GRESHAM

COLLEGE

EXPLORING THE BRAIN

Lecture 15

THE BRAIN AS A COMPUTER

by

PROFESSOR SUSAN A. GREENFIELD MA DPhil
Gresham Professor of Physic

27 November 1997

GRESHAM COLLEGE

Policy & Objectives

An independently funded educational institution, Gresham College exists

- to continue the free public lectures which have been given for 400 years, and to reinterpret the 'new learning' of Sir Thomas Gresham's day in contemporary terms;
- to engage in study, teaching and research, particularly in those disciplines represented by the Gresham Professors;
- to foster academic consideration of contemporary problems;
- to challenge those who live or work in the City of London to engage in intellectual debate on those subjects in which the City has a proper concern; and to provide a window on the City for learned societies, both national and international.

Gresham College, Barnard's Inn Hall, Holborn, London EC1N 2HH
Tel: 020 7831 0575 Fax: 020 7831 5208
e-mail: enquiries@gresham.ac.uk

Gresham Lecture Physic 15
Professor Susan Greenfield

The Brain as a Computer

What is consciousness? As computers become ever more powerful, potentially personalized to HAL-like extremes, so the temptation increases in seeing them as a way to understanding this most tantalising of questions. The idea that non-biological artefacts could be used to throw light on consciousness far outdates even electricity. Ever since the publication of Julien Offroy de la Mettrie's 'L'Homme Machine' over 250 years ago, and its fierce opposition by 'romantic' vitalists, the debate has been rehearsed, recast and rephrased, but never clarified. The 1940s saw the dawn of cybernetics in the 1940s and a quickening if interest in the prospect of brain-like computers and computer-like brains. As the power of artificial systems has escalated beyond all expectation, so has the expectation of what can be discovered about consciousness, using artficial intelligence. Entering the fray are the neuroscientists, backed with remarkable discoveries about the biological brain from the last quarter century, but only recently realising that they can now contemplate consciousness without losing scientific street-cred. If ever the time was ripe for assessing once and for all the connection (if any) between consciousness and computers, it is now.

For the last thirty years brain research has flourished within the paradigm of neuronal communication: 'synaptic transmission'. In brief, a neuron generates an electrical signal due to a transient change in the distribution of ions, and hence charge, between the inside and the outside of the cell. This impulse is then propagated to the end of the neuron whereupon it causes the release of a chemical (a 'transmitter') which diffuses across the narrow gap between cells (the 'synapse'). Once the transmitter reaches the target neuron on the other side of the synapse, it triggers a change in the distribution of ions and thus, in this second cell, causes the generation of a further electrical signal (an 'action potential'). During the 1960s and 70s much was made of the fact that some transmitters triggered the generation of action potentials, 'excited' a cell, whereas others suppressed these electrical signals, 'inhibited' the cell. Inhibition and excitation were seen as the building blocks of brain functioning. How seductive it was to draw parallels with a digital computer, with its on/off switching.

Moreover, some brain processes proved highly tractable to computer modelling. For example the cauliflower-shaped structure at the back of the brain (the 'cerebellum'), plays a large part in the co-ordination of senses and movement needed for sophisticated skills such as driving and playing the piano. It really did look as though the computer was a useful analogy for the brain. Without doubt, some parts of the brain do indeed work like a computer, but how do these processes relate to consciousness? The very skills that the cerebellum enables us to perform are executed without conscious awareness. Obviously when driving, we are not globally unconscious, but we *are* unconscious of making the decision to press the brake when we see a red light. It is no coincidence that the computational approach for modelling brain functions has worked best for processes such as these that are 'automatic', namely unconscious and machine-like.

In contrast, consider movements that are not 'automatically' triggered by an external sensory cue, but rather spring from the inner world of one's individual consciousness. This translation of thought into action is the very link that is weakened in Parkinson's disease: the patient wants to move, but cannot. Parkinson's disease is caused primarily by a lack of a particular transmitter, dopamine, in a certain population of neurons. But other neurons in other parts of the brain also use dopamine, and are not affected in Parkinson's disease. Some such neurons are instead implicated in the totally different disorder of schizophrenia: in this case however, schizophrenia is associated with a functional excess of dopamine. Incidentally, it is because the same transmitter has different roles in different parts of the brain that patients suffer side effects of the drugs used to respectively enhance (L-DOPA) or block (chlorpromazine) the actions of dopamine: Parkinsonian patients can experience schizophrenia-like hallucinations, whereas schizophrenics can suffer from Parkinsonian-like disturbances of movement.

How could the actions of dopamine be modelled on a computer? It would not be good enough to just have a means of exciting or inhibiting nodes. Other transmitters which can, like dopamine, change the electrical signalling between neurons in a comparable inhibitory or excitatory way, nonetheless play no part in schizophrenia or Parkinson's disease. Conversely, dopamine itself can also have a variety of different actions. depending on the molecular target upon which it acts, which in turn determines changes in certain types of ion flux, which ultimately causes differences in inhibition or excitation. In addition, just to make life really tough for anyone attempting to build their own brain, dopamine, as many transmitters, need not be simply excitatory or inhibitory after all. Instead, they can 'modulate' coincidental, pre-existing or potential signals, without themselves having any effect. These biasing actions should not be equated with memory: rather, modulatory signals will last from seconds to minutes, and perhaps hours. Modulation is an increasingly fascinating topic to neuroscientists because it enables the brain to vary its responses incessantly, capriciously and transiently from one moment to the next.

Interestingly enough, it is these modulatory actions of various transmitters that might well be the target of drugs known to modify mood and hence subtely change teh quality of consciousness. Prozac, morphine, amphetamine and LSD all work in different ways and/or involve different transmitter systems, and result in different types of conscious states. Hence there is obviously a strong chemical-selective factor in determining the nature of one's consciousness. It would be hard to see how this chemical selectivity could be preserved in computer models. Admittedly, advanced machines are no longer in the thrall of digital on/off operations, and a silicon retina and neuron have been built with analogue (ie dimmer switch) properties. Yet even so, any one of the various *actions* of dopamine could be factored in, for instance saying it had a 'ten-times enhancing action': but how would a modeler factor in the whole range of different, simultaneous actions of dopamine, whilst at the same time precluding similar actions of non-dopamine transmitters? In short how could the chemical identity ot a transmitter, -with its divergence actions yet molecular distinction from other transmitters with convergent actions,- how could this chemical signature be realised faithfully in silicon?

Moreover, as reflected in the side effects of drugs such as chlorpromazine and L-DOPA, the actions of dopamine are different in different brain regions, so it is not immediately obvious how one would programme in site-specificity, where in addition multi-way chemical see-saws give each region its own pharmacological profile. It is these highly regionalised transmitter balancing acts, the divergence and convergence in transmitter action and transmitter-specific neuromodulation which all endow the brain with an extra dimension and which, in my view, sets it apart in any realistic way from silicon intelligence.

Another computational tack for studying consciousness, adopted for example by Dan Dennett of Tufts University, is to carry on building systems that can do what brains do on the assumption that as the system becomes ever more complex and approximates the complexity of a biological brain, it too will become conscious as a natural consequence of its own complexity. One problem here is that although I (and I assume everyone else) has a feeling about what consciousness is, and a firm conviction that at least one is oneself conscious, the term eludes an operational definition. You might be moving or speaking, but of course you do not need to be. Conversely, movement and speech can be contrived mechanically in the simplest of toys without any but the youngest child imputing an independent awareness. The quintessential feature of my consciousness (and presumably yours and everyone else's) is that it is subjective: everything else is superfluous to its definition.

How then will an outsider test you, and ultimately a computer, for consciousness? Almost fifty years ago the mathematician Alan Turing devised his hypothetical Turing Test: a computer would be deemed to be conscious when an interviewer, with impartial access to both a machine and a person, could not distinguish the two. Modified Turing Tests are now run in US. The modification is to restrict the subject on which the computer/person may be questioned. Even with this massive advantage of not needing to display all the vagaries of broken trains of thought and illogical associations that so characterise human mental actions, the computers have still not fooled anyone. It is a rather sobering thought however, that one human was misjudged to be a computer! Moreover, it is hard to see how an arguably conscious dog or human 1 year old might ever stand a chance of passing.

The Turing test highlights the problem of operational definitions of consciousness. Your first-person personal world need have very little relation to the outside one. The psychologist Donald Mackay expressed this dissociation very well when he pointed out that an actor spouting Hamlet's lines and behaving as Hamlet, did not have Hamlet's consciousness. He was not *actually* the tortured prince of Denmark.

Leaving aside the problem that we would never really know in any case if a computer were conscious, is it at all likely? Certainly there are those who are expecting the new Jerusalem of silicon over-lords any day now. Marvin Minsky of MIT has claimed that artificial systems of the future will 'think' a million times faster than we do, and that they should be regarded not 'them' to our 'us', but rather as our 'mind children', a term coined originally by Hans Moravec of Carnegie Mellon. In a similar spirit, the Nobel Laureate Gerry Edelman has devised a series of 'synthetic animals' called 'Darwin' that with increasing sophistication as the series has developed, move around in a confined space learning about their environment and acting accordingly, with no externally imposed agenda. Edelman reckons that before the end of the next century, synthetic successors to this type of device will be conscious.

No one as yet expects a working model, but to be convincing computationalists should at least be able to outline a theoretical scheme rather than expect us to accept it with the same blind faith that we might accord to believing in fairies. Surely it would be more attractive as a stategy if we knew at least hypothetically, what might be the special extra ingredient that exists in new conscious 'brains' that was lacking in the previous ones. Moreover, just how would the spontaneous dawning of consciousness in artificial brains tell us any more than studying the development of consciousness in real brains?

One answer might be that a researcher could do more with an artificial brain. But if such artefacts were conscious in the same way as animals, tampering with them would entail the same ethical constraints as with their biological counterparts. In a similar vein, silicon colleagues could not be sent to work in an unpleasant or hazardous environment anymore than a human could. So, confronted with a non-biological system, where consciousness had sponateously emerged, how would we learn anything new about consciousness that could not have been learnt from a biological brain?

A modeler might retort that it is best to start with a simple system, even if the resultant consciousness were a mere pale echo of our own. But even if we imagine a system conscious in some kind of cheesey, low grade way, it would be vital to establish what type, or more accurately, what degree of consciousness such primitive awareness might amount to. Granted the first generation at least would have a 'simpler' consciousnesses: but simpler than what, - a rabbit, a flea, a mid-term human foetus? Of course we can imagine conscious machines: but it is a tautology to say that if a synthetic 'brain' were built that were identical (literally) to that of a rabbit, a flea or a mid-term fetus, it would have the corresponding consciousness. The real problem is that no one can hazard a guess at just how complex a simple, 'minimum kit' system has to get in order to be conscious, be it animal or artificial. The modelers' visions tend to be 'functionalist', namely not to produce physical look-alike brains but rather to reproduce the 'function', a modicum of consciousness in a 'simple' system that need not resemble a biological brain at all. The problem here however is that we would need to understand what a modicum of consciousness was, in terms of principles or properties, before we could look for it.

And there's the rub. The ultimate riddle is that as yet we have no inkling as to how consciousness, a first person experience, could arise from a collection of non-conscious elements, be they made of silicon or from the real brain. As a neuroscientist I know certain events in certain parts of the brain cause the sensation of pain, and others feelings of pleasure. But I have no idea how the one actually leads to the other. How could I therefore even dream about replicating this causal connection in an artefact? What principles would I employ? It really is not helpful to assume, as Minsky does, that so long as the system were sufficiently complex, consciousness would suddenly just be spontaneously generated. Even were such a scenario to evolve spontaneously in a silicon system, how would that help us understand the physical basis of consciousness? For both computationalists and neuroscientists the physical basis of consciousness is the Final Frontier, the most challenging question. But surely the answer will not be reached more rapily by going one step back and dealing with silicon systems where consciousness in the first place is, to say the very least, more in doubt than in real brains.

One more argument posed by computationalists, is almost one of default. The rationale runs that the only ultimate alternative to a buildable brain, is to subscribe to the idea of vitalism: that there is some magic spark in living things, initially referred to some two hundred years ago as '*natura naturans*'. Since this life-force would be irreducible and ultimately therefore incomprehensible, it would clearly not be a satisfactory explanation for anyone pressing for a scientific approach to consciousness. On the other hand, *is* computation the only alternative?

I would agree with artificial brain modelers that it is reasonable to assume that consciousness is the emergent property (where the whole is more than the sum of its parts) of non-conscious elements. But it might well be the case that chemical-cellular systems such as those in the biological brain have emergent properties not realistically realisable in silicon, or any other material, save brain tissue. It is understanding the physical, causal basis of how these emergent properties are so generated in the brain that is, in my view, the ultimate challenge for neuroscience.

But let us not be biologist-ist. It is just that it seems intellectually dishonest to accept artificial minds merely as an article of faith. The challenge is for the computationalist to come up with a realistic strategy, however hypothetical at this stage, of how to model variations in the quantity and quality of consciousness, caused by the ceaseless unfolding of specific chemical symphonies in the brain. Artificial neuronal networks can, of course, display an impressive capacity to learn on their own; they achieve feats of problem solving and speeds of calculation that make us look Neanderthal: they can even exploit light-sensitive protein switches. But when it comes to throwing light on the physical basis for consciousness, they do not deliver.