requires consciousness to be explained to him he must be a zombie, in which case we would be wasting our time explaining consciousness to him. Consciousness has many levels, but at the ground floor there are sensations or experiences.

Ian: You mean qualia or secondary qualities?

Ray: Qualia will do nicely. If you do not have qualia then I am wasting my time talking to you.

Ian: So we are talking about qualia?

Igor: Well, that's just a philosopher's word.

Ray: Of course there are many other levels.

Ian: Perhaps I could come back on that later.

Derek: I don't know whether this is a continuation of Richard's reasoning, but can I use the arbitrary term 'normal, healthy, intelligent and balanced human mind'. A person with those characteristics can control the computer rather than the other way around. So the relationship is one of control. Immediately the individual becomes abnormal in that role then the computer may affect an individual with abnormal characteristics and effectively control him or her. I think the notion of control is a way of assessing consciousness.

John: What I have gathered from comments is that the motion should have been 'computer models of the mind are not completely valid' because the speakers have admitted they are valid for quite a large number of activities such as computation itself. Computers do calculations, minds do calculations and so there is a similarity which is not an accident because minds of some kind designed computers to do their work for them, but we are expecting too much if we expect a model of something which is invisible to be completely valid.

Ray: There is a suggestion that in a sense I feel that computational models of the mind are a bit valid. But I do not. They cannot be a bit valid anymore than a person can be a bit pregnant.

John: Do they have to be completely true or completely false?

Ray: Yes, I think there are some interesting points of convergence between computers and minds but not many. We think there are more points of convergence because we are inclined to say things of the kind you have just said – that 'computers do calculations'. They do not. Human beings do calculations with the assistance of computers. Computers are only prostheses; they no more do calculations

than clocks tell the time. Clocks help us to tell the time, but they don't do it by themselves. What is interesting is that computers may be useful heuristic devices so the computational model might be quite good as a discovery device, and this is rather mysterious. How can something that is invalid actually be so useful and so productive of theories that advance our understanding? I think that is a really interesting question. Parmenides produced the rather unthinkable thought that nothing changes. You cannot actually have a thought if nothing changes, because a thought could not come into being and yet his was one of the most potent insights that has ever been and perhaps the foundation of Western thought. So the interesting question is, 'How is it that, even if Igor agrees with me that computer models of the mind are invalid, we both agree that they are incredibly fertile and fruitful?' That is profoundly mysterious and tells us something about human consciousness.

John: Invalid, but fruitful?

Igor: Ray and I decided on the title of the motion in a great hurry via e-mail and I felt when we came in that perhaps it was too harshly worded, because nothing is invalid. So it is too easy a proposition to defeat. I have therefore tried to step back from it a bit and say that the way that some people put the computational theory of mind is invalid, but others make good use of it and therefore it is not invalid.

Sunny: One of the things that always bothers me about these kinds of arguments and it goes back to reading the *Emperor's new mind* is that it seems to me that you are arguing that there is something special about what I would call 'wetware'. I am a physicist by training and I am a physical object, a robot that I might build and claim was conscious would also be a physical object. I am bound by the laws of physics to interact with the world in a certain way and I have a certain level of complexity in my brain because the things that I do are very complex, but fundamentally there is no reason why I should not be able to take that level of complexity and that interaction with the world and build it into a machine and have it do the same thing, unless there is something magic about biology that I am not aware of. I wonder whether you (Ray) are aware of that sort of magic about biology?

Ray: Not at all, and that is why I disassociate myself from the views of John Searle even though he is very hostile to computational theories of mind, because computers can be installed on any number of things, whether it be silicon or human brains. So I do not think there is anything special about wetware, particularly as described by scientists as consisting of neurones and semi-permeable membranes that allow sodium in and potassium out. There is nothing particularly special about that. I think we have a real question here because we do not generally think that we are no different from a pebble. I do not think you feel you are utterly embedded in the causal network.

Sunny: I do believe it is only a difference in complexity.

Ray: Well that is the (wrong) response I wanted because complexity is actually relative to the way you see things. You can describe a pebble as one simple object or as an indescribably complex array of atoms, molecules, or protons, gluons, and leptons. Complexity is always relative to descriptions and not intrinsic to the things being described. There is nothing objective about complexity.

This audience can be seen as a number of people or as a single entity. It could be seen as an audience plus a room or simply as a packed room. Complexity and multiplicity and compositeness are entirely relative to perception and description. They are basically anthropocentric concepts and therefore you cannot insert them or project them into the material world.

Sunny: Not even if you are comparing like with like?

Ray: Not even if you are comparing like with like because the likeness of like is also relative to your observations and descriptions. As Nietzsche pointed out no two leaves are the same.

Igor: There is a problem with words again. The word 'complexity' has several characteristics (or connotations). Used in everyday language it is not all that meaningful, but there were ways in which people use it and (for example) the London School of economics is one of the centres of the European Complexity Programme in which it is used differently. There are people who use the word complexity when emergent properties are involved. These are overall properties that happen as a result of an interaction of a very large number of little elements which is hard to analyse. An example may be the stability of the brain. It is very hard to explain why our brains being full of components can, in natural language, be seen as being very complex. Indeed, I described the brain as the most complex machine on earth. However, it also has a surprising stability which is analysed as a theoretically complex system.

Ian: This is the work of Ross Ashby?

Igor: Ross Ashby and Santa Fe.

Ray: Igor has described the human brain as the most complex object in the world and it is funny that it is human brains that merit that description. It seems to me that if you think about pebbles, there are supervenient properties of pebbles that you could derived from gluons, leptons, protons and so on, so supervenient new properties do not distinguish a pebble from a brain.

Ken: I am really unhappy about this discussion because we seem to be talking about two completely different and incompatible things. I started programming computers over 42 years ago and all I have seen since then is that the boxes have got smaller and have more in them, but what is in those boxes is incredibly simple. There are switches which are either 'on' or 'off'. There is nothing complex about a computer, they are just a lot of little bits. They are useful tools for simulating all sorts of things and I have done a lot of that, but without a human being behind the computer the thing is quite useless and just a pile of junk, whereas the human mind seems to be quite unfathomable. You can talk about how neurons appear to work. I think the cycle speed of the mind is about 238 cycles per minute and the best computer that we have has two processors doing 2000 million cycles per minute and yet it is not in the same league at all. The idea that a pile of junk could replicate the human mind in the foreseeable future seemed quite incredible.

James: I would just like to get back to Sunny's question because I feel that the discussion about complexity is a distraction from that, so to repeat it I would like to ask Professor Tallis, 'what is it that is special about human life?'.

Ray: I think there are two questions: 'what is special about life?', which is partly Sonny's question and 'what is

special about *human* life?' I am not sure I can identify what is special about life, but I have written a great deal in describing what is special about human life. To summarise that in two sentences would be difficult, but essentially it seems to me that human beings are explicit animals, which is to say that they do things that otherwise (just) happen in the universe. They bring about things and they are agents and they are self-conscious. That is what is special and how that came about is interesting, but a long story. There is the question about the difference between living and nonliving things, and interestingly biologists are now getting more and more uneasy about their ability to understand the nature of living things and how they hold together. They are terrified they are going to fall back into vitalism.

James (?): Igor seemed to be talking largely about models of the brain, rather than models of the mind, and it has come down to definitions of consciousness. It could be argued that as long as we are talking about models of brain function we are not addressing the question of computational models of the mind *per se*, assuming that we accept that the mind is not synonymous with the brain by definition. On the other hand I do remember Igor saying a couple of years ago that if computers or machines became conscious they would not be like humans because a conscious machine would be conscious of being a machine just as much as a person is conscious of being a person. I wonder how relevant that is to this debate?

Igor: I will stand by the statement that if an object becomes truly conscious it has to be conscious of what it is and that takes some of the fancy things you will find in the literature out of the game. Concerning your first point: It seems obvious that the point about simulating brains or having virtual models of them is to try to check hypotheses about mind, which to me is the total structure of states that the object develops. Going back to 1956 and automata studies, mind is the state structure of the brain. Getting that right and getting some information from it is my current hundred-year project.

Ray: A hundred years is far too short. It would seem to me that the idea that the mind is the state-structure of the brain would take some unpacking. At the very least, we should ask what is meant by a state-structure and why should the brain have that thing separately and that separate thing become a mind?

Richard: My first question is a proxy question from a lady at the back and I think it is an interesting one. Is the interface between computers and humans so bad that it is responsible for the 500% increase in depression over the last 10 years.

Ray: The reason for the 500% increase in depression over the last number of years may be 500% increase in people's tendency to report that they are depressed.

John (?): I think your notion of validity sets the standard a bit too high to get the conclusion you want. If you have to be either completely valid or not valid at all then it looks like every scientific theory we have ever had has been completely useless because every scientific theory is incorrect to some degree, so we cannot have any notion of progress and I think it is not a very helpful view of validity to say it is either all or nothing. My second point is that I do not see why computers cannot be valid models of mind. If you look at the brain, the very same questions you ask about computers

you could ask about the brain. You could ask how rubbing two neurons together, or millions, or billions, ever produces consciousness. You would be just as baffled and you would not be able to give an answer in that case either. So at most your position should be agnosticism. You should say, 'I don't know whether computer models are valid, we shall have to wait and see, but as it stands now, I cannot see how a computer could be conscious, but I don't see how a brain could be either. We know brains can produce conscious states so maybe it is our ignorance, not something about computation, which is the problem'.

Ray: First of all I do not believe for a moment that stand-alone brains are conscious. But the question of validity is an interesting one and that is why I separate validity from fruitfulness. The trouble with computational theories is that they are invalid in a special way. The phlogiston theory was actually very fruitful although it was wrong. Computational theories of the mind are wrong in a different way. They are conceptually wrong and I find it very interesting how they could be fruitful when they were conceptually wrong. I just find it extraordinarily interesting that most of our cognitive progress has been driven by wrong theories. I have no problem with wrong theories en route, but I am very puzzled as to how they could drive such cognitive progress. Consider Parmenides. He actually lies behind most of modern science: the theory of conservation of matter, theory of conservation of energy; underlying unchangingness of endlessly changing objects of perception; and so on. And yet his ideas were clearly wrong or – worse than wrong – actually unthinkable.

John: But in the past the way that we established that something was conceptually wrong was by coming up with a better theory and, given that we do not have that yet, I think we have to say that the jury was out. The reason why the Ptolemaic view was conceptually flawed was because we ended up with a better view that for a while was not as good empirically, but eventually we saw its (explanatory) superiority. And the reason why phlogiston was rejected was because we came up with something better and until we have that something better maybe we should not be so hasty to throw out something that has not been tried yet.

Ray: Ptolemaic theory was essentially a Fourier analysis of planetary movement and it was pretty good. It was overthrown because it did not predict certain things. When it comes to computational theories of mind I am not too sure that they are exposed to that kind of falsifiability. That is what makes them extremely interesting – and worrying! – to me.

Acknowledgements

The comments made by Ray Tallis are based largely on an article he published in *The Philosophers' Magazine* and we would like to thank Julian Baggini from TPM for allowing us to republish these comments in this format.

Note

1 A debate between Professor Ray Tallis and Professor Igor Aleksander held at Kant's Cavern, The George Tavern, 213 Strand, London WC2, UK
7th March 2007

About the authors

Raymond Tallis was Professor of Geriatric Medicine at the University of Manchester until 2006. His research interests are in epilepsy and stroke and he has published over 250 scientific papers, won many awards (including the Lord Cohen Gold Medal for research into Ageing) and was elected Fellow of the Academy of Medical Sciences in 2000. He has published numerous non-medical books and articles. These include volumes of verse, a novel 'Absence' and over a dozen books on literary theory, cultural criticism, and philosophy, in particular the philosophy of mind. He has received honorary degrees of DLitt from the Universities of Manchester and of Hull for his philosophical writings. He has three books coming out in 2008: 'The Enduring Significance of Parmenides' (Continuum); 'The Kingdom of Infinite Space: A Fantastical Journey Round Your Head' (Atlantic); and 'Hunger' (Acumen).

Professor Igor Aleksander, educated in Yugoslavia, Italy, South Africa and the UK, has taught electronic engineering and computing at many UK universities and has been at Imperial College for the last 23 years. One of the founders of this journal, he has been researching information technology, artificial Intelligence, modelling of the brain and machine consciousness for over 40 years. He has authored over 200 papers and 13 books, the most recent being: 'The World in my Mind, my Mind in the World: Consciousness in Humans Animals and Machines'. He was awarded the year 2000 Outstanding Achievement Award by the Institution of Electrical Engineering for his contributions to Informatics. He is currently investigating virtual neural machines that are capable of emotion-evaluated planning. He spends time in France and Greece where '...the books get written'. He enjoys playing jazz drums, chess and tennis.