On Quantum Physics and Ordinary Consciousness

by Stephen Jones

On the question of whether physics has anything to say about consciousness.

I would like to open by referring you to Michael Lockwood's talk. He seems to be saying that we need a better quantum description of the world because at present the physical view has no room in it for such phenomenal matters as **qualia**, i.e. the feels and qualities of the things that we know; and **meaning**, the factor which makes the contents of our consciousness things we know about; their names, their relations, etc.

Michael Lockwood's talk on "The Enigma of Sentience"

I think what he is saying is that the stuff we know as the qualities of things, the information that we have about things has actual physical existence, qualia "are the very essence of physical being" [Lockwood] and so what does this say about the state of physics' description of the world.

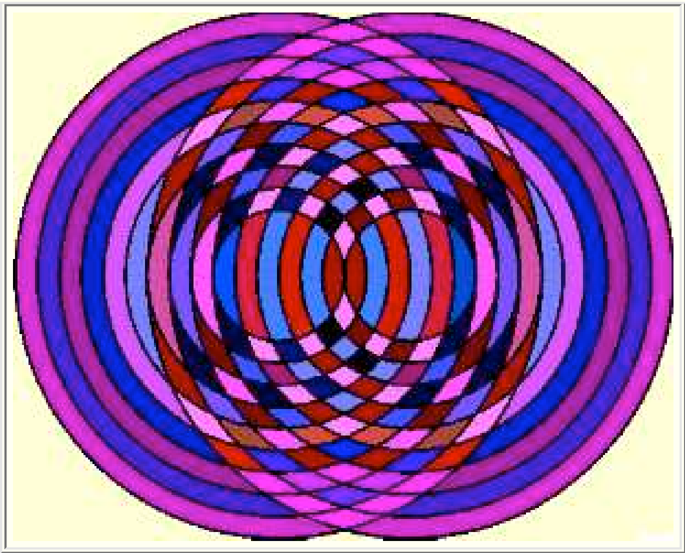
I understand Lockwood to be suggesting that the phenomenal: what we experience and report about; and the physical: what it is that induces and processes the sensations, are two different aspects or **representations** of the same 'stuff'. It might be suggested that these two aspects of the world (its physical emodiment and our experience of it) have a kind of complementary relationship, which others (possibly even Bohr) seem to argue is a relationship analogous to wave/particle complementarity.

On Quantum Physics

The tasks of physics in the early years of this century concerned two matters, one was the macro universe which Einstein dealt with in his relativity theories and the other was the micro universe. Quantum physics is the theory now used in scientific discussion of the micro-universe, that is, the sub-atomic world.

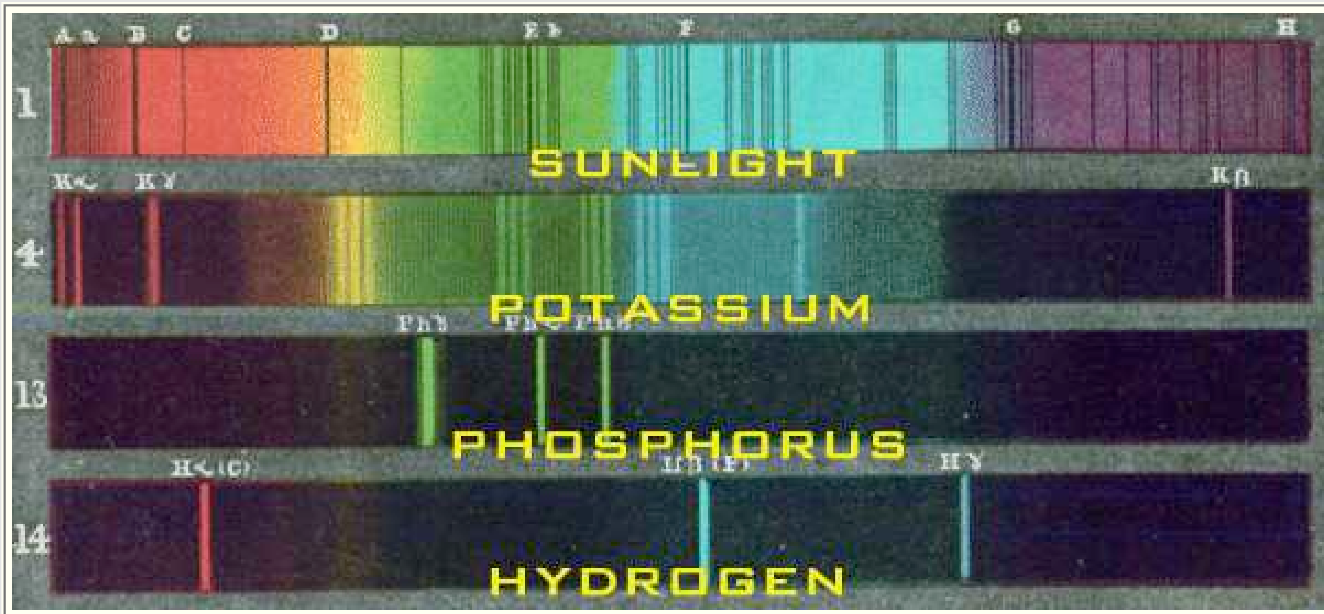
Before it was developed two problems existed for physicists to explicate. One was the problem of the corpuscular versus the vibrational theory of light and the other was the problem of atomic spectra and the discrete, quantised, energy values inherent in electrons in their orbits about the atomic nucleus.

The problem of light was that you could do a number of experiments on light which showed that it had a wave-like or vibrational nature, and you could do a number of experiments which showed that light had a particulate nature. Light's wave-like nature shows up in the interference of light waves passed



through two narrow slits and then allowed to project onto a single screen. These [interference patterns] are analogous to the ripples on the surface of a pond caused by a pair of stones being dropped into the pond. When the ripples mingle, some of them cancel each other out and some of them add together to make a bigger wave.

Light's particle-like nature shows up in atomic spectra and the discrete and consistent patterns of the atomic spectra of different elements. [Max Planck] had shown that an atom when heated to the point of incandescence or when in radioactive decay does not release its energy in a continuous stream, but in discrete bundles or particles such as electrons. It is these particular discrete energy values associated with the release of each electron that are what show up as the lines in atomic spectra and are what became known as **quanta** of energy.



Visible Light Spectra of sunlight and several elements

Now these were quite different sets of experiments showing up mutually exclusive properties of the same thing, namely sub-atomic particles of matter, and of course this created a very difficult problem to be resolved.

Another little matter that was causing big trouble was the discovery

"that it was impossible to describe simultaneously both the position and the velocity of an atomic particle with any prescribed degree of accuracy. We can either measure the position very accurately - when the action of the instrument used for the observation obscures our knowledge of the velocity, or we can make accurate measurements of the velocity and forego knowledge of the position." [Heisenberg,1958, pp39-40].

This is Heisenberg's **Uncertainty relation**. Now, the wave particle duality and the uncertainty relation have a similar characteristic, which is that the kind of experiment you are doing, which involves particular observational instruments, determines the kind of result you are going to get. The measuring instrument gets in the way, or more formally the measuring instrument becomes a part of the system being measurd.

The Danish physicist Niels Bohr was instrumental in getting a co-ordinated view of the implications of all these anomolous descriptions of the sub-atomic world and in the construction of what is now Quantum Physics.

Some comments from Bohr

Bohr, in a talk given in 1938, descibes Heisenberg's "uncertainty relation" by saying that any experiment one might do to determine the "coordination in space and time of the electrons in an atom will unavoidably involve an essentially uncontrollable exchange of momentum and energy between the atom and the measuring agencies" [Bohr, 1958, p19] annihilating any information [any possible knowledge] about that momentum and energy. And conversely any investigation of the momentum and energy of the electron will preclude the possibility of gaining informatioin about the position [the space and time coordination] of that electron. So "experience [knowledge] obtained under (these) mutually exclusive conditions must be (regarded as) **complementary**". [Bohr, 1958, p19].

The formal description of quantum physics might be said to provide a conceptual means for comparing observations. The behavior of microscopic physical systems cannot be described in a language independent of the means of observation.

> "As soon as we are dealing with phenomena like individual atomic processes which, due to their very nature, are essentially determined by the interaction between the objects in question and the measuring instruments necessary for the definition of the experimental arrangements, we are, therefore, forced to examine more closely the question of what kind of knowledge can be obtained concerning the objects." [Bohr, 1958, p25].

We are forced to use classical concepts in the description of the experimental instruments and results, but no results of experiments on atomic objects "can be interpreted as giving information about independent properties of the object" [Bohr, 1958, p26] and must take into account the interaction with the measuring instruments. Information about an atomic object obtained under one set of instrumental conditions may be said to be *complementary* to any information obtained about the atomic object with some exclusively other set of measuring instruments. These complementary informations "represent equally essential aspects of any knowledge of the object in question" [Bohr, 1958, p26]. In fact these complemantary data are necessary for the adequate understanding of many phenomena such as the behavior of light or the electron. The unpredictability of when the electron will be spontaneously emitted in radioactive decay or in the "Schroedinger's Cat" thought experiment is what it is that forces the ascription of a non-causal process within the quantum world. To quote from Heisenberg again:

> "Quantum theory can give us an indication of the probability that the alpha-particle will leave the nucleus in unit time, but it cannot predict at what precise point in time the emission will occur, for this is uncertain in principle. We cannot even assume that new laws still to be discovered will allow us to determine this precise point in time; were this possible the alpha-particle could not also be considered to behave as a wave leaving the atomic nucleus, a fact which we can prove experimantally" [Heisenberg, 1958, p41]

That is, the atomic world is no longer amenable to causal description and can no longer be described using mechanistic principles. The predictable connections necessary in mechanistic descriptions are simply not there. This lever does not connect to that gear chain by this connecting rod.

Now Bohr goes on to say that this situation of complementarity in the properties of the electron is not met with elsewhere in the classical physics but only in psychological investigation. For example he suggests that the properties of living systems are "complementary" to the properties of inanimate objects. He then goes on to say:

> "...the existence of life itself should be considered, both as regards its definition and observation, as a basic postulate of biology, not susceptible of further analysis, in the same way as the existence of the quantum of action, together with the ultimate atomicity of matter, forms the elementary basis of quantum physics." [Bohr, 1958, p19]

Thus Bohr rejects both the mechanistic and the vitalistic views of life. Bohr also speaks against the possibility of reducing the explanation of life to interactions in terms of chemistry and physics on the grounds that

> "the incessant exchange of matter which is inseparably connected with life will ... imply the impossibility of regarding an organism as a well-defined system of material particles..." [Bohr, 1958, p18/19]

similar to ordinary non-living physical matter. Nowadays we use concepts of information and the organisation of a system to show that it is living, so we have removed Bohr's problem of explaining the then mysteries of embryology and development (remember this talk was given in 1938, well before Watson and Crick elucidated the DNA molecule). Whether or not we can properly regard living and inanimate sysems as *complementary* in the way Bohr thinks of that, the idea can definitely be applied to the divergent nature of physical and psychological observation.

Quantum Physics and Consciousness

So why is quantum physics involved in a discussion of consciousness at all? Yes it may well require a non-mechanistic explanation itself but that doesn't tie it in to psychological and phenomenological problems. The first point to make here is that consciousness at the very least manifests through a physical system however complex that may be.
The second point is that the formal description of quantum physics has to take into account the information gathering system as one of its terms. Quantum physics is a theory of knowledge, the knowledge we have of the world.
The third point, and this is the one with the phrase that will be most familiar to you, is that in the orthodox explanation of how one gets from the quantum description of the micro-physical quantum world with all its associated anomolies to the macro-physical classical world with its consistency and stability one has to go from a condition containing all the various potentialities inherent in a particular atomic system to the one actual event that occured on the making of the observation. We have to get from a set of superposed states that exist in potentiality to the actual thing which manifested. This is the **quantum collapse**, the **collapse of the state vector**, and it is the process of something coming from the potential into the manifest. The quantum collapse is the observational act. A physical experiment involves a conscious decision, at the very least, as to what experiment to do.

The observer, the experimenter is necessarily built into the experiment. This is very similar to the effect of the observer in psychological or anthropological investigation and has a consequence which Bohr describes thus

> "...the impossibility in psychical (i.e. psychological) experience to distinguish between the phenomena themselves and their conscious perception clearly demands a renunciation of a simple causal description on the models of classical physics, and the very way in which words like "thoughts" and "feelings" are used to describe such experience reminds one most suggestively of the complementarity encountered in atomic physics." [Bohr, 1958, p21]

I would suggest that Bohr went so far as to imply that the physical and phenomenal worlds bear a *complementary* relation to each other which is similar to the complementarity of position and momentum in the world of the electron. Or perhaps it is more like the complementarity of the wave-like and the particle-like behaviours of subatomic particles, in that these are two systems of description which apply, in superposition, to the same entity. **Superposition**, as used by Schroedinger, has it that the two separately describable sets of properties both hold concurrently.

Schroedinger established the formalism for quantum physics which describes the condition of superposition of the two states potential in the sub-atomic particle/wave being studied. These two states co-exist in the this manner called superposition within the **state vector** and it is the process of quantum reduction to the classical world, otherwise known as the **collapse** of the state vector to which Heisenberg's uncertainty relations apply. This (I think) is what is known as the measurement problem, and it is the point in which the knowledge or observational factor is inserted. Standard quantum physics says that the subjective act of observing, or gaining information from the system is what causes the state vector to collapse into one of the two potentialities hidden in the quantum state.

First I'll refer you to Henry Stapp's discussion of the knowledge term in the quantum formalism and then later I'll look briefly at Roger Penrose's proposal for an objective reduction process which doesn't suffer from the problems arising from the need for an observer who is responsible for the manifestation of the world.
So let's turn to Henry Stapp for his introduction to why the quantum formalism is useful as a basis for a formal description of consciousness.

[see Henry Stapp on The Epistomological Element in Quantum Physics]

Roger Penrose and Objective Reduction

Roger Penrose has probably made the biggest recent impact in discussions on AI and its relation to consciousness as well as on a possible role for a projected new physics in the operations of consciousness. In his two books The Emporer's New Mind (1989) and Shadows of the Mind (1994) he discusses the nature of consciousness and the implications of the search for AI on a science of consciousness. His position regarding AI is that of what is now the standard argument against 'strong' or algorithmic AI. The 'strong' AI position says that:

> "All thinking is computation; in particular, feelings of conscious awareness are evoked merely by carrying out of appropriate computations" [Penrose, 1994, p12].

Penrose uses Turing's concept of **computability** and the result that there are non-computable systems of numbers and their relations; and Goedel's **Incompleteness Theorem** to argue against the idea that it may be possible to construct an artificially intelligent machine, or to simulate human intelligence in a computing system.

[For an introduction to Turing and Goedel see Neural Networks and the Computational Brain.]

The upshot of Penrose's argument is that:

> "Appropriate physical action of the brain evokes awareness, but this physical action cannot even be properly simulated computationally." [Penrose, 1994, p12]

and this leads Penrose to contend that there needs to be developed an extended quantum physics.

Penrose argues that a particular activity of conscious minds, namely **mathematical understanding** cannot be explained within the realm of the classical physics view of the world because it involves the human understanding of, in particular, non-computable numbers. That is, minds can understand things which are not provable within mathematics

His argument seems to depend on the idea that there are things in the world of the mind which are understandings of non-computable mathematical truths. Since quantum physics and classical physics are computable, deterministic procedures; and since Godel's theorem clearly says that algorithmic or computable systems are incomplete, then quantum physics is inadequate to explain the mind. Thus a new (layer of) physics is needed and Penrose offers his theory of the **Objective Reduction** of the quantum state vector as that new aspect to the theory. He thinks that Objective Reduction is a better way of dealing with the mysteries of the measurement problem and the superposition of the two states described in the Schroedinger equation quantum state vector. Very loosely Objective Reduction seems to go something like this:

> If two states exist in quantum superposition each will posess slightly different quantum gravitational fields, which will have slightly different evolution over time. This will induce a divergence in the time evolution of the state vectors of the two states to the point where they become so different that they can no longer co-exist in superposition. Consequently the system's "superposed state would spontaneously jump into one localised state or the other" [Penrose, 1994, p340], i.e. the system will then collapse into one of its potentialities.

Penrose goes on to say that Objective Reduction is a procedure of consciousness. He originally suggested that this possibly happens at the inter-neuron synaptic level (in The Emporer's New Mind, 1989) which (in Shadows of the Mind, 1994) he now doubts because of the scale at which neuron firings occur, and their consequent effect on their environment, the coherence of any quantum system would be hard to maintain, i.e. they function in the macroscopic or classical domain.

The more recent possibility which Penrose canvasses is known as the **microtubule** which is a structure in the cell's **cytoskeleton** (the cells supporting skeletal structure). Stuart Hameroff has done most to elucidate this structure and proposes that objective reduction of the quantum state vector occurs within the very small confined space of the microtubule, and that an orchestrated series of collapses is the source of consciousness.

The Hameroff-Penrose work is a highly detailed analysis of the architecture and scale and possible quantum effects of the microtubule in the neuron. It is probably best that you read their paper: Hameroff, S. & Penrose, R. (1996) Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness.

The main problems with this idea are that no one can see how quantum coherence could be maintained at body temperature; and further that, in that all cells have microtubular structures would not all cells then be conscious? A position which I think Hameroff and Penrose are prepared to accept. For a thorough discussion of Penrose's work refer to the Rick Grush, Patricia Churchland article "Gaps in Penrose's Toilings"

Another version of the possible role for quantum physics in explaining consciousness was presented by Frederick Beck in his paper on quantum selection at the synapse. Beck argues for the synapse as being the point in the neural process which needs quantum explanation. He suggests that it is synaptic transmission which is "the basic regulator of brain activities". [F.Beck, abstract to his presentation to Tucson II]. He proposes that a "quantum trigger", functioning at the atomic level to avoid thermally induced decoherence, regulates synaptic transmission. This trigger effects the capacity for charge tranfer through the post-synaptic terminal via electron tunnelling and Beck ties the low probability of tunnelling events to the low probability of actual neuron firing after any one synaptic transmission event. Somehow he ties this to consciousness by suggesting that the only significant version of this quantum trigger occurs in the processes of Pyramidal cells where they synapse to cells in the uppermost cortical layer. Why this process can be somehow not occurring at all other synaptic transmission events in all other neural cells is not explained.

I'd like to finish with some comments from Paul Davies about what he sees as being the likelihood of a need for a new physics.
Paul Davies on Is a New Physics Necessary?

Some Questions in Conclusion

So where did the idea that a new physics is needed arise from? As Chalmers has suggested; [see Chalmers on the Hard Problem] there is all the stuff we know and will find out about the physical world: the physiological. For example, we can describe how light goes into the eye and is turned into neurosignals by the light sensors in the eye and then is processed for stereoescopy and depth, processed for motion, then processed for identifiable, explicitly encoded, edges and then as we travel upstream into the cortex for the recognised and the novel, for meaning and eventually response. All of these physiologically describable processes are going on, but where is the subjectivity generator or encoded filter or whatever it is.

Is it enough to say that subjectivity is simply a function of the brain, or is there something else needed? and is this a matter of physics? Is consciousness a "field" in some sense? Was Descartes right when he removed the mind from the body and made it something immaterial which communicated to the body through the Pineal gland? [see Some extracts from Descartes] Or is subjectivity simply another class of descriptivity about the world which is "detected through the instrument of the mind"? By which I mean does the phenomenal bear relation to the physical as the wave-like aspects of light do to the particle-like aspects. *Is consciousness superposed on the physiology*? Or is this an unecessary extra layer of description which is better handled using concepts of organisation and complexity, large feedback driven nets. Or worse still does organisation have a physical effect beyond simply the way the physical is hooked up? Does it produce some kind of 'field'?

Now, regarding the brain's capacity to carry out non-computable processes. It is probable that the brain lives on the edge of instability, on the edge of chaos. This gives the brain the capacity to switch states at the drop of a hat, so to speak. A non-linear result is exactly what is necessary in the brain for the capacity to deal with emergencies of one sort or another.

Also, given the extraordinary complexity of the brain and of the social/cultural/linguistic matrix that we live in, any form of oddity is more likely to get an explicit representation in meaning space, so that we can deal with arbitrary symbol attribution comfortably and without becoming too psychotic. (Though I must say one wonders these days just how much arbitrariness we can deal with). In a complex self-regulating system the capacity to deal with anomalies, irregularities and other novel events is utterly essential and made greatly easier by the diversity of processes which can deal with the novelty. The system is capable of dealing with almost anything the world can throw at it. But this is a natural function of complexly organised systems. They are able to handle a huge variety of conditions.

Once we have gained knowledge of the world we gain reflection upon that knowledge and second degree reflection on how we are dealing with that knowledge. "Did I get that right?" In an inconsistent world we have to deal with new events which, so to speak "Do not compute!" That is "I don't understand what is going on!". Also in a brain full of information, memes and ideas, and where memory is at least substantially a process of reconstruction, the merging of formerly distinct ideas within different frames into new ideas (inventiveness) seems unavoidable. This is 'generativity' and is explicitly noticed in Goedel's Incompleteness theorems. [for coverage of Goedel see Neural Networks and the Computational Brain.] The point is that there is nothing beyond the ordinary processes of perception, interpretation, memory and other complex functions of the organised brain, which is needed to account for non-computable results in the brain's activities. So I argue that we don't need to propose a new layer of physics in order to explain consciousness, all we need to do is to get a better understanding of the physiology and its very complex organisation, and its plasticity over time.

References

Bohr, N (1958). Atomic Physics and Human Knowledge. Wiley.

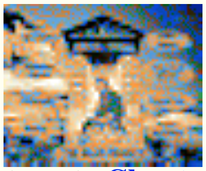
Hameroff S, Penrose R (1996) "Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness." In: Toward a Science of Consciousness - The First Tucson Discussions and Debates. eds S Hameroff, A Kaszniak, A Scott, MIT Press, Cambridge, MA

Heisenberg, W. (1958) The Physicist's Conception of Nature Hutchinson

Penrose, R. (1989) The Emporer's New Mind. Oxford

Penrose, R. (1994) Shadows of the Mind. Oxford

For links to other material on Quantum Physics and Consciousness.

 return to Chapters index	 return to overall index	 return to Introduction and textual index

Michael Lockwood

talks to Stephen Jones at Tucson II.

SJ: What I'd like you to do is talk a little bit about the argument you're putting. It seems to me that what you're doing is putting forward an argument along the lines of: Is there a need for a quantum discussion in this question of what produces consciousness. Am I more or less right there?

Michael Lockwood: Well, yes. I wouldn't say that it was central. My philosophical starting point is that the language of physics simply has no room in it for consciousness as it stands. This has to do with several features of consciousness, one of which is qualia: the feels, the sense and so on, what it's like to have a headache or smell a rose or taste a banana. There just doesn't seem to be room in the theory for those sorts of things. Another aspect of it is the unity of consciousness, the fact that the unity of my mind doesn't seem to be arbitrary and a matter of degree in the way that the unity of any physical object is, ultimately. And a third aspect is the aspect of meaning, that's to say that my thoughts, for example, aren't just meaningless sets of brain events, they actually have reference to things beyond themselves. Indeed it seems to me that meaning is ubiquitous in consciousness. I don't think there's any experience we can have which doesn't involve an element of interpretation, conceptualisation or identification, and I imagine that must be true for the very lowliest organism which has consciousness. Obviously the meaning is going to be of a very basic kind, probably so primitive we don't even have a word for it, but nevertheless I find the notion of consciousness without meaning unintelligable.

But equally it seems to me that the world as described by physics is a meaningless world, once again there just doesn't seem to be room in the framework. So that's one part of it. That's the mystery, what I call the "enigma of sentience". On the other hand, I'm basically optimistic about the whole project of correlating mental states with neurophysiological processes. I don't think I really expect that project to break down. Nor do I think that consciousness is going to be something which somehow exists in parallel with those neurophysiological processes. That doesn't seem to me to make sense either. It seems to me that the logic of the whole project of trying to understand the neurophysiological correlates of consciousness is ultimately going to be that conscious states are physical states, they are firmly embedded in the material world, they are part of the material world.

SJ: Is this the idea of 'supervenience' that I've heard used quite a lot at this conference?

ML: Philosophers use that term a lot. Basically, when they say that mind "supervenes" on the material, what they mean is that you can't have two different mental states, which correspond to the same physical state. And they also mean this in a way that has to do with necessity, that is to say that it's a metaphysical necessity that if anything is in the physical state that I'm in now, then it will be feeling and experiencing just what I'm feeling. And I buy all of that.

But I think that supervenience itself only makes sense ultimately on the basis of an underlying identity. So the moral I draw from what I've just said is, on the one hand, that there's not room in the physical description as it stands for consciousness, but on the other hand, optimism about the project of finding neurophysiological correlates for consciousness. I think that the moral of all that is, yes, conscious states are material states, they are identical with neurophysiological states, but what the existence of consciousness shows is that there's more to matter than meets the physicist's eye. That there's something inadequate about the physical description. Not inadequate simply because it leaves consciousness out, but that the existence of consciousness shows that, in a certain sense, it's systematically incomplete. Ok, does that make sense so far?

SJ: To me there's this whole problem of just what is this extra state or this extension of physics, or that kind of thing, that seems to be one of the major questions that's being bandied around at the moment.
ML: Indeed... Quite.
SJ: So, perhaps you could expand on that a little.

ML: Ok, well, again, part of the story, it seems to me, is that we have to recognise that although we feel that we have a kind of full bodied conception of the physical world. We think we know what we mean when we talk about material objects. I think that that sense that we know what we mean is in large part an illusion. If you look at modern physics, what strikes anybody about it, and, I think, what makes it, in a way, rather off-putting to a lot of people is its incredible abstractness, that it deals basically in mathematical structures. It uses the language of higher mathematics.

Now, what does that mean? I think what that's telling us is that our knowledge of the physical world is effectively only structural. All we know about the world is that it has a certain kind of abstract structure to it. I think that considerations of epistemology, considerations about the theory of knowledge, in a sense, should tell us that we shouldn't expect to be able to know the physical world in any other way. If one accepts the essentially Cartesian perspective, all I really know is what is inside my own consciousness, in my own mind. Everything else is inference. All I can really know about the physical world is that it's a something out there which impacts on my consciousness.

So, presumably there are elements out there which correspond to, and are causally responsible for, elements in here where the perceiving goes on. So in that sense, it seems to me that the only kind of model of the external world that I can have has to be based upon a kind of isomorphism, hmm? An isomorphism. When I look at your face, I assume that corresponding to the different qualia I have about different parts of your face there are things out there. But on the other hand when I think of your face as having a certain colour, and so on, that I think is all projection. What I'm really doing is: I'm taking things that are going on in my consciousness and I'm fleshing out this abstract structure in such a way, as it were to make something real out of it. And in abstract physics, in a sense, one goes on with the abstract structure to greater and greater levels of abstractness, but of course one's capacity to flesh it out simply gives out completely. One's at a loss to picture what's going on when one is talking about projection operators in Hilbert Space or what have you. But in fact, as I say, I think our only knowledge of the physical world is abstract.

But now, the physical world, reality, can't be abstract. It can't be mathematics, the mathematics is just a description. So what we have to suppose here is that the physical world is fully concrete, I mean it is fleshed out. Corresponding to the abstract mathematical structure that we would have in a correct physical theory, there is a concrete physical structure and that physical structure is fleshed out, it has an inner intrinsic nature. The point is that we can't know what that nature is simply on the basis of perception, for the reasons I've given: that we only know it by way of an isomorphism. Similar to the kind of isomorphism that I referred to in my talk: that you have for example, between the pits on the CD and the motions of the piano keys and the pedals when we had that brief recital at the beginning.

Now, the next stage, and this is what I owe to Russell, though it's an idea that really goes back to Kant, and it's made more explicit in Schopenhauer. And, shorn of its specifically Kantian nature, we find it in the mathematician W.K.Clifford. But I got it from Bertrand Russell. What Russell says is, essentially, that all those qualia, all the feelings, the buzz that everybody was referring to today, that seem so difficult to fit into the physical world actually is a bit of the physical world seen in its intrinsic nature. The thought is that, in general, we can only know the physical world abstractly. We can only know its abstract structure. But if we assume that materialism is true, if we assume that our mental states are physical states, then there's going to be a corner of the physical world which we do know. Which we don't know merely abstractly, we actually know what it's like in itself. And we know what its like in itself, we know its inner nature because we are that part. That is the 'us'. That is the mind. That is the 'Cartesian Self' if you like. So, what we're really getting in our own minds is a fragmentary glimpse of the inner nature, which in the physical world in general, is systematically hidden from us. That's part of the story.

SJ: So you're almost saying that the qualia have actual physical being, physical existence?

ML: I'm saying that what we know as qualia is the intrinsic reality which in a sense gets left out in the essentially abstract description which is all that the physical sciences, that includes neurophysiology, ultimately is capable of giving.

SJ: Now, Dennett tries to remove qualia from the conversation.
ML: He does
SJ: And it seems to me that, (and this is my interpretation of Dennett, not Dennett's statement in any sense) he's doing this because the things that are named, which are what qualia usually are, are themselves the outputs of the cellular and neurophysiological processes, the neuronal processes, the neuronal subassemblies a la Greenfield this morning. And that these systems have outputs, and it's those outputs which are the nameable things which are usually accorded as qualia, but I think what Dennett is trying to do is just to say that those are simply the outputs of the systems...

ML: That's not the way I read him. In fact I reviewed his book shortly after it came out. What I take Dennett to be saying is that qualia don't exist. We believe... there are beliefs in their existence, but that's just the story we tell about ourselves. Dennett has this term heterophenomenology. And this is basically, in simple terms, heterophenomenology means the story we tell ourselves about ourselves, right? And qualia are like fictional characters in that story we tell about ourselves. That's his view.

SJ: I guess that's what I mean by the naming - these are the things that are named.

ML: And of course, behind the qualia there are physical things going on. And in a sense the concept of a quale is something that arises in an attempt to make sense of this to ourselves. But nevertheless this is fiction.

SJ: But what you're getting at it that these quale are actually palpable.
ML: I like that word
SJ: they have some existence beyond just being some sort of thing that is named (or is a name).
ML: In a sense they are the very essence of physical being. At any rate they are a manifestation of the essence of physical being, if that doesn't sound too portentous or pretentious. In a sense, what we're grasping in qualia is the very stuff of physical reality, the inner, how should I put it... as it were, the oomph.

SJ: Really your view is more that qualia are the substrate or the base; or that they're the differentiated versions of the base?

ML: Well, they're what flesh out the structure. What we're finding in qualia, what we're sensing is that what we're aware of in being aware of qualia is, in a fragmentary way as I say, is a manifestation of what it is that has the structure. That's the simplest way, the least metaphorical way of putting it. I think that that's literally true.

But then of course, a number of other questions arise. One question is well, alright (and this is what Wilfred Sellars calls the "grain" problem) the trouble is that if it's true that the qualia are actually that which has the physical structure, the kind of physical essence, then how come the structure, the phenomenological structure of consciousness, doesn't seem to match up with the physical structure of our brains as that would be given by the physical sciences. So, there's that problem.

SJ: In what way does it not match up?

ML: Well, that's the point when I quoted Carl Sagan, when the character in his novel says, "think about what consciousness is like, think about what it's like this moment. Does it feel like billions of atoms wizzing about?"

SJ: No, its more like some kind of unified entity, which is where your unity thing comes from isn't it?

ML: Well yes.That goes back to one of the challenges that is presented for a materialist, and that is, as I said earlier, the unity of consciousness. So, as I say, I think the grain problem has to be taken very seriously and I think so does the unity problem. And as it happens I think that quantum mechanics is capable of resolving in principle both the unity problem and the grain problem.

The way in which commonsense conceptualises physical reality, I think, is deeply mistaken. It has tremendous pragmatic value, but if you take quantum mechanics seriously as a universal theory then our commonsense view is way off target. One thing that both quantum mechanics and relativity interestingly have in common is that they say (there are no unique ways of describing reality). You see, I think that question is so nicely crystallised by Carl Sagan, it seems to me there's something that's implicit there. The implication is that there is one, as I put it, canonical way of describing reality. You know, it's like God's story about what's happening. And I think the assumption is: well, God's story is going to involve things like atoms and so on, crucially. So, because we assume rightly that our physics is on the right track, we think that God's story about what's happening will get down to the one true story about reality.
SJ: That's the mind of god?
ML: Well, exactly that's right, that's God's story. Now I think that the message that comes out, both of relativity and of quantum mechanics, is that this very idea that there is a God's story, a uniquely canonical way of describing reality is itself an idea that is effectively discarded. Now if you discard that idea, you can't say well look, here's what I imagine God's story would be like, but, hell, the story that your introspective awareness seems to be delivering ain't at all like that, you know. Therefore you can't match the two.

Now, I think there's no God's story. Both relativity and quantum mechanics basically say there's an infinity of different stories, which are all equally good. And they are related to each other by symmetry transformations. That's to say, they are all equally good because the laws of physics are obeyed whichever story you tell. In relativity, special relativity, you get different stories according to what inertial frame you take as your criterion of rest. Even in classical physics you've got phase space. You can choose different co-ordinates, instead of using position momentum you can use position plus momentum and then position minus momentum. So even there this notion that there are different stories which are all equally good, is beginning to come into focus. In quantum mechanics it becomes crucial.

So, it seems to me that what we really have to see is that the brain (is a physical system). Well first of all, there's a very useful notion that arises, even in classical physics, which I think hasn't really seeped into the imagination of most philosophers, let alone lay people, and this is a simple notion, it's the notion of a physical system. See, I think it's very natural to think, "well okay, if the mind is material, if the mind's really just a manifestation, then, presumably there is, perhaps, a bit of the brain where I live". I mean we were hearing about this, this morning, these intra-laminar nuclei and the thalamus, whatever. There's a bit of it, that is where I live, it is very natural to think. But first of all, one thing we know is we're only a bit of reality. I'm only a fragment of reality. My mind, if it's physical, is only a fragment of physical reality. And I think that it is very natural to think that the way you divide up reality is by a sort of spatial slicing, do you see? Well the point about the notion of a physical system is that it gives you a much more flexible way of slicing reality. What it really says is: take the brain, looked at from the point of view of the physicist, even the classical physicist now, really it's a system that is defined by a vast number of degrees of freedom. This is the notion that degrees of freedom are independent ways in which the system can change state or store energy. A point particle moving in space has three degrees of freedom corresponding to the three co-ordinates. That's a simple example.

This notion of degrees of freedom... The thing is if you've got all these vast degrees of freedom, and any subset of those defines a perfectly good physical system in its own right, and it's a much more flexible way of producing a bit of a physical system. It's really that the notion of a physical system corresponds more to the notion of an aspect of something. But in a sense, this slicing by degrees of freedom is much more fundamental than the notion of slicing by saying the bit to the left and the bit to the right. It's just that we tend to think spatially because space is prominent in the way we visualise nature. But it seems to me that you're getting much more towards the essence of nature if you think of the way God, if we go back to him, that the way God would slice reality would be to slice it along the lines of degrees of freedom.

So a part of reality is just a subset of the degrees of freedom of reality. That's much much more flexible. So that means that we shouldn't necessarily look for a part of the brain where the mind is. It might turn out to be that. But I think the neurophysiologists, neuroscientists are, to some extent, stuck in this very primitive idea of (spatial location). It may simply be that: if the idea is that my mind, my conscious mind, is a subset of the degrees of freedom of my brain, then it can be very spread out. Or it could be different aspects of lots of different bits which are playing a part, do you see?

SJ: You've got this array of physical subassemblies all doing their processing tasks
ML: Yes, quite
SJ: They're producing output which is what I equate with qualia. Now what you're saying, as I read it at moment, is that those qualia actually have some

kind of physical existence.
ML: absolutely...
SJ: beyond simply the abstract numerical output.
ML: They are what make the abstract structure not just abstract structure. They are what have the structure.

There is no way to make this a short story, I'm sorry. Let me bring this to some sort of conclusion. Okay, so you've got your physical system which we can think of in common sense terms, it represents an aspect of the brain. Then quantum mechanics tells us that there are an infinity of different ways in which you can tell the story about that system. The technical term is representations. And these representations actually correspond to different things you might want to measure or observe.

In fact you know, when quantum mechanics was invented, there was a puzzle for a while, because there were two quantum mechanics, there was Heisenberg's quantum mechanics and there was Shroedinger's quantum mechanics. And they both seemed to work but they both looked incredibly different. And it was actually Schroedinger in the first instance who realised, in our modern language, that these were two representations. Heisenberg had an energy representation whereas Schroedinger had a position representation. And they were related to each other by a simple mathematical transformation.

So I take it that there is a consciousness representation. So there's a subsystem of the brain, and there's a consciousness representation. That is to say, that's the story that corresponds to the way things seem to us when we introspect. And there's no reason at all why that story should even so much as mention atoms.

SJ: So consciousness is generating the representation?
ML: Well, I wouldn't say consciousness is generating the representation. It's simply that the representation exists regardless, it's not something that's generated. The representation is just a certain abstract way of describing the subsystem of the brain which is directly manifested in consciousness. That's the thought. There's one representation which is special, because it corresponds to the way consciousness sees itself. Do you see? But its a perfectly correct representation. I mean just looking at the brain as a physical object that representation is perfectly definable in terms of quantum mechanics. The specialness of it of course wouldn't be manifest .

SJ: But does this put consciousness as the substrate?
ML: No...
SJ: Or is consciousness particular and specific to each organised physiological entity?

ML: Well, on the one hand I take it that for every conscious being, there is a physical subsystem, that is to say a subsystem of the total number of degrees of freedom that define the being

SJ: Our representation?
ML: Exactly. Well no, first of all we've got the subsystem, right? and then, if we ask the question well why does that conscious being seem to itself the way that it does? then we need to go a further step and identify a preferred representation. It's not preferred as it were objectively, it's preferred subjectively. If you like it represents the spectacles through which consciousness views its own nature.

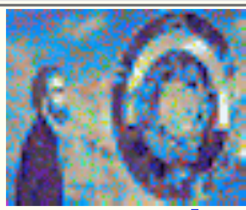
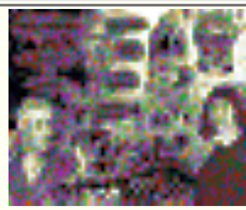
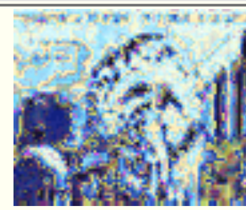
SJ: One of the big problems that I have, in this whole business, which hasn't been discussed very much is the question of culture. The culture, it seems to me has some role in assembling the set of interpretative devices (the spectacles) which to me are what produce qualia. And so...
ML: Oh yes... for example, different cultures that have different colour systems, (Whorf hypothesis)
SJ: Precisely, qualia are culturally relative at least...
ML: I wouldn't say they were culturally relative, I would say they were culturally conditioned.
SJ: Yeah, okay
ML: That is to say: how we grow up, the process of growing up, actually has an imprint, it makes a difference. It makes a difference for the way in which the brain gets wired up! So I think the role of culture here is that it actually helps determine what in detail this subsystem is like.
SJ: Right..., yes...
ML: Well, nevertheless, once it's like that, it's like that in a perfectly objective sense.
SJ: It's a projection onto oneself?

ML: Well, then, I think there's a separate point that, as I say, there is a certain representation (particular to myself). God, as it were, has an infinity of different stories he can tell, he doesn't have just the one story. We just have one story we can tell about ourselves. And that's the story that corresponds to introspection. Do you see what I mean? And if you ask well what does that correspond to physically? That story corresponds to one representation of a particular subsystem of the brain. That's my story. That's what I'm saying. So that's one way in which it seems to me you can have a perfectly good representation where atoms do not such much as get mentioned. Then it goes on and on in a way. But that's probably enough....

SJ: That's great.This whole thing is fascinating.

Links:

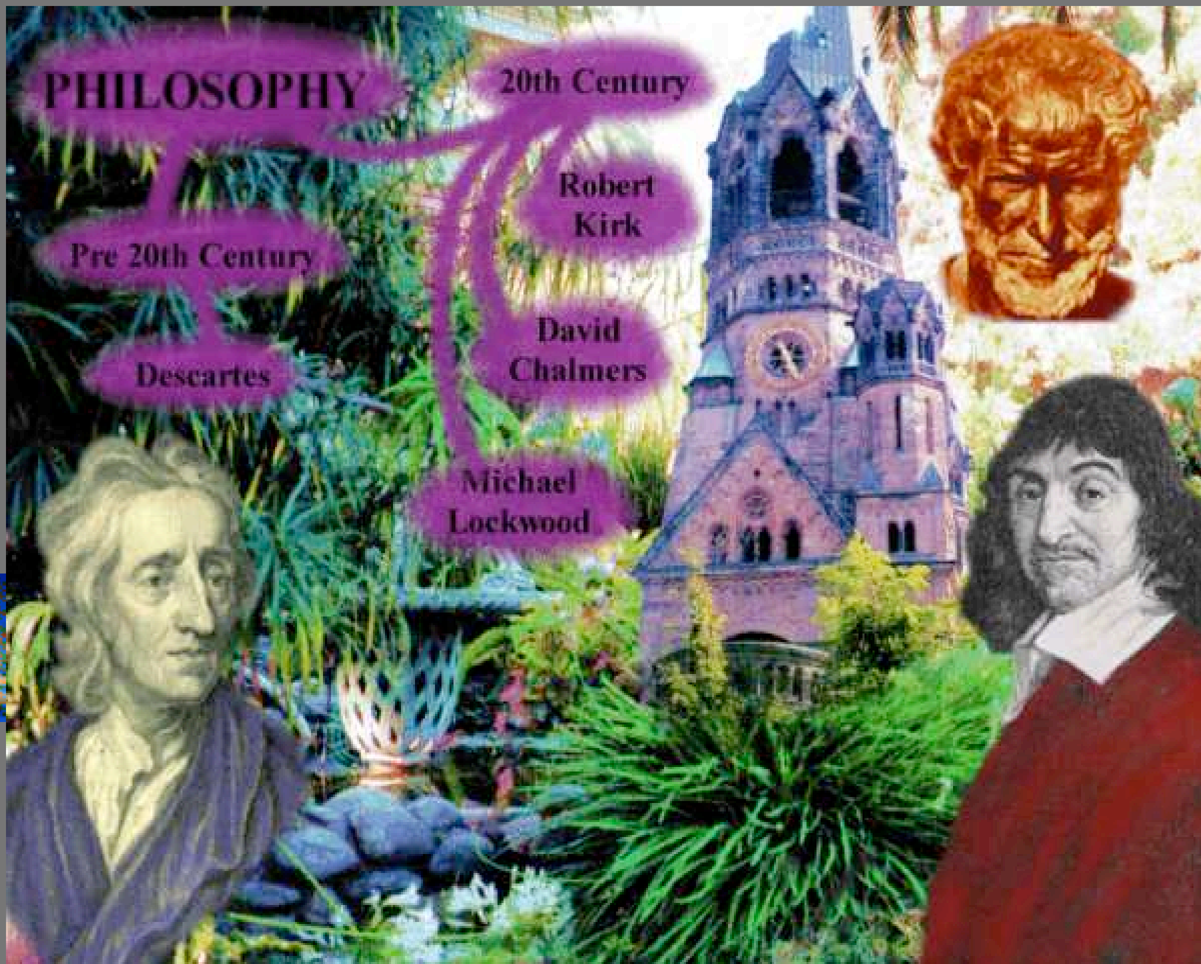
David Pearce reviews Mind, Brain and the Quantum by Michael Lockwood.

 **return to the quantum discussion**	 **return to Philosophy index**	 **return to overall index**	 **return to Introduction and textual index**

The BRAIN PROJECT

Return to [Top Index]

Return to [Chapters Index]

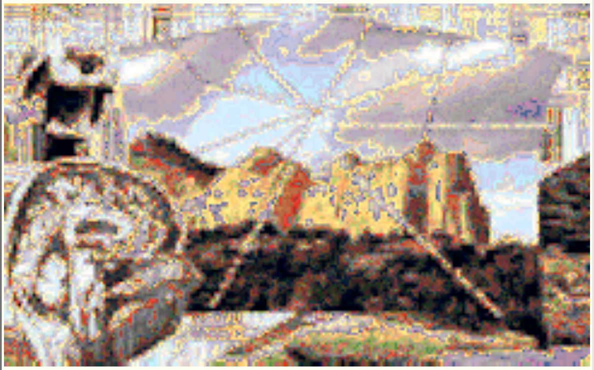
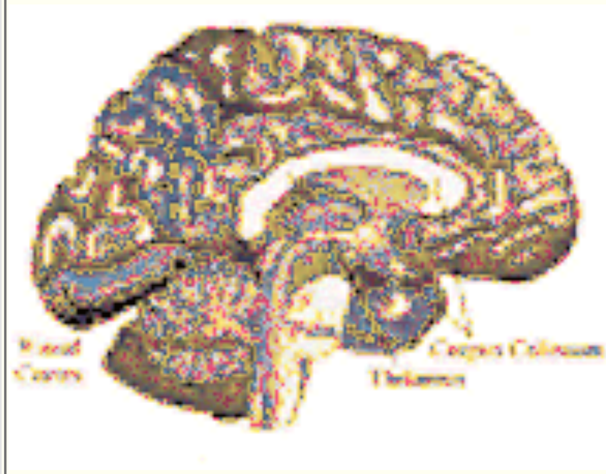
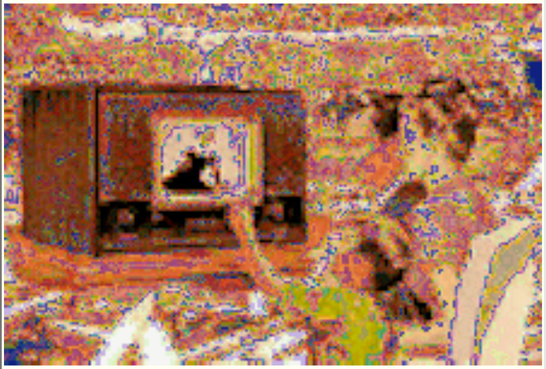


The Brain Project

by Stephen Jones

A Web Project supported by the Creative Development Branch of the [Australian Film Commission]

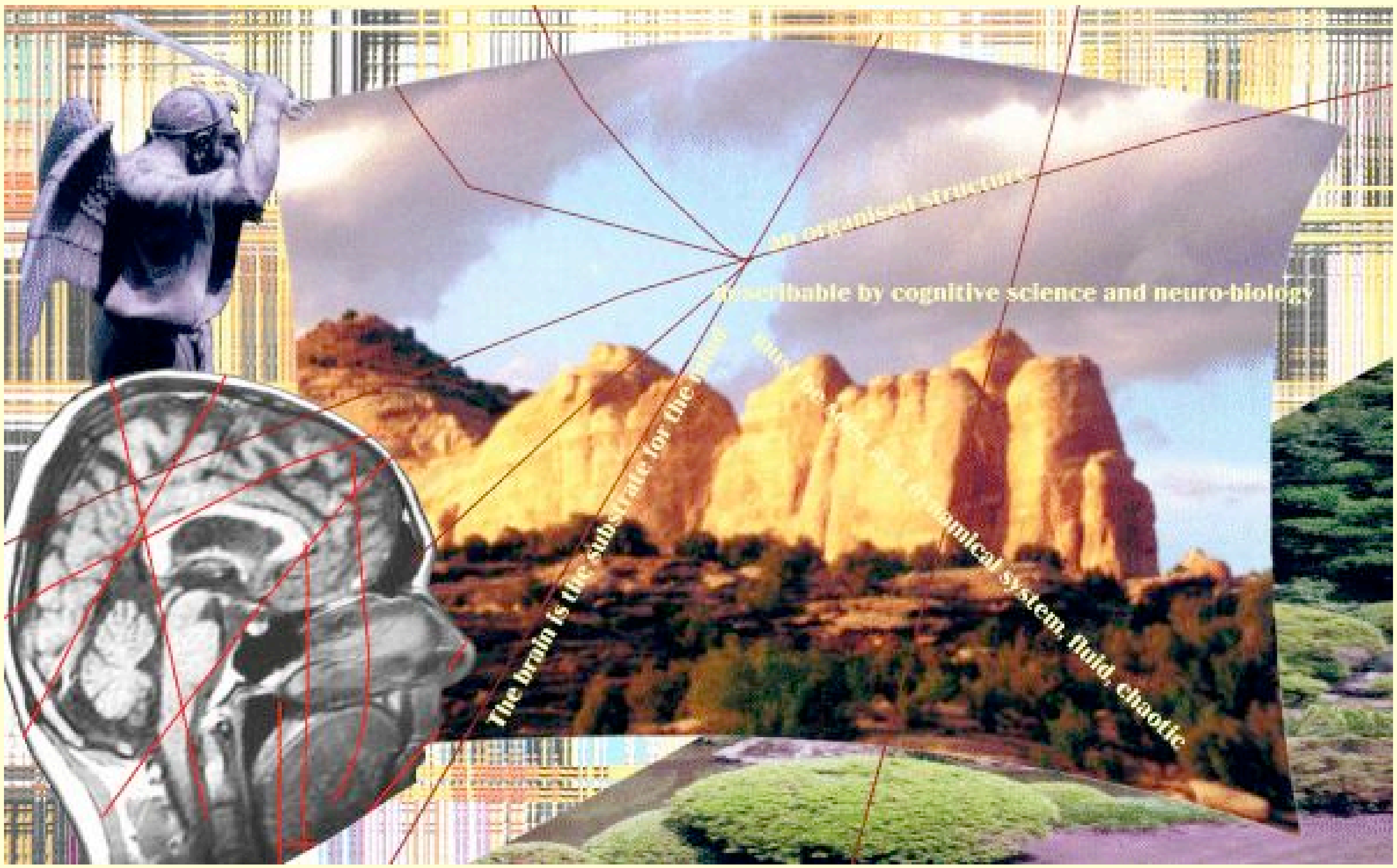
Chapters on various issues relating to the nature of consciousness. Plus papers on video and other matters of interest, including language, cybernetics, interactivity and computing machines.

self-portraits from the inside	**The Brain Project**	**Video** and Other **Projects**
		

Stephen Jones is currently in receipt of a fellowship from the **New Media Arts Fund** of the **Australia Council**, the Commonwealth Government's Arts Advisory body.

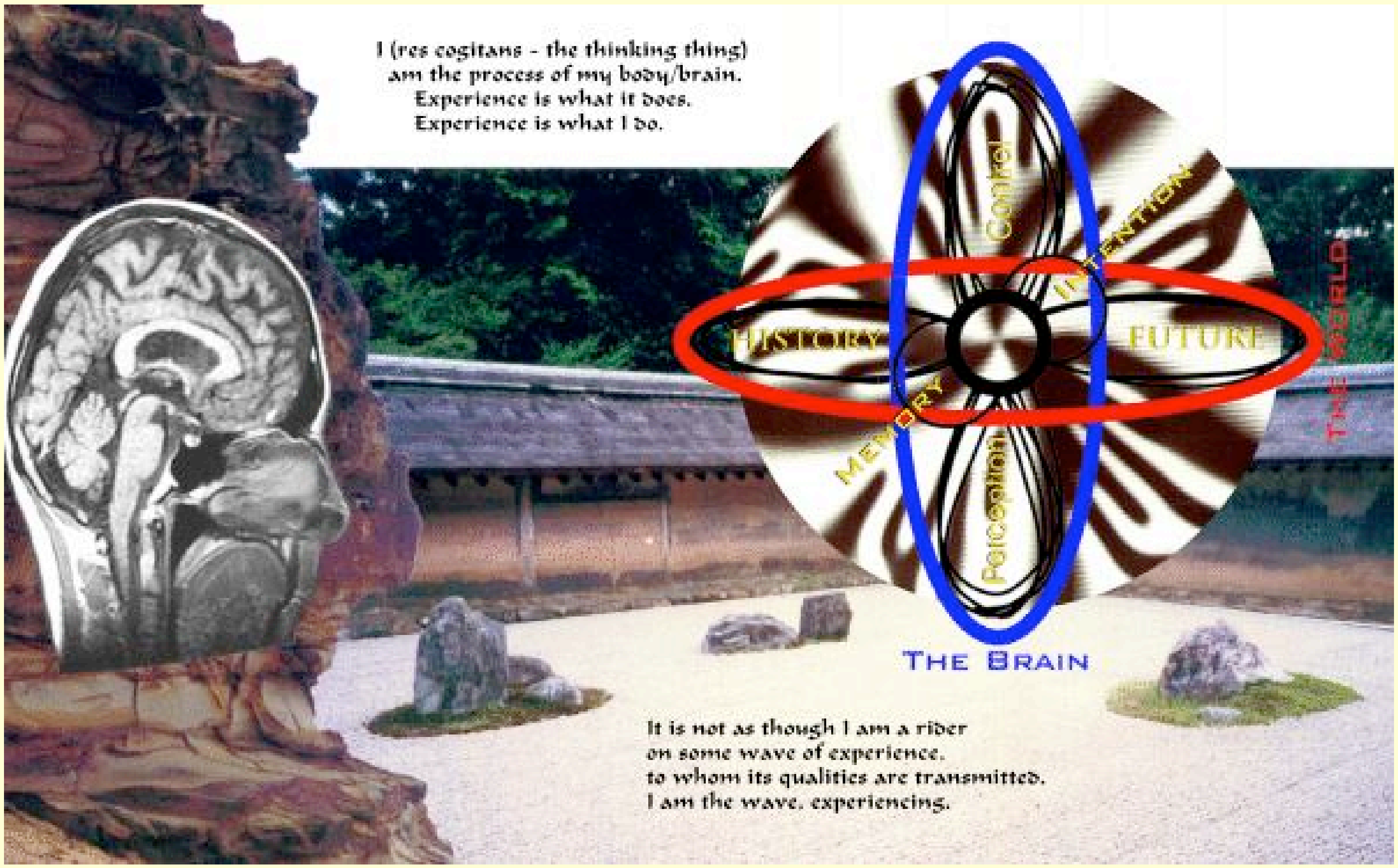
< Culture Domain Home

Self Portraits from the Inside.



Subjectivity seems to be something quite different from the physical world. For a start it is private, experiential, incredibly difficult to communicate and then only by reporting at a third person level. So how is it that one can have subjectivity when, to all appearances, one only has a physical system with which to experience it? Is the physical body all that is needed for consciousness or is there something more involved?

The Brain is the substrate for the Mind ... an organised structure
... describable by cognitive science and neuro-biology.
It must be understood as a dynamical system, fluid, chaotic.



I (res cogitans - the thinking thing)

am the process of my body/brain. Experience is what it does.

Experience is what I do.

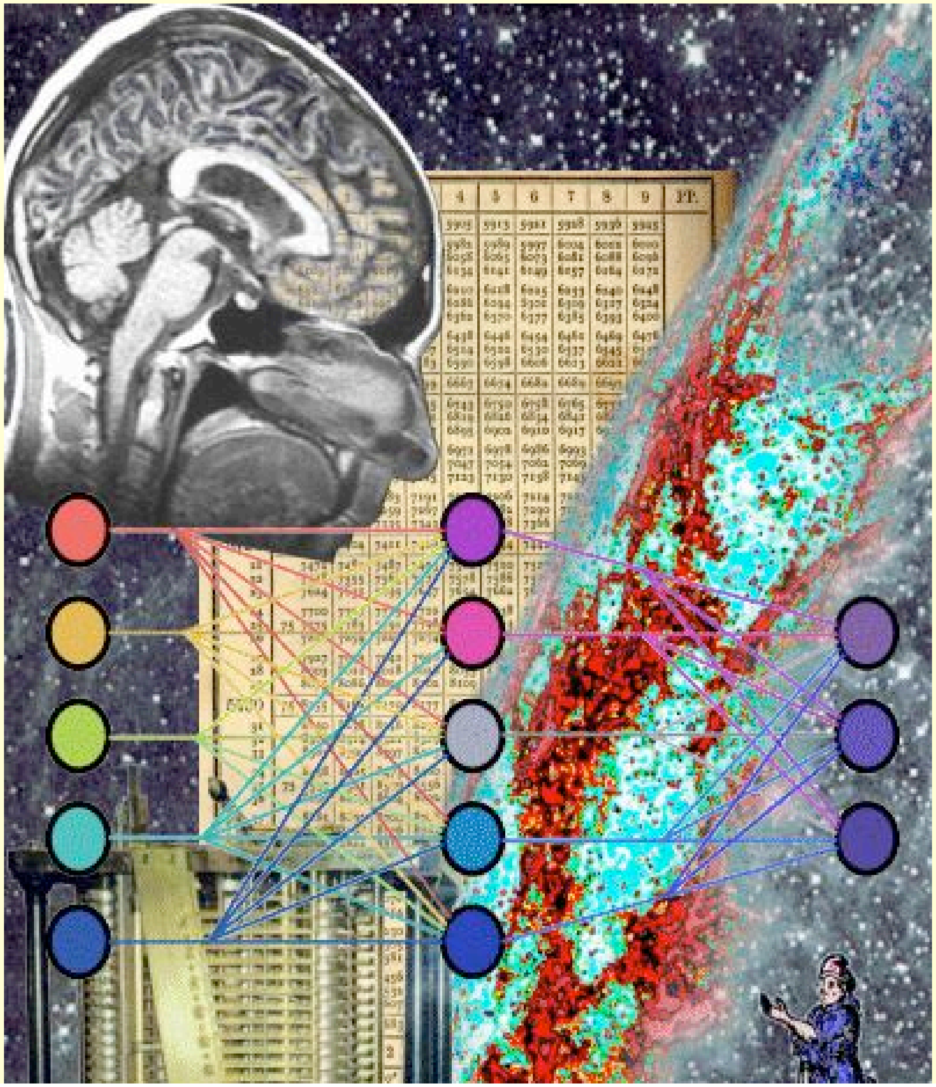
It is not as though I am a rider

on some wave of experience,

to whom its qualities are transmitted.

I am that wave, experiencing.

Subjectivity is essentially the first person experience of the activity of the conscious brain. In other words, it is the experience of being inside these processes of the physiological brain as described within the language of phenomenlogy.

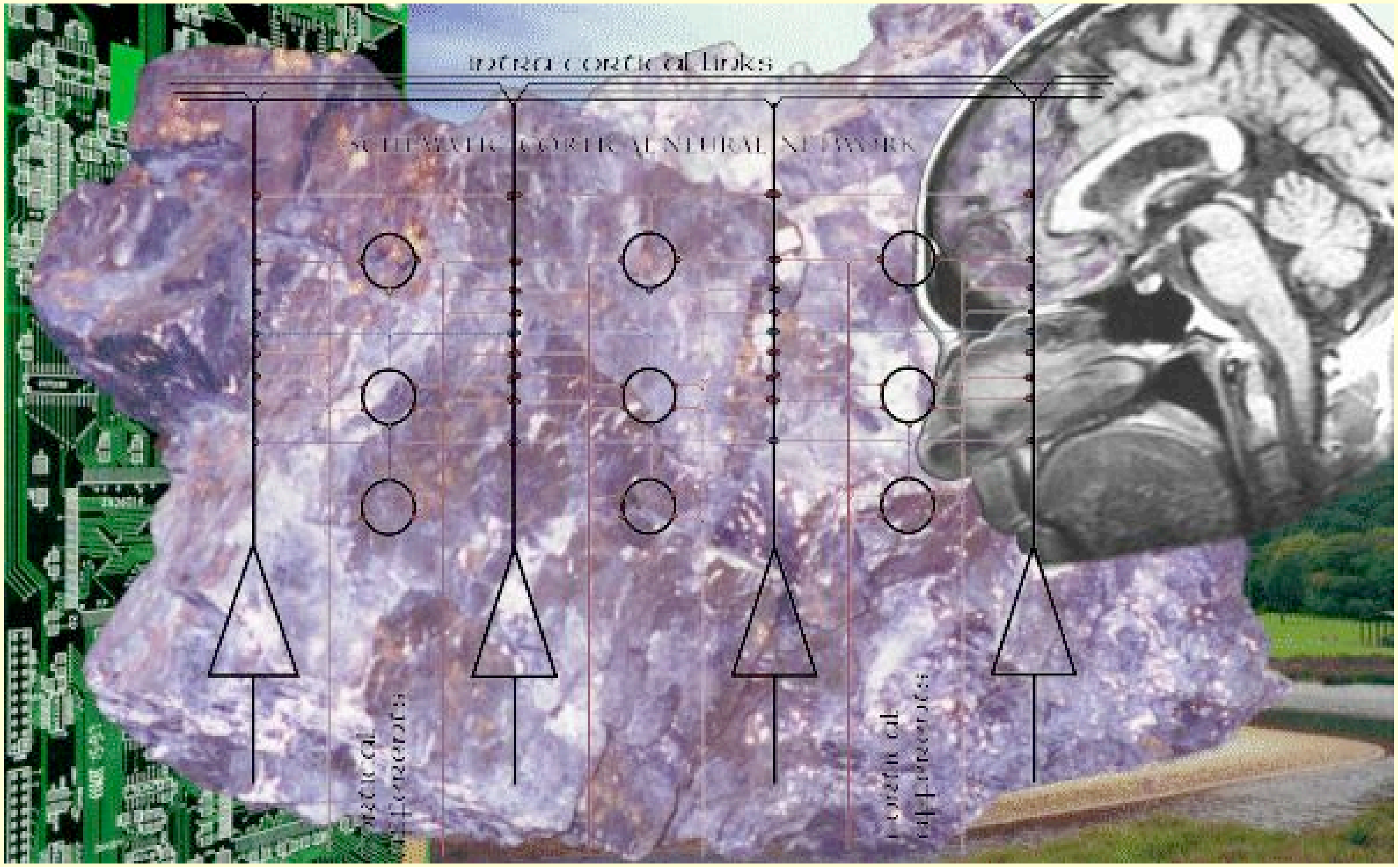


We model the world
in our minds,
we model our minds
in the world
Engines of Analysis
Engines of
Synthesis
What it is that is out
there is not what it
is that is in our
Brains. We
transform and
construct to make
the world
continuous and

contiguous

Our consciousness of the world must be dynamic, in order to keep up with the world, and causal, in order to get what we need from the world, otherwise there is no point in having it. If there is no point in having consciousness then it most likely wouldn't have evolved.

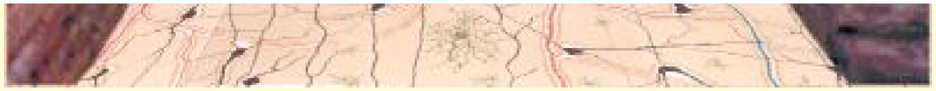
We are looking at a multiply bifurcating neural processing system forming a processing hierarchy with feedforward and feedback pathways enabling feature extraction, quality recognition, object reconstruction, recognition and



naming, cross referencing and global binding enabled by the reverberant structure of output being routed back into input layers and cross-fed into other modality pathways. An extraordinarily complicated but highly organised structure for being dynamically active in the world.

Evolution encourages the development of processing systems for those features of difference which provide information to help the system maintain its presence in the world long enough to produce offspring. Establishment of NN connectionism through self-trained and culturally-trained weighting of synapses will effect the discrimination of features which further enhance survivability.

The world is dynamic: informational change, a form of differencing, is continual; our place within that flux is always changing and thus also generating differences. The informational contents of the neural system is always changing, and so the output 'transforms' presented to higher levels will always be changing, not in topography, but in surface features. These 'surface features' are, so to speak, 'displays' of the dimensions along which some particular feature



extraction process occurs.

The contents of the brain may well be like the contents of a language, codes for those things apprehended, as words are codes, signs standing in for the object. There is only the complex of processes (the patterns of activation, the addressing structures) standing for the object in the brain. The known world is not congruent with what is 'out there' in that everything we know of the world is contained in the processing system which we call the brain. What we know is not the world, but our sensory processing of its waves and disjunctions, a virtual world.

The flow through this network is as dynamic as I am, as dynamic as the world around me and my body might be from moment-to-moment. But the point is I undergo the process, in fact I am the process. Being the process is to experience it. My physiology, my internalised culture and my processing of the present,

experience all this because it is flowing through me. Live. I live it. I am not some rider of the wave, I am the wave. This is an 'identist' position because I can't see any other possible way of viewing the situation. It is completely impossible for me to divorce my experience from my physical system because then I wouldn't be able to experience having a physical system which is perceiving and producing those philosophers' illusions: qualia. I would have thrown the baby out with the 'liquidity'. Why would (how could) all this be going on if it didn't have meaning and wasn't experienced?

The dynamics of the system give its difference and learned experience is meaningfully informational. The whirlpool of feedback gives it endurance and (short-term) memory at whatever scale one is working at at the time. But one is in it, inside it, it is oneself undergoing all this, (there is no 'double aspect', the information is embodied, it can not be any other way). Yes we know the states of our system intrinsically, because it is us, our embodiement, we are not a separate layer observing this thing we are the first-person inside it. We are that process. There is not some experiencer applied to the task there is simply the process, the undergoing, and I have to say, the experience.

return to overall index

Chapters on the brain and the nature of consciousness

The thumbnail images refer to different sections of the site.

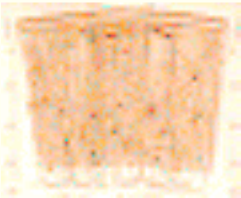
textual index & introduction

there are also:
Papers by other Authors

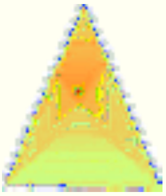
history

chapters

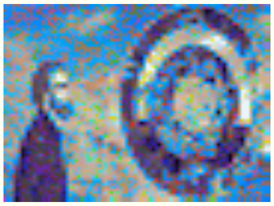
philosophy



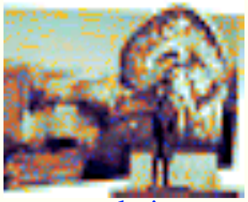
Neuro-physiology



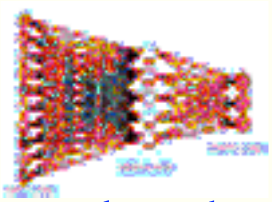
ENTER through
chapters



quantum physics



speculations



neural networks

organised systems

up to top index

A SERIES OF FORUMS inspired by the conference:

TOWARDS A SCIENCE OF CONSCIOUSNESS

held at Tucson, Arizona, 6-13th April, 1996
Presented by Stephen Jones

Topics of the Forums	General Introduction to the Forums 1: Philosophical Issues 2: NeuroAnatomy and NeuroPhysiology 3: Quantum Physics 4: Neural Nets and Artificial Intelligence 5: Cybernetics, Organisation and Complexity, and the cultural milieu.

Background notes on historical ideas of the Brain and the Mind	The classical view The Classical Greek view of the Mind: the Elements and the Humours. The Humours and the rise of Mechanism The shift from Galen's Humours to Anatomy and the Mechanistic description. Extracts from Descartes Extracts from Rene Descartes' philosophical analysis of the Mind and the Brain. Early Neuro-Physiology On the development of Neuro-Anatomy and the Localisation of Function.

Papers by other Authors	The Hornswoggle Problem. by Patricia Smith Churchland. Philosophy, University of California at San Diego, & the Salk Institute. (12 August '96). Does Consciousness Exist ? by Dr. Jayant Sharad Vaidya MS DNB, Academic Department of Surgery, The Royal Marsden Hospital, London, U.K.

New papers of mine	1. Notes and Suggestions towards an Hypothesis of Consciousness 2: What would a Conscious Machine want to do for Itself? 3: A note on a possible physiology of subjectivity, and some comments on what a conscious machine might want to do for itself.

Earlier papers of mine.	Notes on the Cybernetics of Language and Video. (1979) Discussion of social role of the cybernetic model, feedback, conversation and interactivity. With particular reference to Video art. [image intensive] On Animation: The Illusion of Life. (1988) An offering of an hypothesis from a paper delivered to an animation conference:

Just how did this happen ?

In April 1996 I attended the second Towards a Science of Consciousness conference (Tucson II), presented by the University of Arizona in Tucson, Arizona. This gathering brought together many of the major investigators in the field to present papers, to share data and explore their ideas. Discussions and presentations covered the Philosophy of Mind and Knowledge (Epistemology), Psychology, Neurophysiology, Cognitive Science, Computational Science, and Quantum Physics.

From August 14th - 18th, 1996, I presented a series of public forums at The Performance Space Studio, 199 Cleveland St., Redfern, N.S.W. in which I reported on the Conference and introduced the background to some of the major topics presented there.

Each day of the Forums involved presentation and discussion/exploration of a different sector of ideas and debate on the question as to how Consciousness arises. Whilst at the conference I recorded a number of video interviews with some of the main presenters. These formed the basis of this reporting and further work. The video, edited transcripts, supporting materials (books, research papers, graphical and video) were made available and an Internet web connection was installed for the period.

This web site is a continuing development project and will be a regularly enhanced presentation of the materials used in the forums plus materials contributed and generated. I am attempting to provide background briefings, summaries of the main issues of the Conference as well as offer opinion and discussion of those issues and perhaps to synthesise some sort of overall suggestions towards a theory of consciousness. If you would like to comment or contribute a paper please feel welcome.

These sites are for people who are interested in the area but not specialists, as well as those who are specialists and interested in the inter-connections with other fields of thought about consciousness. They should provide an opportunity to mix the communities and stimulate an exchange of ideas.

Contributions to the site are invited from interested people. If you are engaged in the active study of consciousness, or related areas, and would like to make a contribution please contact Stephen Jones. Contributions can be in the form of papers for the site or links and references to other sites. To contribute or comment please email me:

`email:` sjones@merlin.com.au

The presentations at the forums were made with a web page which was on line for pulling up links to other sites and as well used a special helper which controlled a local video server. All the videos were thus available for playback, each segment logged and stored on a dAVE digital disk recorder set up as the video server for random access to material for points of discussion.

This was a physical hypertext show, a physical and virtual system of informational files, performed in the process of thinking about and discussing Consciousness. As presenter I functioned as a hypertext processor, providing the backgrounds and introducing and presenting the video'd talks from some of the speakers.

This series of Forums is part of the **Brain Project**, a development of a documentary opera about consciousness, and is supported by the Creative Development Branch of the Australian Film Commission.

Topics of the Forums (and thus of this WebSite)

1: Philosophical Issues

The philosophy behind, and an overview of, the research programs in the search for and understanding of the processes by which consciousness arises. Some history of the philosophy of mind: Classical to Descartes to now; the current setup of the debate. Getting issues clarified: What we think consciousness is, its definition.

Go to **The Philosophy behind Ordinary Consciousness Part 1: Pre-20th Century** and **Part 2: The 20th Century** which includes talks from **Robert Kirk** and **David Chalmers**.

Robert Kirk, of the Dept of Philosophy, Nottingham University. On the

"Basic Package" in which he describes what it is to be conscious in terms of what he calls the Basic Package. A set of behaviours that all conscious beings have but which is lacked by non-conscious creatures and inanimate objects. The Basic Package consists in the capacity of a creature to gather and use information for itself in the modification of its activities and behaviours in dealing with the world.
Robert Kirk: "The Basic Package"

David Chalmers, of the Dept. of Philosophy, University of California at Santa Cruz. On the "Hard Problem" in which he talks about his division of the question of how it is that we are conscious into the hard and the easy problems. He argues that delineation of the anatomy and physiology of the brain, the description of, say, the visual system or the systems of speech, the physiological pathways of pain, etc, no matter how difficult to carry out are all soluble and therefore of the class of "Easy" problems. The "Hard" problem for Chalmers is: given all the physiology and so on, this still does not explain how it is that we have a subjective view of the world and ourselves. So the Hard problem is: Where does this subjectivity come from? How does all the physiology produce subjectivity? Chalmers speculates that there may be two aspects to information, a physical aspect and a phenomenal aspect.
David Chalmers: "The Hard Problem"

2: NeuroAnatomy and NeuroPhysiology

The basic area of research is **Neurophysiology**. This involves teasing out the processing pathways and systems of the Brain. Intro to the anatomy and physiology of the brain: What do we need to be conscious? Some sort of physical (physiological) system for it to run on. Coupled with some kind of basic set of processes which mean we gather and generate information and use the gathered information in dealing with whatever is the next thing that comes along. Covering neurons, synapses, neurotransmitters; Localisation of function; Neural assemblies and neural systems.

Go to **An Introduction to the Physiology of Ordinary Consciousness** which includes references to transcripts of talks from **Susan Greenfield**, and **Bernie Baars and James Newman**.

Susan Greenfield, of the Dept of Pharmacology, Oxford University. On "Neural Assemblies" in which she posits a system of flexible neural assemblies which recruit available undedicated neurons for the tasks required of day-to-day moment-by-moment consciousness.
Susan Greenfield: "Neural Assemblies"

Bernie Baars of the Wright Institute, Calfiornia and **James Newman** of the Colorado Neurological Institute discuss the concept of a "Global Workspace" in the brain and suggest the extended Reticular Thalamic Activating System as the main "consciousness processor" (my term). The eRTAS sits at the hub of a massive number of neural connections from the sensory systems which are relayed upto the cortex and a massive number of connections from the cortex back to the eRTAS which the cortex seems to use to regulate the flow of information up to itself.
Bernie Baars and James Newman: "The Global Workspace"

3: Quantum Physics

Quantum Physics and Philosophy. If, as philosophers from Descartes to David Chalmers suggest, the mind or consciousness is in some way extrinsic to, or at least not explained by, the neurophysiology: Is there a need for a new layer of physics to explain the emergence of consciousness? Intro to quantum mechanics: the macro world and classical mechanics versus the micro world and quantum mechanics. Particles and waves and "complementarity". The "uncertainty" principle. Is consciousness a physical thing? If so can we describe it adequately within a reductionist framework? If not, is there a need for a new physics to allow us to understand consciousness? Other theories, such as the Penrose-Hameroff theory of "microtubules" are discussed.

Go to **Do we need a new Physics to understand Ordinary Consciousness ?** which includes references to transcripts of talks from **Michael Lockwood**, **Henry Stapp** and **Paul Davies**.

Michael J. Lockwood of Green College, Oxford University. On "The Enigma of Sentience". The language of physics seems to have no room in it for "consciousness", because it doesn't have a way of handling "qualia" or the feel of things, the unity of consciousness and nature of meaning. Nevertheless conscious states are firmly embedded in the physical world.
Michael Lockwood: "The Enigma of Sentience"

Henry Stapp of Lawrence Berkeley Laboratory, University of California, Berkeley, California. On "Consciousness in Quantum Theory". Classical physics is not adequate for describing consciousness since it entails no terms for the conditions that give rise to consciousness. But the basic formalism of Quantum mechanics entails conscious experience as a primitive and as such may provide the basis for a theory of consciousness that is consistent with the physical world.
Henry Stapp: "Consciousness in Quantum Theory"

Paul Davies, Professor of Natural Philosophy, Adelaide University, Sth. Australia. On "Is a New Physics Necessary?". There has always been a link between consciousness and quantum mechanics through the involvement of the observer in the experimental process. This does not imply, however, that quantum mechanics plays a crucial role in the conscious brain. That is more likely to be a function of organised complexity.
Paul Davies: "Is a New Physics Necessary?"

4: Neural Nets and Artificial Intelligence

Computational neuroscience and **neural nets**, etc. Is an **artificial intelligence/consciousness** possible? Can we build an intelligent or a conscious machine? History and development of the artificial neuron.
Go to **Neural Networks and the Computational Brain**

5: Cybernetics, Organisation and Complexity, and the cultural milieu.

Organisation and complexity. Cybernetics, feedback. Culture, language and the development process, the maturation of the brain. Knowledge as a modelling process. Hierarchical systems and organisation paradigms.
Go to **On Complexly Organised Systems and Consciousness**
Also a transcript of a talk by Daniel Miller who is a therapist who considers that it is important to know how therapy might alter the physiology. see Daniel Miller: "Organic Process Therapy"

So where, finally, are we? I think we can say with reasonable confidence that consciousness is a function of the brain, and I argue that it is just that: part of the brain's activity. I.e. the contents of consciousness are identical with the patterns of data-flow through the brain. But we have this reflexiveness which is our view of ourselves and we are led by current cultural paradigms to see ourselves as somehow separate from the physical brain. So the problem has become: is consciousness something, so to speak, foisted upon the physiology by some outside agency or perhaps captured from some "mindfield"? Or is it a necessary function of a highly organised system. My view is that it is the latter. That all the dynamic behaviour of that vastly complex organised system of systems of the brain, given all the feedback pathways and the propagation delays inherent in that, produce a kind of resonating system; an heterarchy of recurrent neural nets which are trained (connected up) by the culture and go live by the processes of interaction internally and externally with all the other similar systems in the world (i.e. other people).

list of Links and other references
Back to top of page

return to overall index

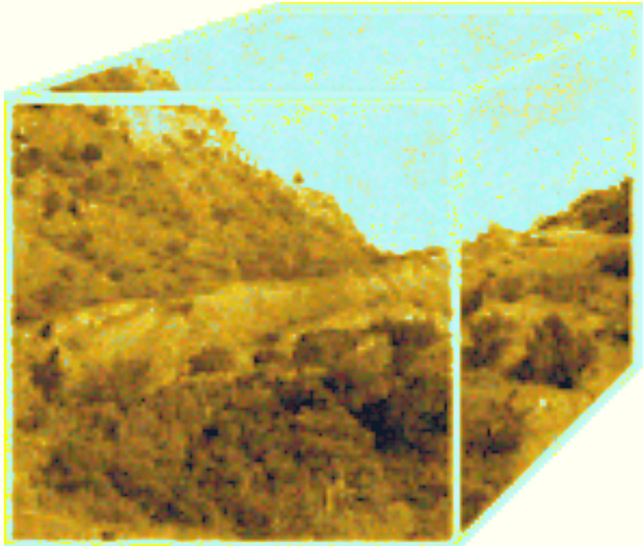
The Classical view of the Mind

Here follows some notes on classical Greek views of the mind and its body.
Part of The Brain Project by Stephen Jones.

Pythagoras

To the Greeks the world was built up from the four elements of the Pythagorean system, fire, air, water and earth. These were linked to the four humours, black bile, yellow bile, phlegm and sanguine. The Pythagoreans had a geometrical conception of the world, presumably derived in some way from the Egyptians' arts of survey.

> "The Pythagoreans erected a system of plane geometry in which were formulated the principle theorems which concern parallels,



triangles, quadrilateral and regular polygonal figures and angles. They discerned many important properties of prime numbers and progressions." (Singer, 1941, pp20-21).

They also reached a concept of irrational numbers, e.g. 'pi' (3.1415...). Their mystical view of the sphere as the perfect figure led to a conception of the earth and the heavenly bodies as spheres.

From the triangle and the square the Greek philosophers developed the concept of the triangular-faced ideal solids as the basic atoms of all things. **Fire was composed of four-faced tetrahedra, earth of six-faced cubes, air of eight-faced octahedra and water of twenty-faced icosohedra.**

Here we find perhaps the earliest known example of the use of technological conceptions to reflect on the nature of the world and the mind.

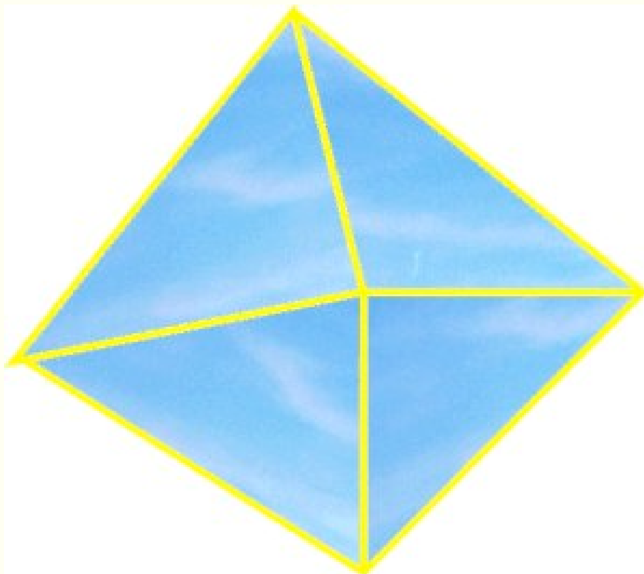
The geometry as a projection of the mind into an ordering process for accurate measurement and division of land, reflects back and provides a basis for thinking about philosophical questions of becoming and essence. As Singer says:

> "The human mind, it must be supposed, is somehow attuned to the processes of nature. We live in a world that is susceptible of mathematical expression." (Singer, 1941, p19).

Plato

Democritus, (a contemporary of Socrates, Plato's teacher) whose theories were opposed by Plato, proposed the idea of the indivisible 'atom'.

> He "not only denied the existence of mind as a separate entity but

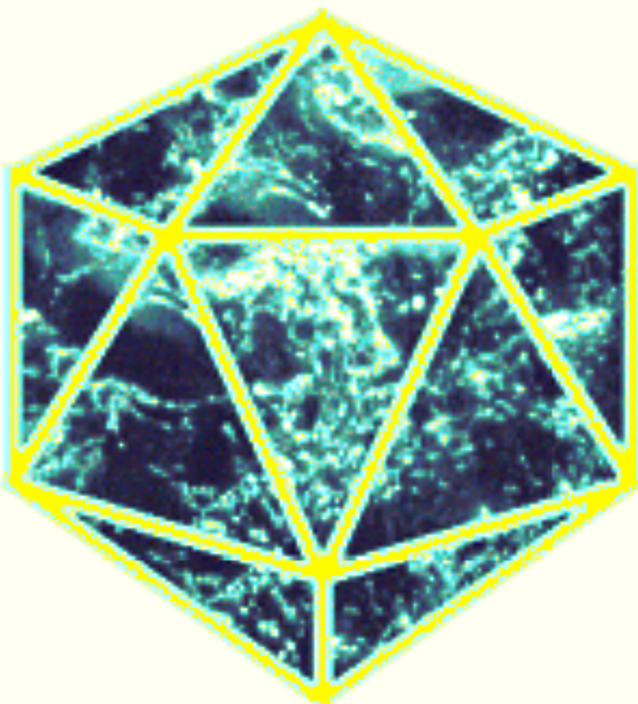


> also assumed the universe to be the result of accident." (Singer, 1941, p33).

Plato carried the Pythagorean system on, (the Pythagorean solids become the Platonic ideal bodies) and with his belief in the perfect sphere of heaven led him to postulate that the rational soul of man "the divinest part of us" must be in the brain, because the head is spherical in shape. He thought of the brain as a gland, and that it produced semen which flowed down through the spinal chord, out through the phallus and into the female vagina, thus rendering women as "flower pots for male seeds", surely a major factor in the exclusion of women from the governance of academic, religious and governmental institutions for 2,000 years or longer.

In the Timaeus, Plato has Timaeus

> "give an account of how the soul moves the body. [shades of Descartes] The soul is



in
movement
and the
body
moves
because it
is
interwoven
with it.
The
Creator
compounded
the
soul-substance
out of the
elements
and
divided it
according
to the
harmonic
numbers
that it
might
have an
innate
perception
of
harmony
and that its
motion
might be
with
movements
well
attuned.
He bent its
straight
line into a
circle.
This he
divided
into seven
circles
(that is the
orbs of the
seven
planets) in
such wise
that the
motions of
the
heavens

are the motions of the soul." (Singer, 1941, p37. from Aristotle's summary in De Anima).

Here is the doctrine of the **macrocosm** and the **microcosm**; as above so below.

Conversely, **Hippocrates**, teaching in another region of Greece, taught:

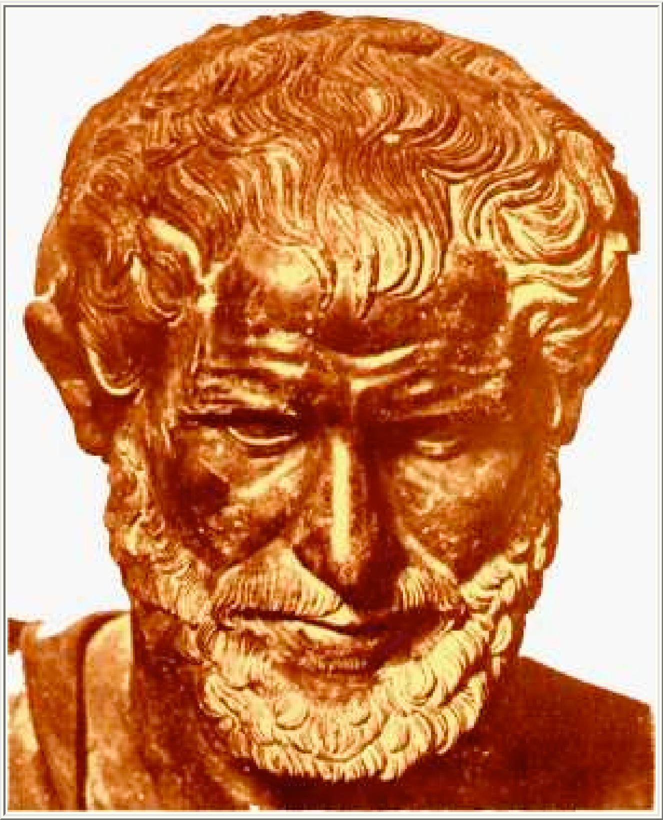
> "Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it...we...think, see, hear and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant."
>
> Note: The first person to locate, on the basis of neuroanatomy, human intelligence in the head was Herophilus of Chalcedon, who flourished around 300 BC. He was also the first to distinguish the motor from the sensory nerves, and performed the most thourough study of brain anatomy attempted until the Renaissance. (Sagan, p13)

Here is the first manifestation of the controversy between the hard, the behavioural science view, the brain does something, it produces semen; and the soft science, or spiritual, view. It is the appearance of the vitalist view: that there is an essence or soul behind everything (Plato's 'ideals'), which governs the development and manifestation of things; and the mechanistic view which suggests the the motions of things can be understood from within those things themselves. The idea of a vital soul being unnecessary to the existence or vitallity of those things.

Aristotle

Aristotle, in De Anima ('On the Soul'), proposed three forms of soul,
1. the vegetative soul possesed by plants in that they grow and decay and enjoy nutriment but hat they do not have motion and sensation,
2. the animal soul

which confirms motion and sensation upon the animals, and 3. the rational soul which is the "conscious and intellectual soul that is peculiar to man." (Singer, 1941, p43). Each higher form possesses in full the attributes of the lower souls so that the human is the sole possessor of the rational soul. (see the discussion of Galen, below)



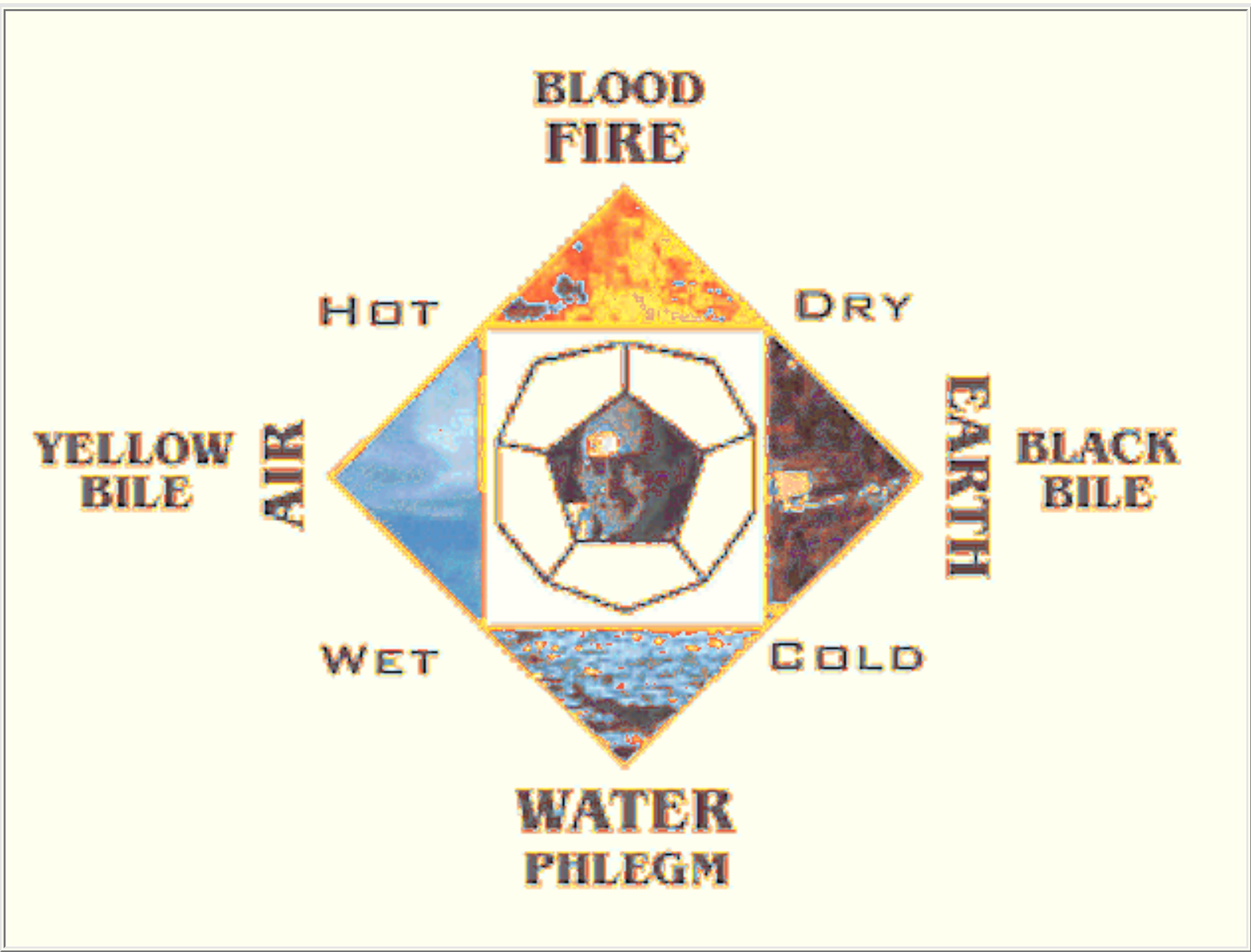
ARISTOTLE

from the **Herculaneum**, 4th Century BC

Aristotle also built his ideas up from the four Platonic solids, (via the four elements manifesting as humours) but added a fifth essence, the 'pneuma' which was carried from the heavens into the lungs and then to the heart, and distributed from the heart to the rest of the body. This 'quintessence' was composed of pentagonal-faced, twelve sided solids called the dodecahedron, (a secret form in the Pythagorean school). In Aristotle's hierarchy of living things, man came at the top and had an extra 'element' which was the body or the fuel of intelligent thought, this was the pneuma and it was evidently (evidence generated by a faulty dissection procedure involving strangulation of the animal) transmitted around the body by the heart. Thus the heart was the seat of the rational soul.

Erasistratus (300-260 BC), an anatomist working a century after Aristotle; on finding three tubular structures going to every organ of the body - an artery, a vein and a nerve - decided that the **pneuma** (quintessence) was conveyed via the carotid arteries from the heart, up to the brain and then flowed from the brain down through the nerves to all organs. The brain was seen as the vessel for distilling the pneuma.

Galen

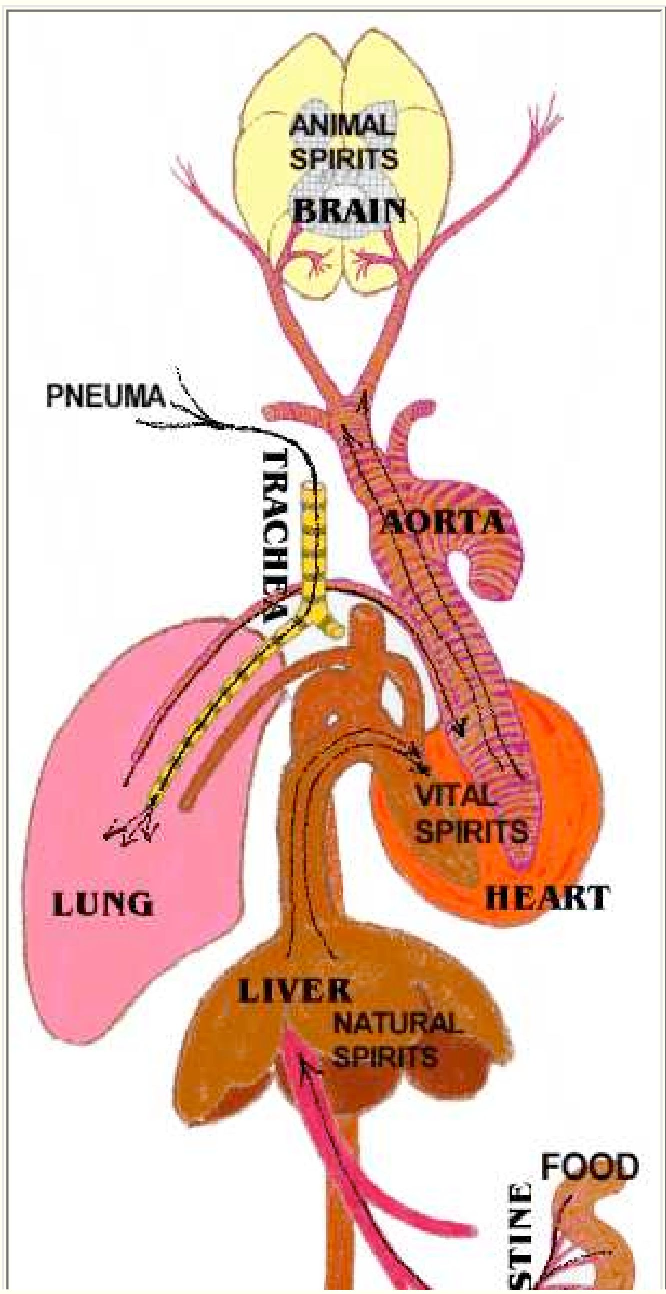


The Four ELEMENTS in association with the four HUMORS and the four QUALITIES

Then comes **Galen** (AD 129-199), whose ideas formed the basis for all medical discussion until virtually the 17th century. Galen summed up all the ideas then current, building up a paradigm which incorporated the atoms of Pythagorus, the four elements of Empedocles, the perfectly spherical divine universe of Plato, the 'primum mobile', the vegetable, animal and rational souls, and the quintessence of Aristotle, the four humours of Empedocles and Hippocrates and Erasistratus' notion that nerves are hollow tubes which carry quintessence from the brain, and added his own discovery of blood in the arteries with the contention that arterial blood was intermixed with quintessence. Galen wrote

"In the universe there are four elements - fire, air, water and earth; and in the living body there are four humours, black bile, yellow bile, sanguine and phlegm. Out of the excess or deficiency or misproportion of these four humours there arise disease; by restoring the correct proportion diseases are cured" (Bergland, p40)

Galen further assigned the three largest organs of the body to be the seat of the three Aristotlean souls; the liver was the seat of the **vegetative soul**, the heart was the seat of the **animal soul** and the brain was the seat of the **rational soul**.



The brain received 'pneuma' from the lungs via the arteries and converted it into animal spirit which it pumped out through the nerves to animate the muscles`

The Pneuma is drawn in from the world, enters the lungs through the trachea and is conveyed to the heart via the pulmonary artery. Food is converted to **natural spirits** in the liver. The natural spirits are then conveyed to the heart via a "hepatic vein" where much of it is cleaned and returned to the lungs but some is mixed with the pneuma and sent to the brain as the **vital spirits**. The function of the brain was to distill the vital spirits still further, converting it to **animal spirits** which was an ethereal substance passed into the (then thought to be) hollow nerves and distributed throughout the body, thus animating the body.

Galen's physiological teaching (after C.Singer *Greek Biology and Greek Medicine*)

> Note: Galen "distinguished three ventricles: one at the front of the brain, divided in two; one in the centre; and one at the rear. He noted that if the substance of the brain was cut at a given point, the animal did not lose consciousness or movement. For this to happen, the section had to penetrate as far as the ventricles. A lesion of the posterior ventricle had the most disturbing effect on the animal. Galen demonstrated that the brain played the central role in controlling bodily and mental activity andtivity originated in the cerebral substance itself...For Plato and Galen the rational soul had its seat in the brain." (Changeux, 1985, p7)

The rational soul had the functions of imagination, reason and memory, and these functions were assigned to the ventricles. Because the function of the brain was to distribute 'animal spirit' throughout the body, it seemed obvious that the fluid filled ventricles should be the major functional units of the brain, rather than the white matter and the grey matter surrounding them.

So, summing up: the brain was the seat of the rational soul. It received vital spirit from the heart, mixed into the 'sanguine' humour (blood). The brain then separates the animal spirit out and stores it in the ventricles, distributing it throughout the body via the nerves. This fluid travelled, via the nerves, to muscles and organs to control all the bodies activities. The rational soul was considered responsible for imagination, reason and memory.

Commentary

It seems to me that we always use the stuff we generate about the world, in trying to understand the world; as a model from which we can gain greater and greater differentiation of the things being considered. The brain works by recognising the differences between sensory data, sameness drops away by being habituated out. The ideas of proportion and ratio, similarity of nature and process in things of different level (as expressed in the macrocosm / microcosm) provide the basis for metaphor and modelling to elucidate nature and thus allowing finer and finer differentiation between things. We use the projections and constructions of our minds to feedback upon ourselves. It is geometry which provides the data of these earliest models of human being.

References:

Bergland, R. The Fabric of mind, 1985

Changeux, J-P. Neuronal Man, 1985,

Singer, C. A Short History of Science, 1941

Singer, C. *Greek Biology and Greek Medicine*, 1922

Sagan, C. Broca's Brain,

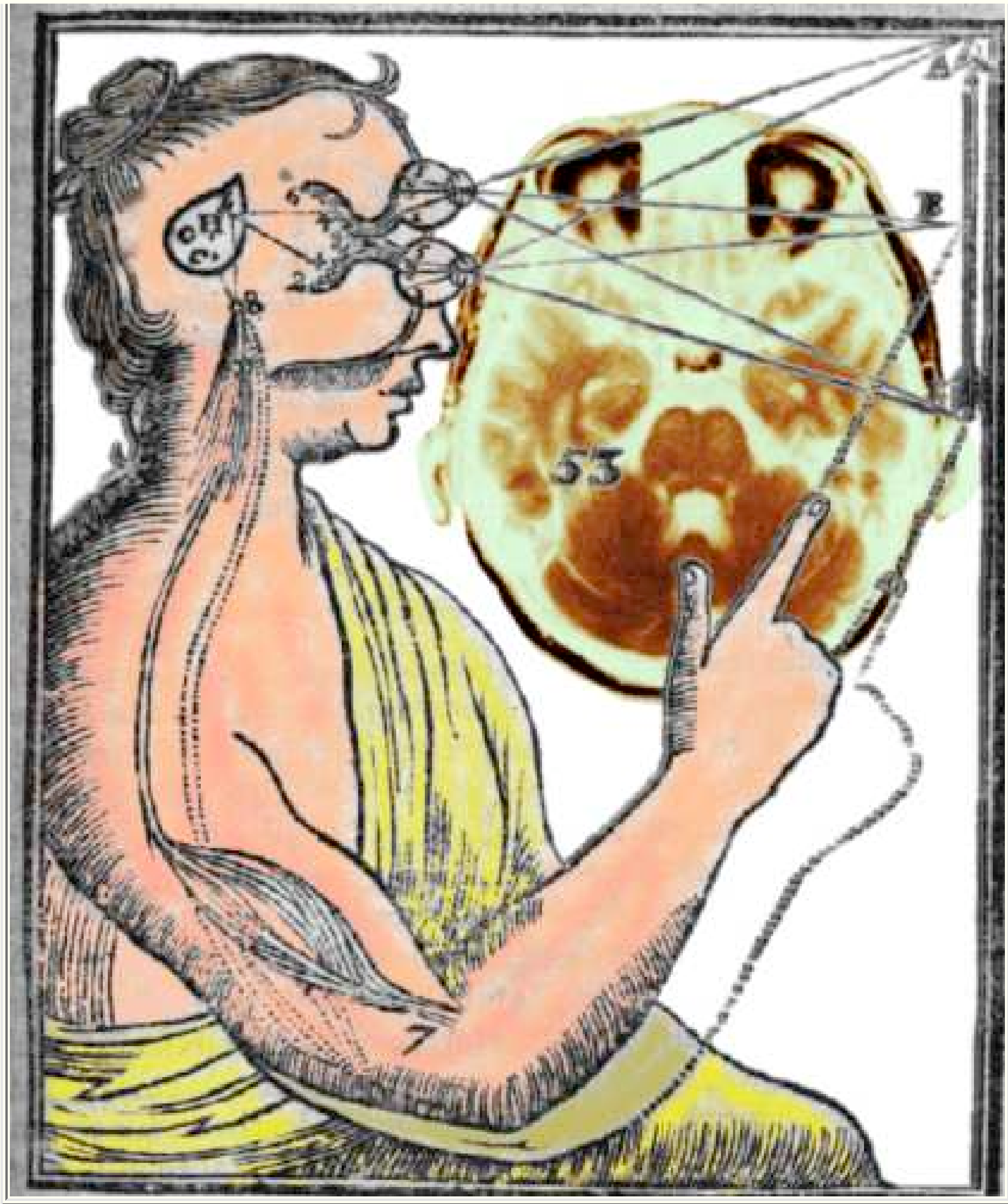
 **return to top level index**	 **return to History index**	 **return to Chapters index**

Rene Descartes

Extracts from Rene Descartes' philosophical analysis of the Mind and the Brain

Part of The Brain Project by Stephen Jones.

The brain, the nerves & the hydraulic analogy



on the hydraulic automaton

It was Descartes who brought about the separation of the mind and body and established dualism as the predominant framework. Descartes developed a view of the animal and human body as an automoton:

> "And as a clock composed of wheels and weights observes not less exactly all the laws of nature when it is ill-made and does not tell the hours as well as when it is entirely to the wish of the workman, so in like manner I regard the human body as

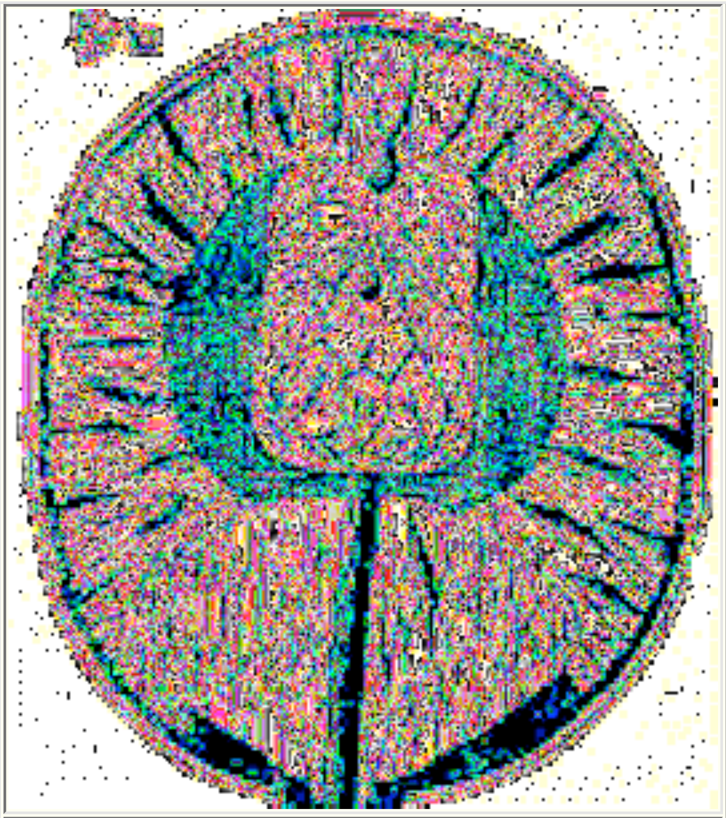
> amachine so built and put together of bone, nerve, muscle, vein, blood and skin, that still, although it had no mind, it would not fail to move in all the same ways as at present, since it does not move by the direction of its will, nor consequently by means of the mind, but only by the arrangement of its organs." [Meditationes, vi].

Descartes saw the brain as an organ of humours, fluids coursing through the nerves which drive the body mechanically. Descartes, having reduced the body to mechanics, has to locate the mind, so for him the soul resides in the pineal gland, a small, single (i.e. untwinned), apparently vestigial body at the base of the brain, behind and between the eyes.

on the motions of the body

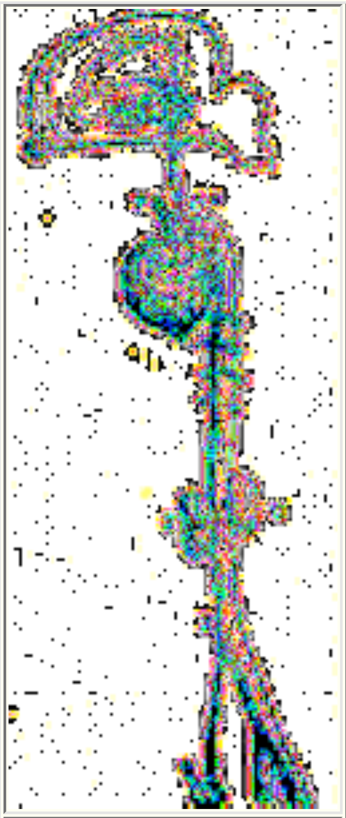
The sanguine spirits enter the cavities of the brain (the ventricles) and from there

> "enter the pores (or conduits) in its substance, and from these conduits proceed to the nerves. And depending on their entering (or their mere tendency to enter) some nerves rather than others, they are able to change the shapes of the muscles into which these nerves are inserted and in this way to



move all the members. Similarly you may have observed in the grottoes and fountains in the gardens of our kings that the force that makes the water leap from its source is able of itself to move divers machines and even make them play certain instruments or pronounce certain words according to the various arrangements of the tubes through which the water is conducted."

"And truly one can well compare the nerves of the machine that I am describing to the tubes of the mechanisms of these fountains, its muscles and tendons to divers other engines and springs which serve to move these mechanisms, its animal spirits to the water which drives them, of which the heart is the source and the brain's cavities the water main. Moreover breathing and other such actions which are ordinary and natural to it, and which depend on the flow of the spirits, are like the movements of the clock or the mill which the ordinary flow of the water can render continuous. External objects which merely by their presence

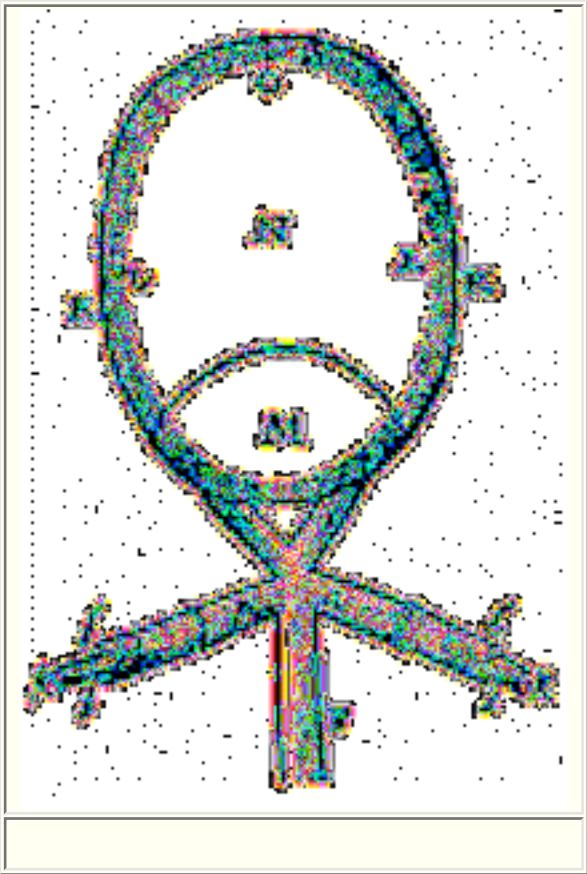


act on the organs of sense and by this means force them to move in several different ways, depending on how the different parts of the brain are arranged, are like strangers who, on entering some of the grottoes of these fountains, unwittingly cause the movements that then occur, since they cannot enter without stepping on certain tiles which are so arranged that, for example, if they approach a Diana bathing they will cause her to hide in the reeds; and if they pass farther to pursue her they will cause a Neptune to advance and menace them with his trident; or if they go in another direction they will make a marine monster come out and spew water into their faces, or other such things according to the whims of the engineers who made them. And finally when there shall be a rational soul in this machine, it will have its chief seat in the brain and will reside there like the turncock who must be in the main to which all the tubes of these machines repair when he wishes to excite, prevent, or in some manner alter their movements." [Descartes: Traite de l'Homme (1664) (T.S.Hall transl.)]

on the sensations of the body

"To understand, next, how external objects that strike the sense organs can incite [the machine] to move its members in a thousand different ways: think that [a] the filaments (I have already often

told you that these come from the innermost part of the brain and compose the marrow of the nerves) are so arranged in every organ of sense that they can be very easily moved by the objects of that sense and that [b] when they are moved, with however little force, they simultaneously pull the parts of the brain from which they come, and by this means open the entrances to certain pores in the internal surface of this brain; [and that] [c] the animal spirits in its cavities begin immediately to make their way through these pores into the nerves, and so into muscles that give rise to movements in this machine quite similar to [the movements] to which we are naturally incited when our senses are similarly impinged upon.



that the mechanical view is the only one necessary

"I say, that these functions imitate those of a real man as perfectly as possible and that they follow naturally in this machine entirely from the disposition of the organs - no more nor less than the movements of a clock or other automaton, from the arrangement of its counterweights and wheels. Wherefore it is not necessary, on their account, to conceive of any vegetative or sensitive soul or any other principle of movement and life than its blood and its spirits, agitated by the heat of the fire which burns continually in its heart and which is of no other nature

than all those fires
that occur in
inanimate bodies."

on the difference between man and machine

Descartes refers to those who might be acquainted with the automata and, in respect of the human body, suggests

> "Such persons will look upon this body as a machine made by the hands of God, which is incomparably better arranged, and adequate to movements more admirable than any machine of human invention. Were there such machines exactly resembling in organs and outward form an ape or any other irrational animal, we could have no means of knowing that they were in any respect of a different nature from these animals; but if there were machines bearing the image of our bodies, and capable of imitating our actions as far as it is morally possible, there would still remain two most certain tests whereby to know that they were not therefore really men. Of these the first is that they could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others. The second test is, that although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it would be discovered that they did not act from knowledge, but solely from the disposition of their organs. Again, by means of these two tests we may know the difference between men and brutes." [shades of the Turing test] Descartes, Discourse on Method (Part V) (1637) (transl. J Veitch).

on the cogito, the thinking thing

Having carried out his radical doubting of everything, Descartes seeks out one point which may be seen as being certain.

> "I shall proceed by setting aside all that in which the least doubt could be supposed to exist. I suppose, then, that all the things that I see are false; I persuade myself that nothing has ever existed of all that my fallacious memory represents to me. I consider that I possess no senses; I imagine that body, figure, extension, movement and place are but the fictions of my mind. What then can be esteemed as true? I was persuaded that there was nothing in all the world. Was I not then likewise persuaded that I did not exist? Not at all; of a surety I myself did exist because I persuaded myself of something. But [if] there is some deceiver or other, very powerful and very cunning, who ever employs his ingenuity in deceiving me, then without doubt I exist also if he deceives me, and let him deceive me as much as he will, he can never cause me to be nothing so long as I think I am something. I am, I exist, is necessarily true each time I pronounce it, or that I mentally conceive it."

What attributes can Descartes have without requiring a body in which he can have no certainty? Obviously not any of the movements or sensations of bodies, but

> "What of thinking? I find here that thought is an attribute that belongs to me; it alone cannot be separated from me. I am, I exist, that is certain. But what [kind of] thing [am I]?...a thing which thinks."

But what of these other 'uncertain' things? Bodies and sensations and the like.

> "...solely by the faculty of judgement which rests in my mind, [do] I comprehend that which I believed I saw with my eyes. [Whatever] error may still be found in my judgement, I can nevertheless not perceive it thus without a human mind. Bodies are not, properly speaking, known by the senses or by the faculty of imagination, but by the understanding only, and since they are not known from the fact that they are seen or touched, but only because they are understood, I see clearly that there is nothing which is easier for me to know than my mind." Descartes: Meditations on the First Philosophy.

on the rational soul

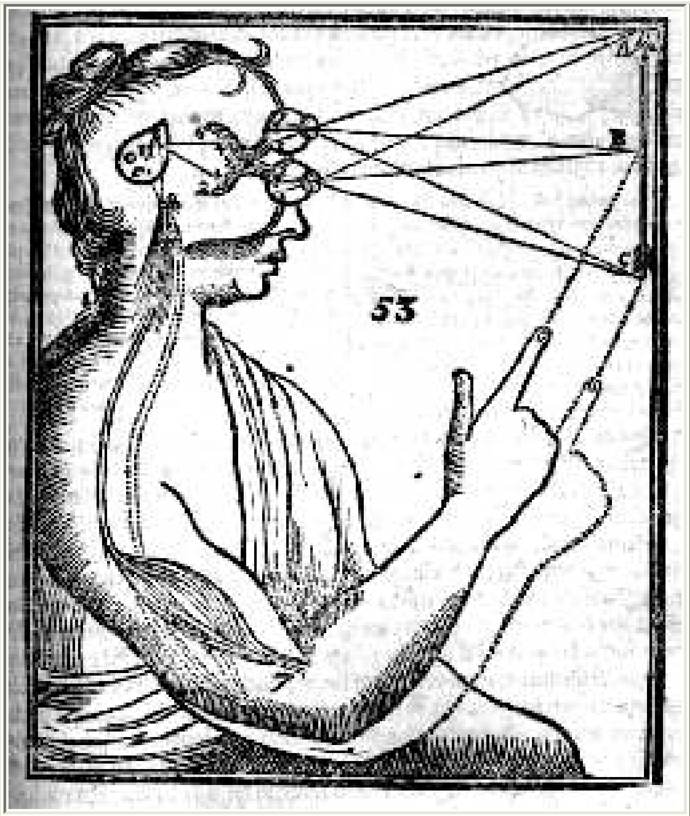
In reference to the rational soul Descartes declares

> "I had...described the reasonable soul, and shown that it could by no means be educed from the power of matter...but that it must be expressly created; and that it is not sufficient that it be lodged in the human body exactly like a pilot in a ship, unless perhaps to move its members, but that it is necessary for it to be joined and united more closely to the body, in order to have sensations and appetites similar to ours, and thus constitute a real man."

on images vs encoding in the brain

Descartes establishes that we do not see by the direct transmission of images to the brain but of a coded version of the image;

> "...you must conceive the nature of these images quite differently...for since [the philosophers] have no notion of the images except that they must be like the objects they represent, they cannot possibly explain how they can be produced by these objects, and received by the external sense-organs, and transmitted by the nerves to the



brain. Their sole reason for the assumption is that they have noticed that a picture readily induces us to think of the object depicted, and have thus thought we must be led to conceive of the objects that affect our senses by tiny pictures formed within our head. But we have to consider that thought may be induced by many things besides pictures - e.g. by signs and words, which in no way resemble the things signified." Descartes: The Dioptrics (transl. Anscombe & Geach)

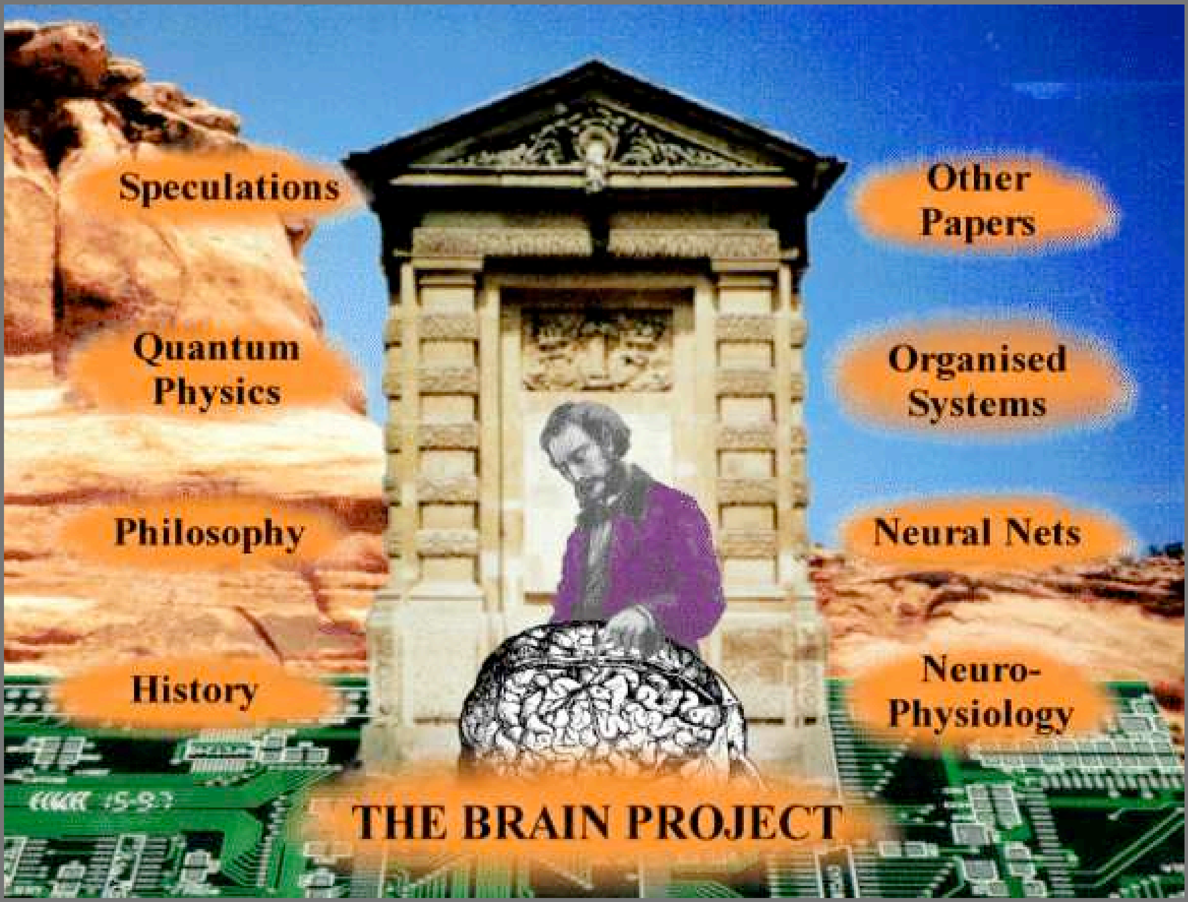
for a full text of The Discourse on the Method of Rightly Conducting the Reason, and Seeking Truth in the Sciences by Rene Descartes.

 **return to Philosophy index**	 **return to top level index**	 **return to History index**	 **return to Chapters index**

The BRAIN PROJECT

Return to
Top Index

Return to
Historical Background Index



The BRAIN PROJECT

Return to
Top Index

Return to
Chapters Index

Early Neurophysiology

Some notes on the development of early Neuro-anatomy and Neuro-physiology with special reference to the localisation of function.
Part of **The Brain Project** by Stephen Jones.

The beginnings

Available technology and procedures have always constrained the gathering of information and the development of theories. Cultural conditions and the frameworks of thought, coupled with political restrictions mould the way we think about things and modulate the way we look at matters of investigation, gather data and develop theories. For the Greeks and medievals anatomy was not readily acceptable and only rarely was it possible to do it on human bodies. Most of the work was done on animals. The Church froze real investigation until the rennaiscance and for long after, so that it took a difficult and slow path to get past the animal-spirits/pneuma model of the operation of the brain. Thus Descartes reproduces Galen's system within the medium of anatomy, i.e. he supplies a mechanical system of tubes and fluids as an analogy for the arteries, veins and nerves leading to and from every organ, within which to carry and distribute the Galenical humors and spirits. But realising that this mechanical description alienates the idea of the soul (which would create a serious problem for him with the Church) and also recognising that we do have this 'mind', Descartes constructs a dualistic system of mind as an ethereal thing connecting into and controlling the body via the pineal gland. Descartes separated the soul from the mechanical body, linking them via the pineal gland.

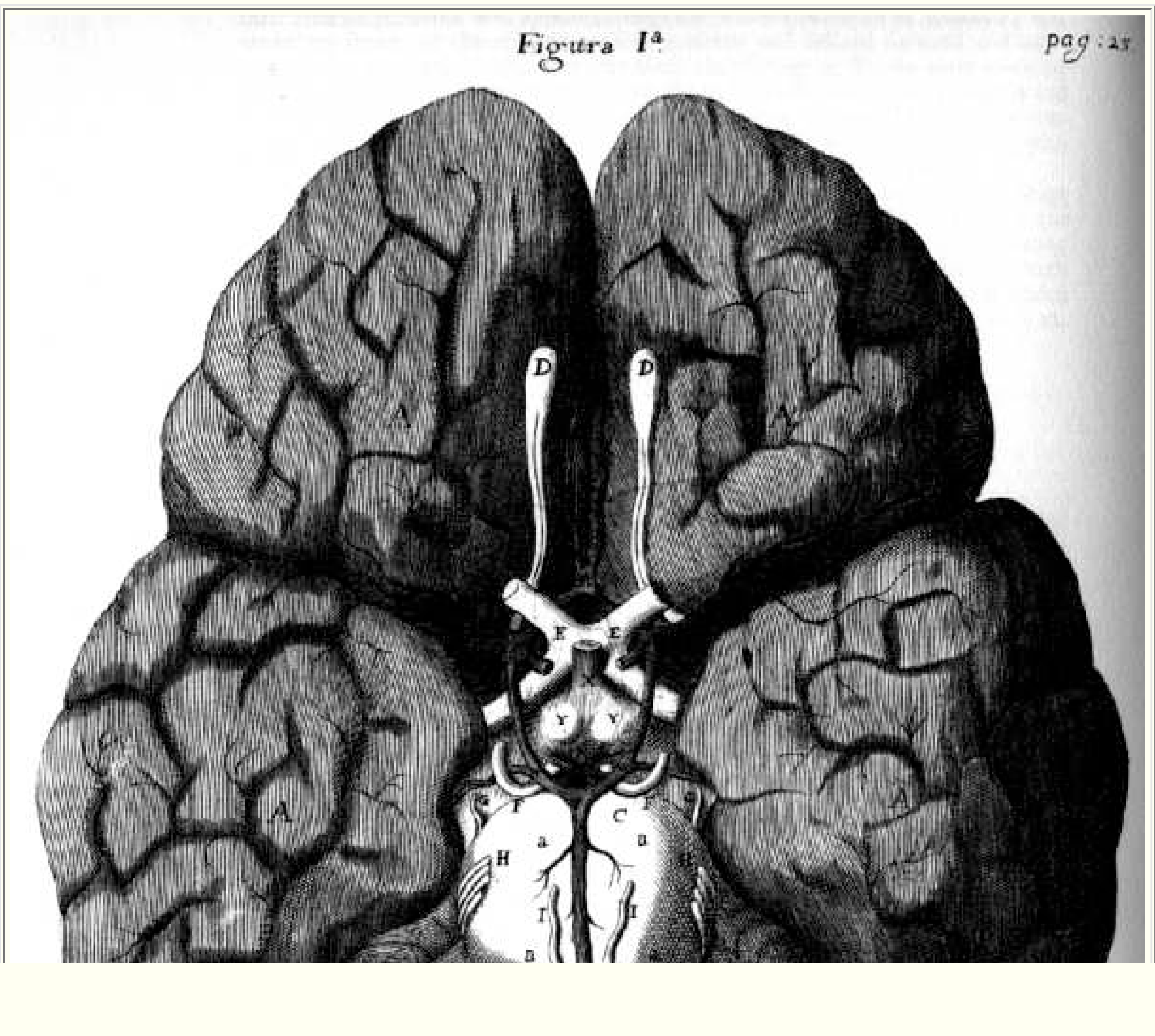
In terms of setting the philosophical agenda, this dualistic system, this separation of the mind from the body, could not have been a worse error. It completely misguided the development of a philosophy of consciousness as being somehow not part of the otherwise mechanistic view of the human body and left the mind as something which even now is often considered to be too hard a problem for us to tackle. The idea of a 'soul' or 'spirit' or, as it became in physics, an 'ether' always provided a sort of escape clause whenever the religious element was threatened by the discoveries of scientists right up to the 20th century. One wonders if the desparate need to maintain control over the mental framework of the culture isn't still as strong as it's always been.

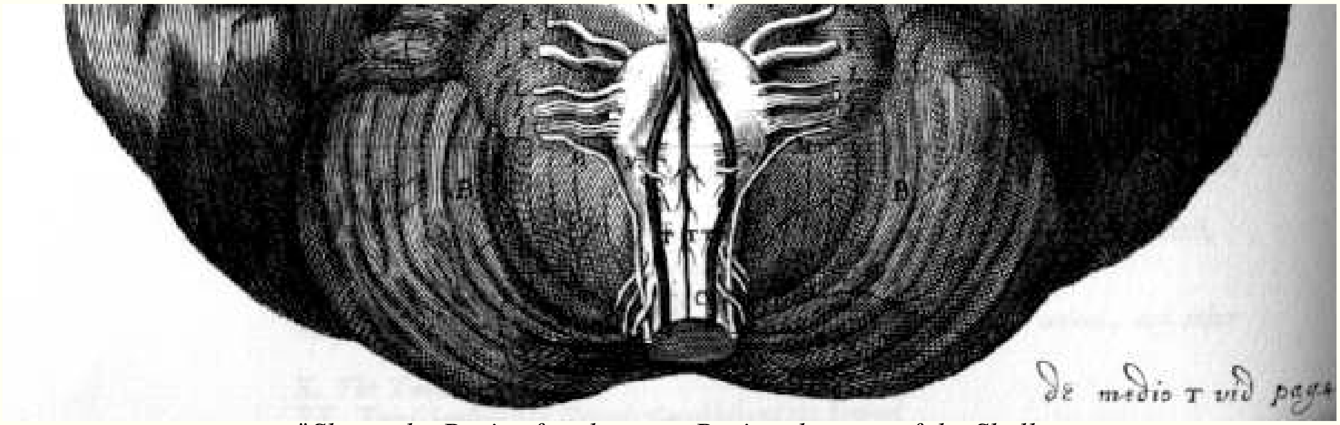
The invention of the microscope allowed the nerve to be distinguished and the emphasis of ideas on the seat of the mind, the 'rational soul', was shifted to the grey and white matters, for they were found to be made of the same matter as the nerves.

Thomas Willis

The English anatomist, Thomas Willis (1621-1675),

> "produced the best pictures of the brain so far. He showed that the folds of the cerebral cortex covered a certain number of subcortical centers, such as the striatum, the thalamus and the corpus callosum that unites the two hemispheres. He distinguished a cortical 'grey matter', responsible, in his mind, for animal spirits, from a deeper 'white matter', distributing the spirits to the rest of the organism, to which they give sensation and movement...[But] Willis still accepted the idea of an immaterial, reasoning soul unique to man, somewhere beyond the blade of his scalpel." [Changeux, 1985, p11].





"*Shews the Basis of an humane Brain taken out of the Skull, with the Roots of the Vessels cut off*."
from **The Anatomy of the Brain** by Thomas Willis, M.D. London, 1681.

Thomas Willis assigned the function of linking the mind to the brain to the **striatum**. He was also one of the last of the animal spirit theorists (along with Descartes), and carried out experiments showing that if the blood was prevented from reaching the brain then

> "nerve function ceased because vital spirits could not reach the ventricles for conversion into the essential animal spirits." [Stevens, p13].

Willis summarised the workings of the animal spirits:

> "...the animal Spirits are procreated only in the Brain and Cerebel, from which they continually springing forth, inspire and fill full the medullar Trunk: (like the Chest of a musical Organ, which receives the wind to be blown into all the Pipes) but those Spirits being carried from thence into the Nerves, as into so many Pipes hanging to the same, blow them up and actuate them with a full influence; then what flow over or abound from the Nerves, enter the Fibres dispersed every where in the Membranes, Muscles, and other parts, and so impart to those bodies, in which the nervous Fibres are interwoven, a motive and sensitive or feeling force. And these Spirits of every part are called Implanted, forasmuch as they flow not within the Nerves, as the former, with a perpetual flood; but being something more stable and constant, stay longer in the subject bodies; and only as occasion serves, *viz*. according to the impressions inwrdly received from the Nerves, or impressed outwardly by the objects, are ordained into divers stretchings or carryings out for the effecting of motion or sense either of this or that manner or kind. [Willis, 1681, p126]

and seems to be saying that the animal spirits flood the system of the nerves and muscles. By pressures outwardly placed on the flood motion is effected and by inward pressure sensation is received. (see the notes on Descartes: the extract on sensations; for his version of this idea).

Elsewhere Willis speaks of the motions of the muscles:

> "Therefore as to the Musculary Motion in general, we shall conclude after this manner, with a sufficiently probable conjecture, that the animal Spirits being brought from the head by the passage of the Nerves to every Muscle (and as it is very likely), received from the membranaceous fibrils, are carried by their passage into the tendinous fibres, and there they are plentifully laid up as in fit Storehouses; which Spirits as they are naturally nimble and elastick, where ever they may, and are permitted, expanding themselves, leap into the fleshy fibres; then the force being finished, presently sinking down, they slide back into the Tendons,..." [Willis, in Stevens, p13]

He seems to find some problem with the idea of this spirit being a fluid when he considers the speed with which sensations and muscular actions occur and so he likens the animal spirit to rays of light filling the passages of the nerves and muscles

> "...the animal Spirits flowing from the Medullar substance into the nerves, are as it were rays diffused from the light it self, and the other Spirits every where abounding in the Fibres, are as so many lucid particles included and implanted in the Air, which are actuated by the former, and being stirred up by them into motion, perform the acts both of the sensitive and locomotive Faculty." [Willis, 1681, p126]

Nevertheless, Willis distinguished some of the subcortical centres, such as the thalamus and the striatum. He showed the corpus callosum linking the two hemispheres, and he distinguished the white matter of the deeper areas from the grey matter of the cortex. He also dilineated eight major pairs of nerves at the base of the brain, including the optic nerves, the smell nerves and the nerves which control the movements of the eyes.

But the vital spirits, though they begin to have a more subtle composition now, were still distilled in the cortex and distrbuted through the body, to which they gave sensation and movement, by the subcortical layers.

The discovery of electricity wrought the first real change in the concept of 'animal spirit', but even for Galvani the system was still an hydraulic one, though the fluid was of a rather finer nature.

Pierre Gassendi "at the beginning of the 17th century ... announced... that animals also have a memory. They can reason and possess other psychological characteristics similar to those of man; therefore, they must have a soul. What is more, for Gassendi, the soul is not situated in any part of the body." [Changeux, pp12- 13].

Then in the 18th Cent. the 'dualism' of Descartes was challenged when Julien Offray de la Mettrie wrote, "that it was quite possible to remove the soul from Cartesian theories without losing much, and that man himself could be put in the category of [the] mechanical animals [the product of a fad for 'automata']. For Pierre Cabanis, the "brain secretes thought as the liver secretes bile." The theory of the immaterial nature of the soul disappeared progressively from works devoted to the brain sciences." [Changeux, p13]

Meanwhile, the English philosopher, John Locke, developed a theory "according to which mental faculties and instincts derived sensations." [Changeux, p14].

Physiognomy / Phrenology

Meanwhile, the anatomists are trying to analyse the activities of the brain into functional areas. "Different parts of the brain were now seen as controlling or analysing different body functions, though 'higher mental activities' were still without a clear-cut home." [Rose, p 41].

We have the folly of Phrenology, which "claimed to be able to distinguish, through the skull, separate brain regions for each of a large number of human faculties." [Rose, p 40]. But we also have the beginnings of modern neurophysiology.

Franz Gall (Austrian) speculated in the 19th century,

> "that the cerebral cortex represented the highest level of the brain and that the development of this area characterised mammals and man."

He also noticed that

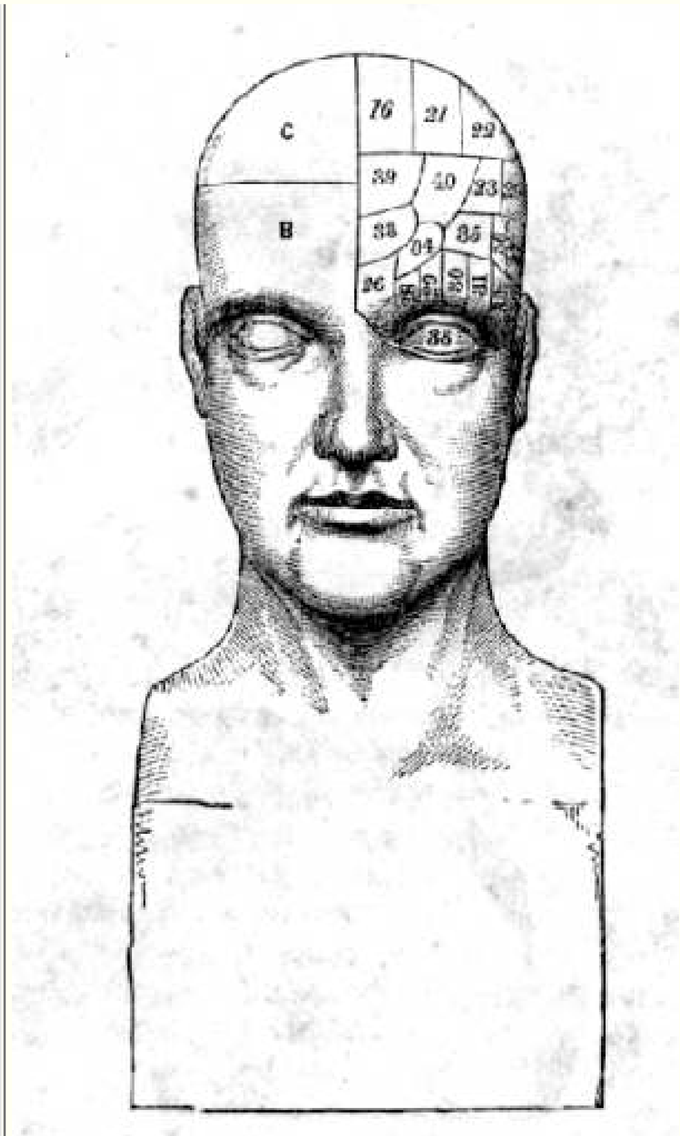


Diagram 1.—Front View of the Model Head.

the cortex was made of the same matter as those nerves in direct contact with the organs. Gall's intention was to analyse and localise cerebral functions from a physiological point of view. For Gall, humans possessed innate, essential and irreducible "moral and intellectual faculties" [Gall].

"Drawing on everday language, the biographies of famous men, and descriptions of mental disturbances and manias", he generated a list of twenty seven faculties that he said "represented the exaggeration of one or other of these faculties."

The sex drive, maternal behaviour, aggressiveness, verbal memory, understanding of words and the sense of location and spatial relations featured on this list. Gall collected the skulls of criminals and busts of famous men. He postulated that the suface of the skull would mirror the exaggeration of functional areas in the cortex so that he could correlate bumps in the skull with faculties prominent in certain individuals and created a map locating particular mental functions in particular areas of the cortex. This became Phrenology, the mapping of the bumps on a person's head would give indications of the capabilities and personality of the individual. [Changeux, p14].

Gall was a 'materialist', abandoning the idea that mental activity lay in an immaterial soul. For this, he was persecuted by the Church, hounded out of Vienna and exiled to Paris.

Ablation experiments

Gall was seriously challenged by Paul Flourens, (French), who developed 'ablation' (or surgical removal) as a procedure. By removing anatomically defined areas of the cortex of an animal, and watching its behaviour afterwards, he could localise certain functions. For example, by removal of the cerebellum, the animal lost the ability to coordinate movements; or "discrete lesions in the medulla, near the posterior ventricle, disturbed certain vital functions, such as respiration." [Changeux, p16]. Flourens wrote:

> "a large section of the cerebral lobes can be removed without loss of function. As more is removed all functions weaken and gradually disappear. Thus the cerebral lobes operate in unison for the full exercise of their functions." [Flourens].

> "The cerebral cortex functioned as an indivisible whole ... [housing] an "essentially single" faculty of perception, judgement and will...the last refuge of the soul." [Flourens in Changeux, p17].

But Flourens' experimental technique was not very accurate, he experimented on birds or lower vertebrates (in which the cortex is not particularly differentiated) and his behavioural studies were too rudimentary for an adequate investigation of the faculties catalogued by Gall. Better anatomical work by Francois Leuret and Pierre Gratiolet mapped the folds and fissures of the cerebral cortex, and demarcated and named the frontal, temporal, parietal and occipital lobes.

A new procedure using "natural experiments" such as cranial damage and lesions began to provide a means of studying the localisation of faculties in humans. In Paris in 1861, Paul Broca presented the case of a man who, while appearing to have all his faculties otherwise, possessed no ability to speak. On the man's death, Broca performed an autopsy, revealing

> "a lesion situated principally in the middle part of the frontal lobe of the left hemisphere...By establishing rigorous correlations between anatomical and behavioural features, Broca gave the first demonstration of the discrete cortical localisation of well- defined faculties." [Changeux, p19].

A lesion of the left frontal lobe of the left hemisphere will cause a loss of the language faculty, known as `aphasia'.

In 1909, Korbinian Brodmann, using data from both monkeys and humans, established fifty-two areas in the brain which showed some sort of functional identification, especially sensory projection areas and motor areas.

But the principle of cerebral localisation of function must not be applied in too narrow a way.

> "For Hughlings Jackson...the more a process was complex and voluntary, the more it mobilised multiple cerebral areas. A cortical lesion disorganised an ordered sequence of

physiological processes rather than destroying a cortical centre". [the idea of a centre should be replaced by the idea of a 'preferential integration focus'] [Changeux, p21].

The Clockwork model

Descartes' and Willis' hydraulic models of the brain and the nerves though differing in detail were derived from the hydraulically driven animal models of the waterworks at Versailles and other pleasure gardens. These technological examples provided a descriptive with which those writers could explain what they were discovering. Waterwork and later clockwork automata were very much the fad of the times and Vaucanson, the master automata builder, even suggested an "artificial man". Although the empirical philosophers of the 18th and 19th centuries expanded the mind into the whole of the brain, nevertheless a mechanical, clockwork view of the brain retained, driving the body through hydraulic distribution of fluids throught the nerves which were seen as tubes through which the humours flowed. (for further material on Automata see: Neural Networks and the Computational Brain)

References:

Changeux, J-P. Neuronal Man 1985

Flourens, P. Experimental Researches on the Properties and the Functions of the Nervous System in the Vertebrate Animals. Paris, 1824.

Gall, F. Sur les fonctions du cerveau et sur celles de chacune de ses parties. Paris, 1822-25.

Rose, Steven. The Conscious Brain, London,1973

Stevens, L.A. Explorers of the Brain,

Willis, T. The anatomy of the Brain. London, 1681

 **return to top level index**	 **return to History index**	 **return to Chapters index**

Neural Networks and the Computational Brain

or Matters relating to Artificial Intelligence

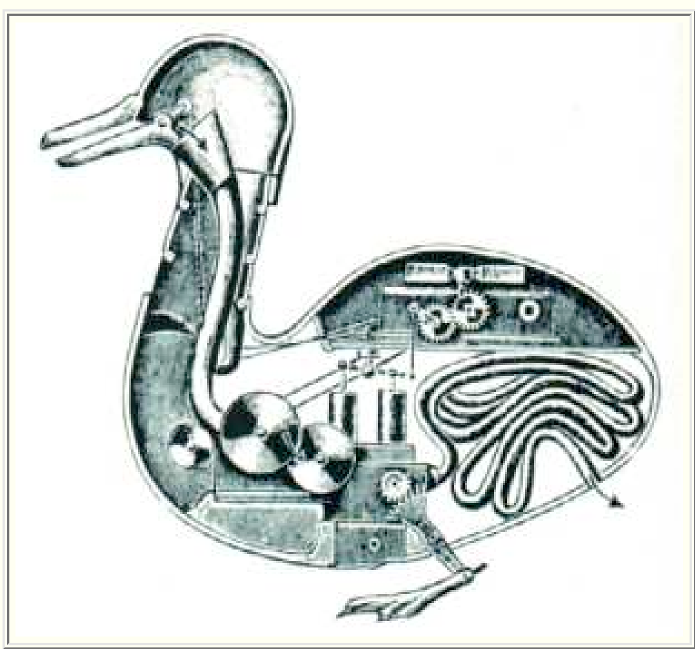
by Stephen Jones

Automata

The idea that we might be able to produce an artificial intelligence or perhaps even a conscious machine has had a long history. Harking back to the time of Descartes there was a great deal of activity in producing hydraulic automata for the pleasure gardens of the wealthy. These were hydraulic devices which, for example, might respond to a person stepping on a specially constructed flagstone in a garden pathway by triggering a cupid sculpture to spray water over that person.

Using hydraulic and clockwork models many automata were produced emulating in some way the activities of animals or humans. Many of the mechanical devices of the 17th and 18th centuries echoed aspects of human and animal motion and behaviour. The mechanistic view of the world developed greatly and natural philosophers felt that all human behaviour could be explained by mechanical models. In 1680 an Italian, and student of Galileo's, Giovanni Borelli, published De Motu Animalium (On the motion of animals) a study of the mechanical action of the muscles. In France in 1748 de la Mettrie's L'Homme Machine (Man a Machine) was published in which he claimed that all human behavior including the mind had mechanical explanation. This work was burned as atheistic and is still considered by historians of science as unecessarily extreme. [for example see C. Singer A Short History of Biology, 1931, p357] One should note that there was also a great deal of opposition in some academic quarters to this mechanistic view which was expressed under the framework of 'vitalism'.

In the same period Vaucanson produced a number of quite successful toys which emulated some activity or another of an animal or bird. Sir David Brewster in his book Letters on Natural Magic provides a description of Vaucanson's duck:



It "exactly resembled the living animal in size and appearance. It executed accurately all its movements and gestures, it ate and drank with avidity, performed all the quick motions of the head and throat which are peculiar to the living animal, and like it, it muddled the water which it drank with its bill. It produced also the sound of quacking in the most natural manner. In the anatomical structure of the duck, the artist exhibited the highest skill. Every bone in the real duck had its representative in the automaton, and its wings were anatomically exact. Every cavity, apophysis, and curvature was imitated, and each bone executed its proper movements. When corn was thrown down before it, the duck stretched out its neck to pick it up, it swallowed it, digested it, and discharged it, in a digested condition. The process of digestion was effected by chemical solution, and not by trituration, and the food digested in the stomach was conveyed away by tubes to the place of its discharge." [Brewster, 1868, p321]

The possibility of the automaton has enticed engineers in the western world for many centuries providing many an exhibit at fairs and expositions and as a feature of tales and novels from the Golem to Frankenstein. The robot workers of Karel Capek's R.U.R and Fritz Lang's Maria in Metropolis provide memorable 20th century examples.

Neural Network theory & non-reducibility of brain operation to the neuron

Research into potential systems of artificial intelligence now looks to the brain for models rather than looking to technology for ideas from which to model the brain. A number of scientists are looking at the development of artificial intelligence from the basis of a developing understanding of the architecture of the human brain. This work is now represented in two interlocking disciplines: Computational neurobiology: which involves understanding human/animal brains using computational models; and Neural Computing: or simulating and building a machine to emulate the real brain. The analysis is made on two levels: coarse grained, examining and elucidating networks of interacting subsystems which is largely a neurophysiological activity; and fine grained, building theories and models of actual artificial neural networks as subsystems.

By the 40's enough work had been done on describing the behaviour of the neuron for psychologists and mathematicians to make a serious attempt at a mathematical theory of the neuron, both natural and artificial.

The artificial neuron

The original neural network was based on work by Warren McCulloch and Walter Pitts published in 1943. They built up a logical calculus of sequences of nerve connections based on the point that a nerves' action potential only fires in an all-or-none manner if the treshold for that nerve has been exceeded.

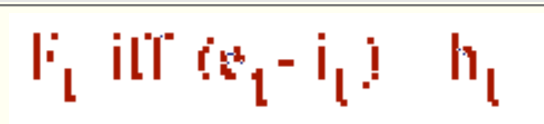
They produced an artificial logical neuron network consisting of three kinds of neurons

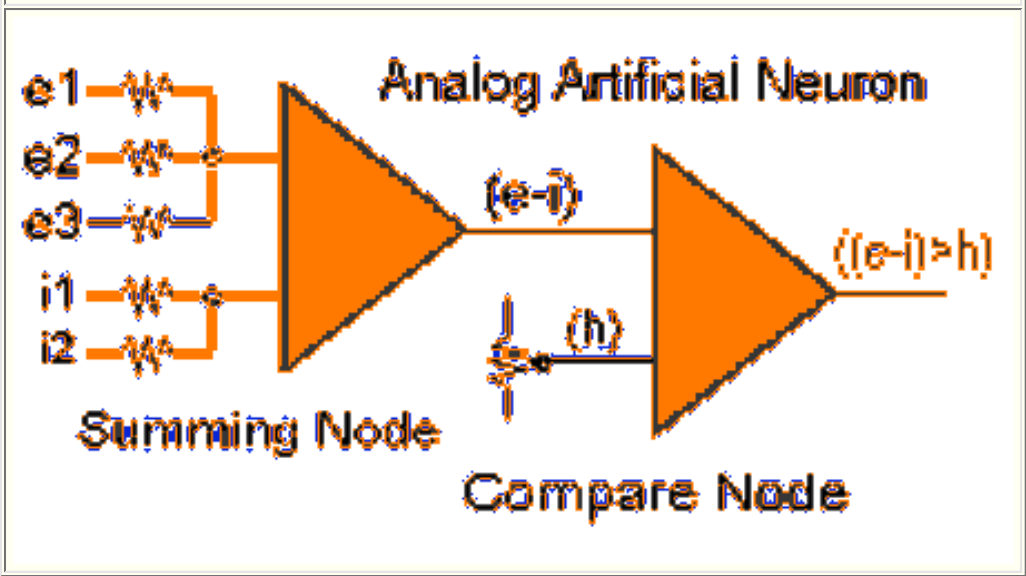
1. Receptor, afferent or input neurons which receive the impulse to fire from a sensor.
2. Central or inner neurons which are synapsed onto from receptor and other neurons and synapse onto output and other neurons.
3. Effector neurons which receive impulses from both inner neurons and directly from receptors.

They described a set of rules for the operation of the neurons:

1. Propagation delay is assumed to be constant for all neurons,
2. Neurons fire at discrete moments, not continuously.
3. Each synapse output stage impinges onto only one synaptic input stage on a subsequent neuron.
4. Each neuron can have a number of input synaptic stages.
5. Synaptic input stages contribute to overcoming of a threshold below which the neuron will not fire.

An artificial neuron is set up to fire at any time **t** if and only if **(e-i)** exceeds **h**,
where **e** is the number of excitatory synapses onto it at time t, **i** is the number of inhibitory ones and **h** is the firing threshold for that neuron.





Given a clearly defined set of input and output conditions it is possible to create an arbitrarily complex neural network from the three types of neurons, with appropriate thresholds at the various synapses of the network.

Compared with biological neurons

McCulloch and Pitts suggested that this network may as well describe the functioning of a human nervous system as much as it might describe an automaton. Nevertheless, the whole system is deterministic. The network is a scanning device which reads the input to output transform specification as if it were a dictionary, the 'meaning' of every possible input 'word' is determined by the dictionary of associated inputs and outputs in its repertoire.

> "Given any finite dictionary of input stimuli and their associated meanings or output responses, we can...always make (on paper) a scanning device or neural network capable of consulting the dictionary and producing the listed meaning or response for each input 'word' denoting its associated stimulus." [Singh, 1965, p158]

The associations of input to output are altered by altering the pattern of interconnections between neurons of each layer.

This is really a look-up-table device using neurons to carry out logic hardware functions, all its input and output are predetermined, for each set of possible inputs and interconnections there is a fixed result. Obviously human intelligence is not so fixed, and there will always be shortfalls in any strictly defined neural system. Active human neural systems learn and adapt to the culture in which they grow, so the McCulloch and Pitts neuron is inadequate to describe what is really going on, but networks starting at this level can be set up to learn and adapt.

Jagjit Singh in his textbook on information theory speaks of the potential behaviour repertoire of natural neural systems as being impossible to reduce adequately to unambiguous description:

> "Whether any existing mode of behaviour such as that of the natural automata like the living brains of animals can really be put 'completely and unambiguously' into words is altogether a different matter...Consider, for instance, one specific function of the brain among the millions it performs during the course of its life, the visual identification of analogous geometrical patterns. Any attempt at an 'unambiguous and complete' verbal description of the general concept of analogy, the basis of our visual faculty, will inevitably be too long to be of much use for even drawing (on paper) neuron networks having the wide diversity of visual responses the natural automata normally exhibit as a matter of course. No one in our present state of knowledge dare hazard a guess whether such an enterprise would require thousands or millions or any larger number of volumes. Faced with such a descriptive avalanche, one is tempted to substitute the deed for the description, treating the connection pattern of the visual brain itself as the simplest definition or 'description' of the visual analogy principle." [Singh, 1965, pp171-2]

These neural networks are essentially digital, computer-like models having profound differences from real neural systems. For example, in real neural systems the pulse trains carrying quantitative sensory information seem to be coded in pulse frequency modulation form, rather than digital representations of number; also the depth of connectionism seems to be much more efficient in our neural operations. That is, the number of layers of neurons: sensory input, processing and output (efferent) layers; is much less than appears necessary with artificial neural nets.

McCulloch and Pitts also spoke of neuron nets having circular interconnections in which "activity may be set up in a circuit and continue reverberating around it for an indefinite period of time, so that any realisable (result) may involve reference to past events of an indefinite degree of remoteness." [McCulloch & Pitts, 1943] thus producing a regenerative process which might be akin to learning and to memory.

In considering the differences between biological systems and automata von Neumann examined the problem of self-reproducing machines. He discerned that in systems below a certain level of complexity the product of those systems would always be less complex than the system itself, but with a sufficient degree of complexity the system can reproduce itself or even construct more complex entities.

> "Since the physical basis of mindlike qualities resides in the patterns of organisation of biological materials occurring naturally in animals, there is no reason why similar qualities may not emerge (in the future) from patterns of organisation of similar or other materials specially rigged to exhibit those qualities." [Singh, 1965, p202].

One should note here that it is this statement about 'mindlike qualities residing in the physical' which it is the task for computational physiologists to prove in this work exploring neural nets.

As Charles Sherrington has remarked,

> "It is a far cry from an electrical reaction in the brain to suddenly seeing the world around one with all its distances, its colours and chiaroscuro." [Singh, p203]

and in Penfield's work of direct electrical stimulation of the exposed cortex, the patient

> "is aware that he moved his hand when the electrode is applied to the proper motor area, but he is never deluded into the belief that he willed the action." [Singh, p204]

That is, there will be action co-ordinating or integrating centres 'above' the direct control networks. The stimulated versus the willed movement are distinguished as having different antecedents. The complex systems of neural nets are organised hierarchically with layers of processing nets projecting to higher "integrating" layers and so on up to the cortical planning and control layers. Also many layers use descending projections to control what they are being fed in the way of information. This prevents swamping and allows attention and concentration on particular processes.

W.Ross Ashby, Warren McCulloch, Grey Walter, and Norbert Wiener at a meeting in Paris (from Latil, P de: *Thinking By Machine*, 1956)

Coping with failures

John von Neumann, in attempting to produce a useful description of a reliable computing system added the idea of **redundancy** into the neural network in order to bring it more into line with the inherent unreliability of the physiological neuron net. Redundancy is a matter of using several copies of the same device with their outputs going to a majority decision device, so that if any one device fails the system still has enough functioning copies of the device to keep going. He showed that by using many redundant components in a circuit one could make an automoton with an arbitrarily high degree of reliability.

If any particular neuron net of moderately reliable organs is triplicated and the results of the three sent to a **majority decision organ** then the latter will give a highly reliable result equivalent to the result from a perfectly reliable single organ. But this leads to an uncontrollable proliferation of organs in something as complex as the brain. Von Neumann resolved this problem by increasing the number of input lines to each organ in the net such that the misfiring of a small number of components cannot cause a failure of the whole automoton or network. The redundancy in the system is increased considerably in order to control errors.

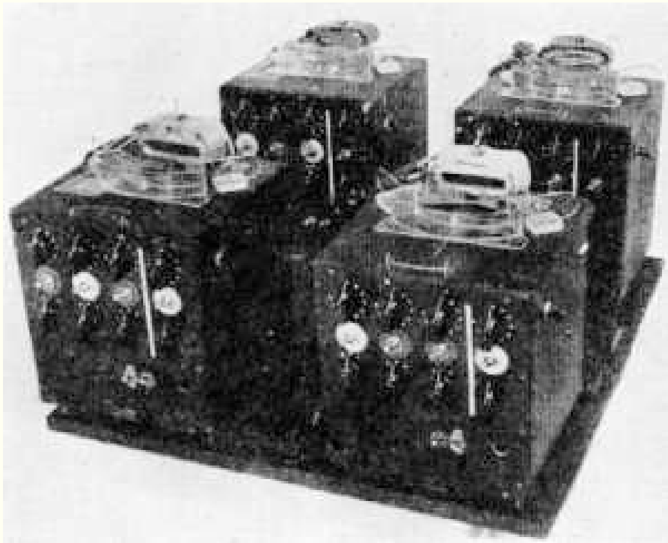
Coping with disturbances

The first thorough exploration of the behaviour of a mechanism in emulating living nervous systems was carried out by Ross Ashby in Great Britain and published in his book Design for a Brain in 1952. He was interested in the problem of how a dynamic system achieves a range of behaviours which may be said to show **stability** within the **limits** of survival for that dynamic system as well as **adaptability** to changes in the environment of that system.

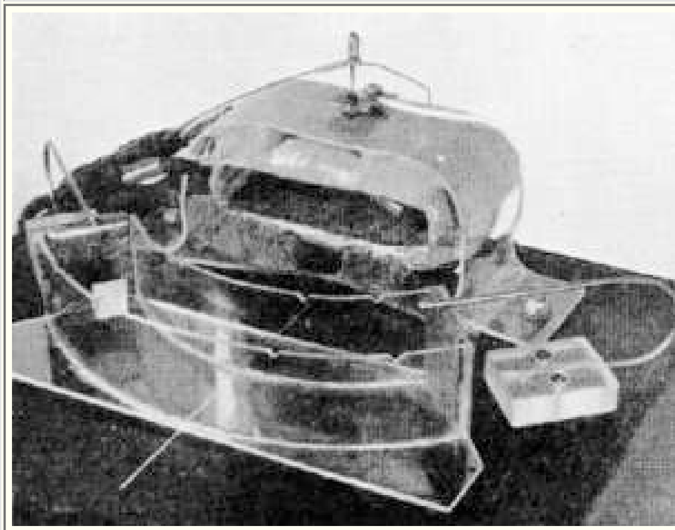
When a system is perturbed by the occurence of an input the system's response to the perturbation will be determined by previous experience (training). If the response is not exactly appropriate to the input then an **error** will occur. This error could be catastrophic for the system especially if the system is not particularly adaptive. In order for the system to adapt to a changing range of inputs it must be able to accommodate the errors. The incorporation of the difference between the actual response and the required response is known as **feedback** self-regulation.

Ashby built an electro-mechanical system employing a set of four pivoted magnets and an arrangement of electrical connections and impedences. The effects of the position of each magnet were routed to the other three magnets via a number of parameter altering devices, viz. selection switches and motion constraints. With any change in the operating conditions the positions of the magnets would automatically shift until the original specified conditioin of stability was re-established. This machine was the Homeostat and demonstrated an operating procedure which he called **ultrastability**.

Ashby used feedback in this self-regulating mechanical system he called the homeostat so that it would reach a stable state no matter how serious the **perturbation** of the inputs. Ultrastability is the capacity of a system to reach a stable state under a variety of environmental conditions. But the probability of stability being achieved decreases



Ashby's Homeostat (from Ashby, W.R: *Design for a Brain*, 1952)



Detail of magnet and coil from Homeostat

steadily as the system becomes more complicated. A large and complex system is very much more likely to be unstable.

Ashby showed that within a large complex system when any input disturbance is handled by only a small subset of the full array of subsystems in the system that disturbance will not affect the stability of the overall system but will, in fact, be easily accommodated by the system. If the subsets of input handling devices are different for differing ranges of input disturbances then the behaviours of the system will be spread over the system and no one input disturbance can take over or disturb or cause the failure of the full system. Ashby calls this the **"dispersion of behaviour"**, ie. responses to a range of inputs may be said to be "dispersed" over the system. Within Ashby's framework each of these input handling subsystems will be ultrastable and the complex will be **multistable**.

Obviously this is what our brains do with our various modes of sensory faculties. Ashby's theory is applicable to living animals: the entire array of possible environmental disturbances is grouped or dispersed into separate sensory systems and these systems are specialised to filter all but a very specific subset of inputs. An animal is thus built up of a number of ultrastable subsystems in a dispersed organisation. The animal's behavioural adaptation to new stimuli will reach appropriate responses much more quickly than in a single ultrastable system which had to generate responses to the full array of possible input/perturbing conditions. This sort of behaviour occurs similarly between the system and its environment and between subsystems within the multistable system. Ashby is suggesting that adaptive behavior and goal-seeking behavior in animals is handled via this principle of multistability (a system of ultrastable devices in a dispersed behavior system).

The adjustable synapse

Returning to neural nets. The McCulloch & Pitts neuron had a fixed **threshold**, so McCulloch developed a model with a variable threshold which in a network provided a means of changing the internal organisation of the network, making it more able to self-adjust to changing environment.

D.O.Hebb suggested a synaptic threshold modification principle:

> "when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency as one of the cells firing B is increased." [Hebb, 1949]

The idea of an adjustable synapse allows an artificial neuron network to go beyond the process of simply making decisions based on a look-up-table or the execution of a set of logical rules as in an ordinary computer. The network can tailor its response to, or "interpret", its input by **adjusting the weighting** of each synapse in adding to a threshold so that new responses can be made to variations in the input conditions. If the actual output of the network is compared with the desired output then an error value can be determined which can be incorporated into the weighting of the synapse

There is an almost biological principle of adaptation to conditions in operation here, the internal organisation of the learning machine can be altered according to the 'feedback' of an error value from the output or result of the process. This is the principle embodied in F.Rosenblatt's **Perceptron**.

The Perceptron consists in a net of sensor units feeding to a set of association units which feed one or more response units. If the sensor units feed enough 'yes' votes to the association unit to which they are mapped to exceed the threshold of that association unit then it will be excited or 'fire'. When enough association units fire so as to exceed the threshold of the reponse unit to which they are mapped, then the response unit will fire. If the result is correct then the tresholds of the response units will be left as they are, but if the result is incorrect then the thresholds of the response units will be modified. This process is iterated enough times for the response unit to give a correct response to the input of the whole Perceptron system. Thus the Perceptron is said to be 'trainable'. The output of the network is affected by altering the weighting or the value contributed by each connection.

> "In sum, the essence of the training scheme is to reduce selectively and iteratively the influence of active inner units which tend to give unwanted response, and enhance the influence of those that tend to give the desired response...after a certain number of trials a stage is reached when no further adjustment of weights of inputs into a response unit is required to secure correct identification by the machine in all subsequent presentations of picture patterns to it." [Singh, p228-9]

Mismatch

The next problem for neural nets, particularly for pattern recognition machines, is to allow for some level of mismatch, i.e. recognition on the basis of **similarity**. Humans handle the similarity problem with ease but a machine, especially an essentially classificatory machine, will have to go through a wide range of image transformation and generalisation. Image transformation employs shifting, rotation and scaling and in neural net systems is very neuron intensive, far to much so for it to be a viable model.

Pattern recognition by classification, and, with the inclusion of probability factors, by similarity is massively neuron intensive given the huge number of possibilities that any system might encounter. But if the system starts in a massively overconnected way, thus being provided with many more options than will be needed once trained, and abandons connections which are not used, then the neuron count can be kept down considerably once trained. But this arrangement would suffer from not being very flexible and being unable to account for new variations not encountered during training. The human brain is incredibly flexible and able to accomodate novelty, which none of the standard feedforward neural net models are able to do.

There is a model developed by Uttley for a kind of machine based on a conditional rarity model designed on classification and conditional probability principles with vast overconnections. This model does acheive the necessary economy of units - with chance connections. Because of the overconnections, there is initially ambiguity of discrimination, with units failing to recognise their associated input representations to a high degree. As information is accumulated those connections which carry little information become less effective until they are disconnected. Ambiguity is then eliminated and the system learns to discriminate, meeting some of the physiological and psychological facts rather well. This model has some considerable similarity to what occurs in the maturation of an infant's brain. The infant's brain is massively oversupplied with neurons and millions die off as the infant brain matures.

Simulation of biological neural nets

The strength of interconnections in a network of neurons determines how the network will respond as a whole to a particular input; "the pattern of connection strengths represents what the network knows." (Ferry, 1987, p55) The connections are bidirectional allowing for feedback circuits.

It now seems generally accepted that the brain's power arises spontaneously from huge numbers of neurons highly interconnected and processing information in parallel. These neuronal assemblies are defined as groups of 'neurons that are simultaneously active in response to a certain input'. But it is incredibly difficult to study these assemblies physiologically, getting enough electrodes into a small enough space to study enough neurons is currently out of the question, so finding an assembly and then showing its synchronised operation and its spread is incredibly difficult.

The neural net approach developed by McCulloch and Pitts was a hardware approach or was carried out on paper in mathematical and logical procedures. For a long time this work received very little attention and only in the last 15 years has a simulation approach emerged which models in the computer the interconnections and the interactions and simulates the activity of the nerve 'assembly' or the neural network. Given the massive numbers of synapses onto it that any one nerve might have, some of which are excitatory and some of which are inhibitory, and given the modulations of neurotransmitters across those synapses, the triggering of that nerve depends on the summation of all those inputs exceeding the threshold of that nerve. A cell's response might be a graded response according to the strength of the overall inputs or it might be an all-or-nothing response based on a threshold. All the outputs might be feedforward type processes or some might feedback into layers of cells preceeding the layer of the cell being considered, thus controlling its behaviour.

These systems of simulated neural nets can exhibit learning when based on D.O.Hebb's rule for learning developed in 1949, such that:

> "the connections between cells that are active at the same time will be strengthened, increasing the probability that the first cell will excite the second cell in the future. Connections between cells whose activity is not synchronised will be weakened. Synchronised patterns of firing that occur repeatedly will eventually become stable representations (or memories) of the inputs that give rise to them, and can be reactivated by only partial inputs." [Ferry, 1987, p56].

Other aspects of brain function modelling being explored include how can one maximise the number of representations that a given network can hold? Obviously, maximising the number of connections, it may be that every neuron in the brain is only a very few neurons away from every other, (in the realm of 5 neurons distant). But then how do these overlapping assemblies operate without interfering with each other? Possibly some form of negative feedback is involved in preventing an active assembly from becoming to large.

In a neural network simulation system each neuron takes account of the weightings of the outputs of its upstream neighbours to determine its own state. Of course, it is not the neuron somehow deciding of itself that it should look at what its neighbours are doing, this connectivity is already in place and is, so to speak, woken up by use. Each neuron exists in a net of similar neurons and they are wired up, axon to dendrite, synaptically connected, forming self-programming processing subsystems. These networks are inherently robust - answering von Neumann's earlier call for reliability in artificial computing systems: if some neurons malfunction the overall function of the network is not affected.

Information is encoded in neural connections rather than separate memory elements, as unique patterns of interconnections. Also the system learns 'spontaneously' because it alters the strength of particular interconnections according to the repitition of use of those interconnections and the array of possible experiences provided and possible solutions arrived at, i.e. training. In computer terms it might be thought of as 'self-programming'.

This training process: the alteration of weightings on particular synapses in the network; can also be acheived by a recurrent or feedback network in which an error weighting is generated by comparing the actual output to the desired output and then fedback into the weightings of the synapses of the processing layer of neurons.

Summary

To rehash the neural net idea I want to quote from Tank and Hopfield's article in the December 1987 issue of Scientific American.

> "A biological neuron receives information from other neurons through synaptic connections and passes on signals to as many as a thousand other neurons. The synapse, or connection between neurons, mediates the 'strength' with which a signal crosses from one neuron to another. Both the simplified biological model and the artificial network share a common mathematical formulation as a dynamical system - a system of several interacting parts whose state evolves continuously with time. Computational behaviour is a collective property that results from having many computing elements act on one another in a richly interconnected system. The overall progress of the computation is determined not by step-by-step instructions but by the rich structure of connections between computing devices. Instead of advancing and then restoring the computational path at discrete intervals, the circuit channels or focuses it in one continuous process." [Tank & Hopfield, 1987, pp62-63]

One might liken this activity to the human process of consensus decision making, where a problem is discussed until everyone involved knows enough about it for a decision to evolve from the range of opinions held by individual members of the group. The 'computational surface' of the nodes in a neural network shifts according to the weightings of each node. The weightings alter with training, i.e. through exposure to examples of the kinds of problems to be encountered by the particular network, and the solutions develop in the form of a kind of best-fit. "The network carries out the computation by following a trajectory down the computational surface. In the final configuration the circuit usually settles in the deepest valley" to find the best solution. [Tank & Hopfield, 1987, p67]. This approach is good for perceptual problems and for modelling associative memory. Using Hebbian synapses we can develop a model for learning: "synapses linking pairs of neurons that are simultaneously active become stronger, thereby reinforcing those pathways in the brain that are excited by specific experiences. As in our associative-memory model, this involves local instead of global changes in the connections.

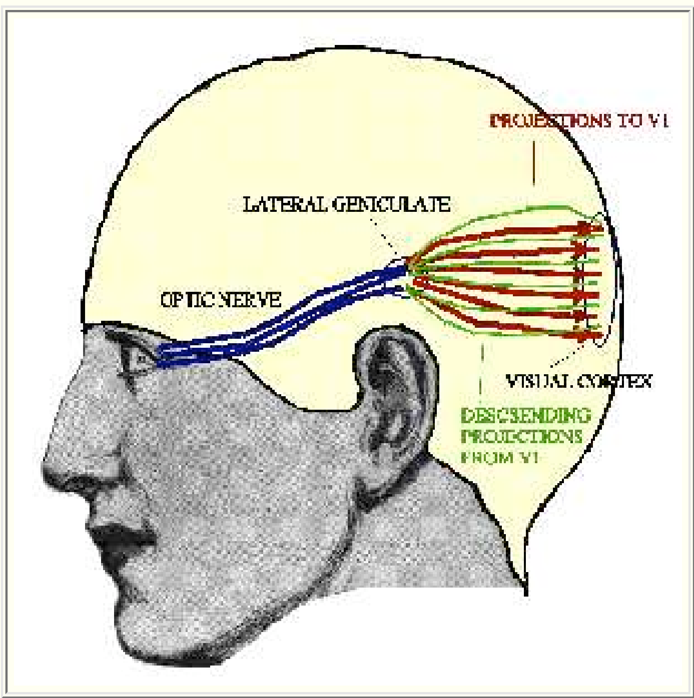
Modelling real networks

Computational neuroscience is one of the major areas of investigation into what it is that brings about consciousness in what we know to be the extraordinarily complex but highly organised networks of neurons in the human brain. At the Tucson II conference Paul Churchland, of the University of California at San Diego, asked:

> "Have we advanced our theoretical understanding of how cognition arises in the brain?
> Yes: Through **artificial neural networks** that display
>
> a) learning from experience,
> b) perceptual discrimination of inarticulable features,
> c) development of a hierarchy of categories or framework of concepts,
> d) spontaneous inductive inference in accordance with past experience ("vector completion").
> e) Sensorimotor coordination between sensory inputs and motor outputs
> f) short term memory with information selective decay time
> h) variable focus of attention
>
> [from a slide in Churchland's presentation at Tucson II]

Churchland took us briefly through (a "cartoon version" of) the visual system.

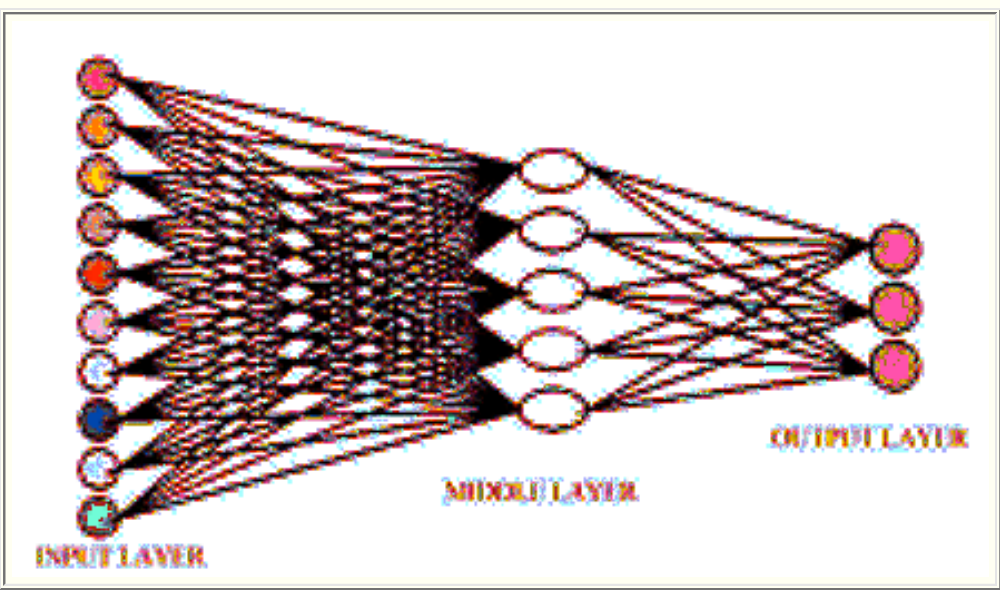
A pattern of light: "... a representation on your retina is transformed by going through a trillion little



synapses into a new pattern at the LGN that is projected forward to V1 (the primary visual cortex), (where it is) transformed by another population of synaptic connections in to a third pattern. It rather looks like the basic mode of representation in the brain is in patterns of activation across populations of neurons, and the basic mode of computation is transfomation from one pattern to another pattern, to another pattern. Transformations which co-opt relevant kinds of information. Information that is relevant to its day-to-day behavior." [from Churchland's presentation at Tucson II]

This is very similar to the structure of a neural net (of course !, given that they were designed from actual neurons).

One of the primary problems being used in neural network development is that of face recognition, i.e. attaching a name to the face. Churchland presented work done by Garrison Cotrell's group at the University of California at San Diego, using a feedforward neural network havng 80 cells in the inner layer, which did a pretty good job of the basic face recognition task. He then mentioned some of the ways in which it failed and how one might deal with these failures using recurrent or feedback weighting of the connections in the network and discussed how this relates to some aspects of consciousness such as short-term memory.



Churchland's description of the face-recognition network: The **Input layer** is made up as 64 x 64 neurons (4096 neurons) consisting in photocells having a photograph of a face projected onto it ("being stimulated to an appropriate level of brightness"). The **midddle layer** consisted in eighty cells. The **output layer** had eight cells which give 8 bits to identify: face/non-face; male; female; "name" (5 bits); The network was trained up on about 11 faces and a small number of non-faces, with a number of examples of each face, and it did very well on the three kinds of distinguishing it had to do. When shown a test set of novel faces it did about 10-15% less well than it did on the learned set. Still a remarkable performance and not far down on our own sort of performance.

But this network cannot discriminate ambiguous images (like the duck-rabbit illusion). To paraphrase Churchland: What a feedforward neural network does is embody an input/output function, with a unique output for every different input. To achieve something like the handling of ambiguity we need something more than feedforward networks. So he introduces "recurrent pathways" which bring contextual information from the rest of the system of the brain and feed it back into the network. This allows the network to "modulate its own responses to perceptual input" These recurrent pathways are the channels for the feedback information which we have discussed above. For example there are a very large number of descending pathways from the visual cortex back to the LGN, more than there are projecting from the LGN up to the visual cortex.

Recurrent pathways were originally introduced into neural nets as a form of short-term memory. They also provide a level of directability and handling of amibiguity as well as answering some of the other desiderata for a theory of Consciousness. In the brain, the best candidate for a neural correlate of consciousness is the thalamo-cortical system which is a massive recurrent network centering on the thalamus (see Newman and Baars on the thalamo-cortical system)

Artificial Intelligence

Of course, all this neural network work has another intention besides elucidating what it is in humans, or biological systems in general, that produces perception and attention and consciousness; and that is: Is it possible to build an artificial consciousness, a silicon system which might display some kind of consciousness? And even more importantly how do we test the machine we built to see if it really is conscious?

Two tasks had to be carried through before real computing machines, let alone intelligent machines or AI's, could be developed. First was the software problem: developing systems of algorithms for problem solving, and secondly, the hardware problem: a theoretical machine had to be produced which could deal with these algorithmic systems and treat them as instructions for its operation.

Dealing with **the hardware problem** first:
Alan Turing is the name most associated with the development of electronic computers. He showed that it would be possible to build universal, or

reprogrammable, computing machines by developing an abstract version which became known as the Turing Machine. In this device a machine reads a paper tape on which symbols or spaces are written. These symbols tell the machine what to do next after reading the symbol: the symbols may then be erased or new symbols may be written, and then the tape is stepped on to the next symbol. This provided a conceptual base for the development of machines which could be supplied lists of instructions on how to solve a problem, which could then proceed as instructed and generate a solution to that problem at which point the machine would halt. He also showed that some mathematical problems were not 'computable', i.e.were unable to bring the universal machine to a halt, and were therfore not amenable to solution using an algorithmic process. [An algorithm is a fixed set of instructions which if carried out correctly will solve the problem they are designed to solve no matter how much time it might take.]

His other great contribution was the development of the **Turing Test**. Turing considered that a universal problem solving machine was something which could be constructed in any of a variety of materials. In the 19th century Charles Babbage had proposed an 'analytical engine' which would have been contstructed with mechanical components and driven by steam. Turing was building machines using electrical components, but he considered that in principle it wouldn't matter what the machine was built of so long as the operating principles were of the nature of his universal machine. So a flesh and blood machine would be possible and given that the existing flesh and blood machines were, more or less, intelligent, then a machine built in other materials might also be intelligent. But how would you know whether this possible machine, once built was truly intelligent? Thus the Turing Test: a machine and a human are placed in a room, with an interregator outside the room. They communicate to each other by some agency, e.g. a teletype machine, which doesn't reveal clues such as voice quality. The interregator asks questions of the two in the room and if the interregator is unable to tell which of the two is the source of the answers then the machine can be said to be intelligent. (Or, of course the human can be said to be dumb, which is a comment that Jaron Lanier made in his presentation at Tucson II.) So the idea is that if a machine can be made to behave in a manner indistinguishable from a human then that machine should be described as intelligent.

So this is the prime test that an artificial intelligence would have to pass.

The other problem was **the software issue**:
The software task for the putative designers of artificially intelligent machines was the production of a mathematical system which reduced reasoning to a mechanical process, first in arithmetic and then in more generalised systems of reasoning. So mathematicians of the second quarter of the 20th century spent a good deal of effort trying to develop formal logical systems which could provide general problem solving algorithms. These were to provide a basis for a consistent theory of arithmetic (known as number theory) and would be later employed in programming computers. But the Czech born mathematician Kurt Goedel caused a considerable upset when he showed that any formal logical system was necessarily incomplete. This is known as Goedel's theorem or **Goedel's Incompleteness theorem**

and works like this:

> In any formal system S, which is consistent, there can be a proposition which denies the provability of that proposition (of itself) within the system; i.e. the statement "this statement cannot be proven within S" can exist within S. Since this proposition can exist then it must be true, which denies its 'not provable' status, and therefore produces an inconsistency within what is supposed to be a consistent system. Thus no formal system of propositions can be complete.

This result has quite extraordinary consequences but these have been very differently interpreted by various people. The first interpretation has been that it provides an avenue for the existence of freewill in the world. Another interpretation, which is slightly more relevant to our discussion here, is that it has been suggested that no computing machine will be capable of becoming intelligent in the way that humans are because the formal algorithmic systems; i.e. the systems of programs, that a computer is constructed with can never provide the sort of mathematical 'understanding' or 'truth finding' capabilities that humans have. But this is to take a very narrow and simplistic view of the ways of designing machines as well as intelligence.

Now, I have another view divergent from this and that is that what the Godel result indicates is that to consider intelligent machines as being restricted to formal mathematical systems of propositions, that is algorithmic programming, is to severely misunderstand both human intelligence and the implications of incompleteness. The point is: Human intelligence is **generative**, i.e. it is capable of constantly producing new ideas, new sentences of language, new creations of art, new musical productions, etc, etc... and this is because an intelligent system is necessarily incomplete. It is in this idea that the possibilities for intelligent artificial constructions lie. A machine which is capable of passing the Turing Test must not only be able to pass a maths exam, but it must also be able to make up a new story about the neighbours or worse tell a lie about itself, that is it must be generative, always able to produce new sentences about itself and any other content of its system or its context.

There were two presentations at the Tucson conference which used an idea similar to this.

One was from **Steve Thaler** of Imagination Engines, Inc. in which he introduced a neural net construction in which the standard input stimuli are removed and replaced by various kinds of noise generation within the network cavity itself. The "internal noise...is shown to produce a temporal distribution of distinct network activations identical to that in human cognition" [Thaler, abstract of Tucson II presentation, and see Holmes, 1996]. The nets he uses generate new versions of existing structures by stochastically altering the weights of the internal connections of the net. Many of the results will be useless but some will be useful, and these can be easily selected out. For example, Thaler uses his neural nets to develop everything from new designs for motor-cars to composing all possible pop-songs.

The other was by **Daniel Hillis** of Thinking Machines Corporation, who are developers of massively parallel computers. Hillis spoke about his attempt to use evolutionary techniques to simulate the design of a very complexly organised machine.

In his presentation to Tucson II Hillis outlined a number of the current arguments against intelligent machines based on the failure of algorithmic computation to produce true intelligence. Then he presented a technique of evolving machines which, more or less randomly, may or may not come near to solving the problem being set. The most successful of these machines are then selected out and 'married' by a sexual combination process and new machines produced which are tested, and so on around the cycle for many thousands of generations. As he says the circuit diagram may be impossible to read but they are the most efficient machines for solving the particular problem ever produced.

Finally, I'll give **John Searle** the last word on "Can a computer be conscious?". He says that if you define a machine as a physical system that performs certain functions then our brains are machines: "I am a machine that thinks...The question: can a computer think? obviously has a trivial answer, Yes. My brain is a computer. Listen 2 + 2 = 4, that's a computation. So I am a computer that thinks." But conversely, computation as defined by Turing (as symbol manipulation) does **not** constitute thinking or consciousness. [Searle, presentation to Tucson II]

References

Ashby, W.R. (1952) Design for a Brain. Wiley

Brewster, Sir D. (1868) Letters on Natural Magic. W. Tegg.

Churchland, P.M. (1988) Matter and Consciousness. MIT Press

Churchland, P.M. (1989) A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. MIT Press

Churchland, P.M. (1995) The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. MIT Press

Ferry, G. (1987) "Networks on the Brain". New Scientist. 16 July 1987, pp54-58.

Goedel, K. (1931) "On Formally Undecidable Propositions of Principia Mathematica and Related Systems I" in Davis, M.(ed) (1965) The Undecidable. Raven Press

Hebb, D.O. (1949) The Organisation of Behavior Wiley

Hillis, W.D. (1985) The Connection Machine. MIT Press

Hillis, W.D. (1985) "The Connection Machine". in Scientific American, June, 1987, pp86-93

Holmes, B. (1996) "The Creativity Machine". New Scientist. 20 Jan.1996, pp22-26.

McCulloch, W.S. and Pitts, W.H. "A Logical Calculus of the Ideas Immanent in Nervous Activity" in McCulloch, W.S. (1965) Emodiments of Mind. MIT Press.

Searle, J. (1992) The Rediscovery of Mind. MIT Press

Singh, J. (1965) Great Ideas in Information Theory. Dover.

Tank, D.W. & Hopfield, J.J. "Collective Computation in Neuronlike Circuits", Scientific American, Dec 1987, pp62-70.

Turing, A. (1950) "Computing Machinery and Intelligence" in Mind, 59

von Neumann, J. (1958) The Computer and the Brain. Yale University Press.

von Neumann, J. (1966) Theory of Self-Reproducing Automata. University of Illinois Press.

To return to the discussion of organisation and complexity.

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

Susan Greenfield

talks to Stephen Jones at Tucson II.

SJ: Susan, during your paper today you spoke in some detail of the concept of a neuronal assembly which by its plasticity might provide a Rosetta Stone in the search for the neural correlates of consciousness. Could you outline your ideas about neural assemblies for me?

Susan Greenfield: I think we should start with two assumptions and they're not particularly controversial assumptions. Let's assume that consciousness isn't beamed in from outside, and let's assume that biological brain is the only entity that we know of, at least at the moment, that generates consciousness. If you go along with that, then our problem is to somehow cater for consciousness using the physical brain. We already know there is no fixed area of the brain which is the centre of consciousness, and similarly we know its not distributed evenly over each and every individual brain cell. So really the only framework left to us is that of the neuronal assembly. But a neuronal assembly that is highly transient. A neuronal assembly where the components can belong to a number of assemblies and where they can do other things as well as generating consciousness.

So my starting point is to somehow work out what forces or factors would control the formation of transient neuronal assemblies. And how these would relate to consciousness. And I think a good starting point is perhaps a little unscientific. That is, to stand back a bit and to say well, before we rush in with our electrodes or our lesioning equipment, lets look first at what we want of the brain. What do we want the brain to deliver that's going to cater for consciousness. What properties of consciousness could there be? And to my mind there are three very fundamental properties of consciousness that we're going to have to make allowances for.

First, as we've just seen, is that there is no centre for consciousness. So, if we're postulating individual groups of neurons we have to say that it's potentially spatially multiple. At the same time most of us would say that we only ever have one consciousness at any one time. So the first property of consciousness, as I see it, is that it's spatially multiple, but temporally unified.

The second issue is again a very simple one, to my mind. Which is that we are always conscious of something. By definition, if you're conscious of nothing, you're unconscious. So we somehow have to think of an epicentre, like the stone in a puddle or the epicentre of an earthquake or boss in the middle of a load of telephones. Something that triggers a neuronal assembly to be large enough to somehow mediate consciousness in some way.

The third property of consciousness again, I think, is a very straightforward one, that in itself isn't particularly original but just hasn't ever been stated formally before. Think about animals or foetuses or children. It's often been a riddle as to whether they're conscious or not, and how conscious they are relative to us. To my mind, let's take the foetus, it would be very odd to postulate that as soon as it's born suddenly all the console lights go on and suddenly something dramatic happens to the child to make it conscious. We know that the brain during birth is no different just before or just after birth. If, on the other hand, a baby is conscious in the womb, then when does it become conscious? Did it suddenly become conscious? Or is it postnatally, when it's two years old that it suddenly becomes conscious? All these are very odd scenarios. And they've proved a riddle to many people.

I think the way around that is to say simply that consciousness grows as brains grow. That it's not an all or none phenomena. Rather it grows in proportion to the sophistication of the brain. And if you went along with that idea then you'd also be able to explain animal consciousness. That a rat would be not as deeply conscious as a cat, which wouldn't be as deeply conscious as a chimpanzee, which in turn wouldn't be as deeply conscious as say George Bernard Shaw or Van Gogh. So, my third possible property of consciousness is that it's a continuum. That is to say, it's continuously variable.

If we put those three properties together one can then go back to the brain. What we are looking for, then, is neuronal assemblies that are forged transiently. Which are triggered by different epicentres: by things in the outside world or in more sophisticated brains, inner thoughts, in some way. And that vary in size, and according to the variation in size of the neuronal assembly, then you would have a certain degree of consciousness at any one moment.

Now we actually have evidence that neuronal assemblies do exist and that they are highly transient and highly dynamic. There's a group in Israel, led by somone called Grinvald who's worked with voltage sensitive dyes. Now what this means is that you can watch neurons light up in large populations and this has an advantage over conventional techniques where you record electrical activity as it goes on in the cells. And what this group is showing is indeed if you shine, lets say something very simple like a spot of light at a frog, then you don't just have the simple response of the brain, but one that grows over time gradually with more and more recruited neurons. Just like the stone in the puddle: concentric rings generated ever outwards as they grow. Moreover you can have variation in size, according to competition. Its a bit like arm wrestling, where if you shine another light that will limit the formation of the initial neuronal assembly.

So we know neuronal assemblies already exist. What could actually govern their form? The extent of the neuronal assembly? I think it would be obviously the strength of the stimulus: how bright the light was, in a simple example. Also, the degree of connectivity. Imagine a boss going in and wanting to phone up all his or her managers or sub-managers, clearly telephone wires would have to be there. And finally, the receptivity. Now this is something people who don't normally work on the brain find it hard to think about because you think about neurons as rather passive boring little blobby things that just fire action-potentials, signals, or don't fire. But in fact they can be enormously biased. And the analogy might be: think of a telephone network ready to receive the boss's phone call if a rumour had gone round that there was to be a pay rise, everyone would be very keen to answer the phone, and instead of staring out the window or arguing with their neighbours or perhaps ignoring the phone altogether, they'd be much more alert, ready to pick up the phone.

Now, the brain has that equivalent of a rumour, if you like. I.e. something in itself which means nothing, but can bias responses. And that is mediated by a variety of chemicals that are released almost like fountains being released. So in themselves they do nothing but if while they're there a neuron is excited then it will enhance the response. Now these transmitters, these chemicals that fountain out from the brain are actually the targets of gluta-modifying drugs, such as anti-schizophrenic drugs and antidepressant drugs such as prozac. So we know they do play an important part in consciousness. And to my mind what they could do is they could provide a third factor in determining the extent of the neuronal assembly. In that they could dictate how readily a neuron was corralled up, however transiently, to be recruited into this transient neuronal assembly: corralled up, recruited by the stone in the puddle, the epicentre, the boss on the phone.

So I would suggest therefore that we do have a Rosetta stone, a neuronal correlate of consciousness, and we can actually use that to journey in either direction. Either we could start with those factors and see how they relate to phenomenology, to certain charicatures of consciousness. Or we can travel back in the other way we can look at certain conditions of phenomenology, certain types of consciousness, and see if we can explain it in terms of neuronal substrates.

So let's go first in one direction. Imagine abnormally small neuronal assemblies. That could be due to any of three factors, as we've said, the epicentre not being very strongly stimulated. That might occur for example in dreams. Poor neuronal connectivity; that occurs very definitely during development, because we're born with very few neuronal connections compared to when we're two or three years old. Or the state of arousal. The amine availability; that would be determined by our biorhythms or perhaps an aberration.

Now for each of those situations I can give you an example. We've seen childhood for poor connectivity. I think a weaker perceptual stimulation, where things aren't flowing in from the out side world can produce dreams. And interestingly enough there is the very strong association between dreams and childhood. A child will spend a lot more time in dreaming sleep than us adults. In fact at a certain stage, I think its 26 weeks in the womb, its entire conscious state, its entire life is spent in dreaming sleep, in REM sleep. The third state I think which is similar to dreaming and childhood is schizophrenia. In this case, as in childhood, and in dreaming, you're very much the passive recipient of the outside world, and I would argue that this is because you're neuronal assemblies are too small to generate inner resources to have inner associations, extensive memories, extensive interpretations of what's happening. Rather you're processing in a very banal and literal way what is coming in. We know in schizophrenia for example they are very poor at attempting proverbs, if you say "a rolling stone gathers no moss", and ask them what that means they'll say, well it means that a stone when it rolls doesn't gather moss. Rather like a child would interpret the proverb.

So, I would say, this profile of small neuronal assemblies belongs to a range of conditions: childhood, schizophrenia, dreaming. Imagine when you wake up from a nightmare. Although you tell yourself it's a nightmare, it's very hard to rationalise it away and not experience a horrible type of consciousness and I think perhaps being a schizophrenic might be like that a lot of the time. Certainly for a child it's like that. My small brother when he was about four was convinced there was a dragon under the bath, and however much we tried to rationalise it away, to appeal to his inner sophisticated brain, because it didn't have much connectivity, you see, he just wasn't convinced. He just knew there was a dragon under the bath and that was the end of it. Rather like when you wake up from dreams and you're frightened.

So that would be a charicature I think of where we can look at different factors, what we could call charicatures of a small assembly, from different aspects of life, but see that nonetheless there are certain factors in common in the type of consciousnesses produced. By the same token, abnormally large assembly would result in, one would predict therefore, in the outside world being very remote, very grey, with an inner cognitive, an inner abstract type of thought dominating. I think that could be the case with depression. In fact we know that with depressives one of the biggest features that they display is that the outside world is seen in shades of grey. So that is, if you like, the sort of things I think we'd be able to do if one were to buy into this model: of going from the physiology, of manipulating things with drugs, arousal levels, the ontogenetic status of the brain, or dreaming states, and then see how they can be explained in those terms.

Or we can go the other way. Let's take a sensation which can be a type of consciousness. Pain is a good one. See if you could explain that in terms of different neuronal assemblies. I think the interesting thing about pain is we tend to think about pain as being all or none, and that everyone has the same type of pain, but its far more fascinating than that. We know that pain thresholds, that is your sensitivity to what you see as pain, vary enormously with an individual throughout the day. We know that people, say, on battlefields will not register pain until a lot longer afterwards. We know that if you tell someone that they're about to feel pain, the anticipation leads to a perception of pain far greater than if someone isn't told they're going to have pain. We know there's phantom limb pain, that people feel pain in the source of something that isn't there. We know we've also got morphine, which is one of the oldest analgesics, one of the best analgesics in the world, acts in a very interesting way. People say they still feel the pain, it just doesn't matter any more.

So this suggests to me, that perhaps pain, or the depth of pain you feel, is reflected in the depth of our conscious assembly at any one time. So, for example if you were told you were going to feel pain and had time to create a lot of assemblies, a lot of scenarios in your mind, so that your neuronal assembly is larger. Similarly, if you're in an accident or on the field of battle, so much is going on that a whole multitude of neuronal assemblies are jostling and pushing each other out of the way. So there isn't time to form a large neruonal assembly. Similarly, the action of morphine might be, and this could be something one could try experimentally, might be to reduce the efficiency of neurons to wire up to each other, to create a sizeable neuronal assembly. Hence the analgesic affect. One's still aware of it, as one might be aware of a weakly shining light, but because it can't trigger a large assembly it is literally not significant. That would fit with how we describe pain in the first place, which is in terms of association: such as pricking, burning, stabbing. One thinks of pain always in metaphorical terms.

So I think to a certain extent we could draw up, at this stage, a rather crude table of, on the one hand physiology things: that is, amine levels which are reflecting arousal; degree of connectivity, which reflects the ontogenetic status of the brain; epicentre strength, which reflects the degree of stimulation coming in and the number of connections. So for everything there is a parallel between the physiology and the folklore. Or we could start at the other end, and look at alzheimers disease and schizophrenia or dreaming or childhood. In both cases you could journey in and express those conditions in terms of size of neuronal assembly. And I think in the future, the very near future we're going to have the technology to look at the size of the neuronal assembly. With a sufficient time and space resolution we could actually start to make accurate predictions of using that as an experimental Rosetta Stone.

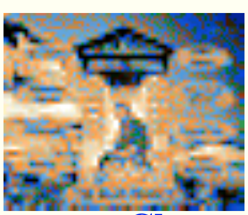
Links

"The mind's pot- pourri" by Susan Greenfield: announcement of Journal of Consciousness Studies

Dances with neurons by Susan Greenfield; a review of Conscious Experience Edited by Thomas Metzinger

---

To return to the physiology discussion.

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

An Introduction to the Physiology of Ordinary Consciousness

by Stephen Jones.

"You are but one function of the brain."

[Rodolfo Llinas at Tucson II]

The reigning question seems to be: Does consciousness stand alone, separate in some way with respect to the brain, or is it an intrinsic aspect of the brain's functioning? If consciousness is intrinsic then is there some sort of place within the brain that we can isolate as being the seat of consciousness? Or do we have to look more for an arrangement distributed over a larger scale? It is these latter two questions that those physiologists interested in consciousness are asking.

The Homonculus

Daniel Dennett, a philosopher from Tufts University in Boston and author of Consciousness Explained, discussed the issue of whether there is a particular organ of the brain in which the sensory data coming in through the eyes to the primary visual cortex is converted further in order to make it what we see and report about, i.e. to make it "qualia". This is really the idea that there should be some particular organ of the **self** in the brain.

In looking at the connectivity in the visual system Dennett dispenses with the idea of the **homonculus** or a little man inside the brain which does the consciousnesss work. As Dennett says, it is an empirical fact that there is not an homonculus in the brain, and even if there were we would still have to go through the same investigation of the homonculus' consciousness, and so on *ad infinitum*. For Dennett:

> "The work done by the homonculus in the Cartesian Theatre ... must be distributed in both space and time within the brain." [Dennett, from his presentation to Tucson II]

It seems pretty unlikely that there would be a particular organ of the brain which takes the incoming, by-now-integrated (necessarily so, if this model were to work), data flow from one's body and the world and converts it (puts it up on the screen of the minds' eye) into the qualia laden subjective world that we report. Each layer of processing is the subject of each pattern of data from the input sensors in its stream, transforming the pattern into salient features appropriate for the next layer. A multitude of processing subsystems take part in all the activities of the every-day. They interweave their connectivity throughout the brain, supplying divergent and convergent information pathways wherever necessary.

The N.C.C.

So the problem for those neurophysiologists who are interested in consciousness is in discovering the neurophysiology of mental processes. If there is no one organ of consciousness: what is actually going on in the brain and what could the distributed systems of the brain be that, at least, underlie consciousness? Put another way: What are the **neural correlates of consciousness** (which phrase is often reduced to the N.C.C.)? What should the search for the NCC provide us? and will it provide us with an explanation of consciousness? For example, the neural correlates of seeing are the primary visual pathway and its divergences. But vision is not the be-all and end-all of consciousness (the blind are after all perfectly conscious). Drawing on David Chalmers presentation to Tucson II: Being conscious involves having, and being able to report having, information which is in some way 'globally available' to an organism. If a neural mechanism of 'global availability' were found empirically, then such neural mechanism could be an NCC. Chalmers calls this a "bridging principle" between consciousness (the phenomenal) and neurophysiology.

Bridging Principle: **- Consciousness ~ Global availability** **Empirical work:** **- Global availability ~ Neural process N** **Conclusion:** **- Neural process N ~ Consciousness** [slide from Chalmers presentation at Tucson II]

There are several candidates for a neural mechanism of global availablity, including Llinas' 40Hz oscillations and Baars' "Global Workspace".

A full story about the NCC should simultaneously: **- 1. explain availability** **- 2. underlie consciousness** **(Though it probably won't explain consciousness!)** [slide from Chalmers presentation at Tucson II]

Chalmers also, doesn't think that there will be a single NCC. It is more likely that consciousness is a distributed 'activity' and involves many different processing systems handling data in many modalities which becomes available within many forms of representation and description.

We are searching for the embodiments of subjectivity, and we must ask whether or not that is all we need to explain what it is that brings about consciousness.

Distributivity

The neurophysiologist Joseph Bogen argues that consciousness is subjectivity, and he comments that you can't see subjectivity; it's like looking for the wind, you can only see its effects. Bogen suggests that we look for a centre (a nucleus) that has **distributivity** (i.e. widespread inward and outward connectivity) as a site that produces subjectivity as consciousness. He is referring to the Intra-laminar Nuclei which is a subpart of the Thalamus. We will return to Bogen and the thalamus later, once we have looked a little at the anatomy of the brain and the physiology of the nerves.

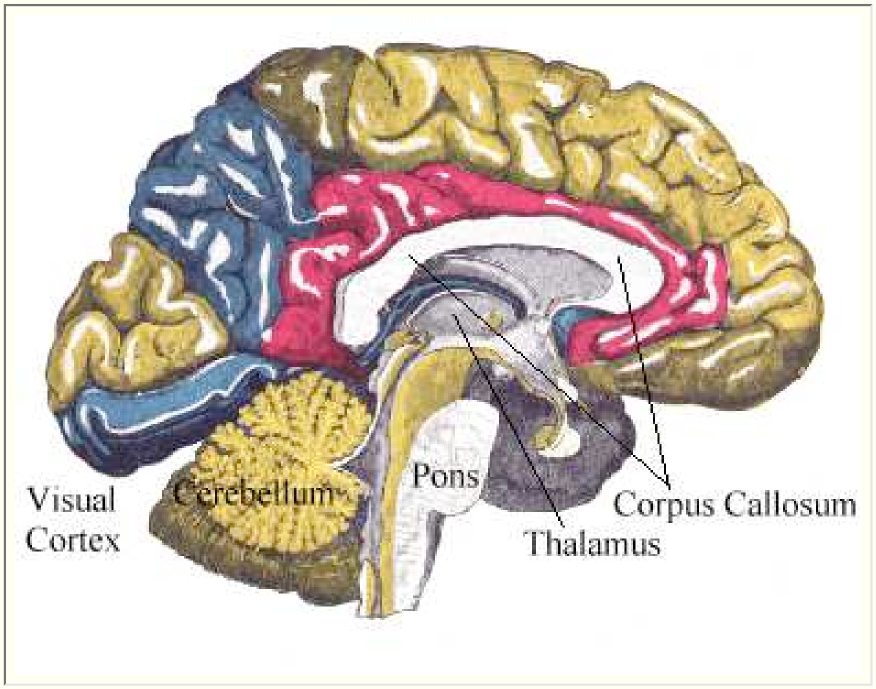
Neurophysiology

So what are the roles of neuroanatomy and neurophysiology? To quote from one of the great neurobiologists, David Hubel, who with Torsten Wiesel did most important work in elucidating the anatomy and physiology of the visual system:

> "Anatomy seeks to describe the various elements of the brain and how they are put together; physiology asks how the parts function and how they work together." [Hubel, 1979, p40]
>
> "The number of nerve cells, or neurons, that make up one's three pounds or so of brain is in the order of 10(to the power of 11) (a hundred billion) give or take a factor of 10. The neurons are surrounded, supported and nourished by glial cells, whose number is also large. A typical neuron consists of a cell body, ranging from about five to 100 micrometers (thousandths of a millimeter) in diameter, from which emanate one major fiber, the axon, and a number of fibrous branches, the dendrites. The axon may give off branches near its beginning and it often branches extensively near its end. In general terms the dendrites and the cell body receive incoming signals; the cell body combines and integrates them (roughly speaking, it averages them) and emits outgoing signals and it also serves for the general upkeep of the cell; the axon transports the outgoing signals to the axon terminals, which distribute the information to a new set of neurons." [Hubel, 1979, p39]

Physiology is impossible without the data of anatomy, but then we must ask what the structures anatomy describes are for?

Well, let's do a quick precis of **neuroanatomy**.



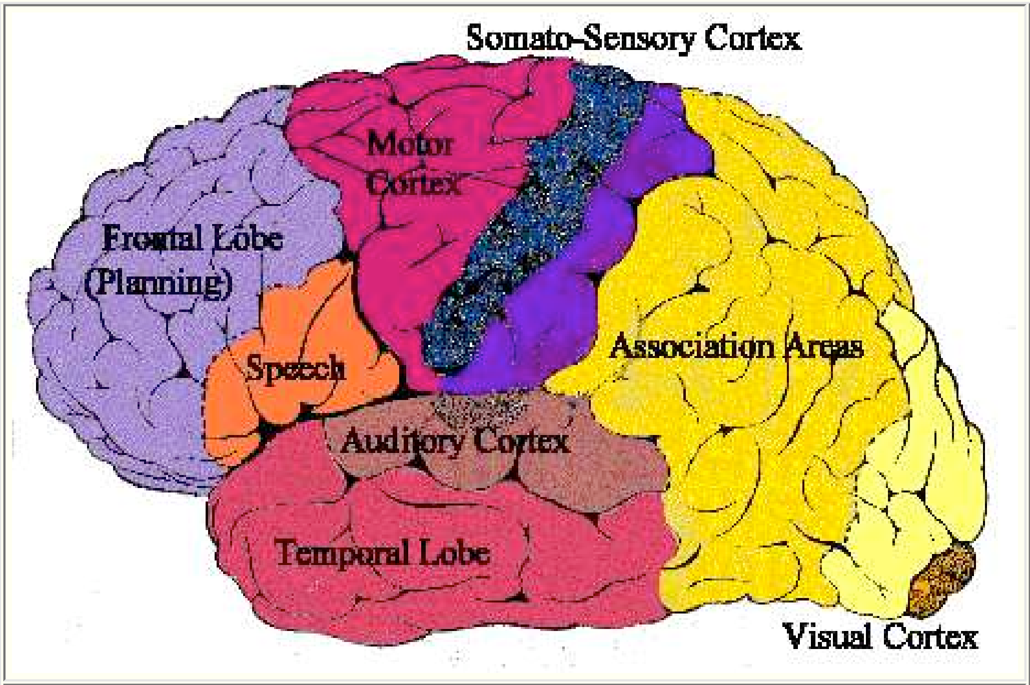
Sagittal view of Brain

The brain is generally divided into three layers of architecture which house the processing subsystems for its variety of functions:

The **Hindbrain** (or old brain) at the base of the brain provides the connection into the spinal cord. The **medulla** which controls some of the vital functions such as the heart and respiration. And the **cerebellum**, which is involved with the co-ordination of movement with sensory and somatic data from the rest of the brain and the body. It might be called the integrator of bodily control.

The **Midbrain** Contains the **tectum** which is involved with visual reflexes and detection of movement and the **reticular formation** which is an array of structures concerned largely with basic arousal, vital survival reflexes and the **thalamus** which distributes sensory and motor information to the cortex, and receives control information from the cortex.

The **Forebrain** or the **cortex** (also known as the cerebrum): overlaying the rest of the brain, greatly enfolded and larger by far than the older parts of the brain. Describing its functions working from the front of the brain to the rear: In the foreward regions of the cortex, the **frontal lobes** are involved in planning and control of movements. Behind this is the **primary motor cortex** which is involved in the control of movement. Then the **somato-sensory cortex** which receives information about the body senses. Behind that are the **association areas** involved in memory and interpretation of sensory data. The **primary visual cortex** is at the rear of the brain and the **primary auditory cortex** is at the temporal lobes of the cortex.



Cortex from the left side

Running through this again from a slightly different angle, the various subsystems of the brain are bundles of nerves gathered into specific functional groupings and chains of interconnections which are beginning to be well understood. These bundles of nerves are the organs of the brain.

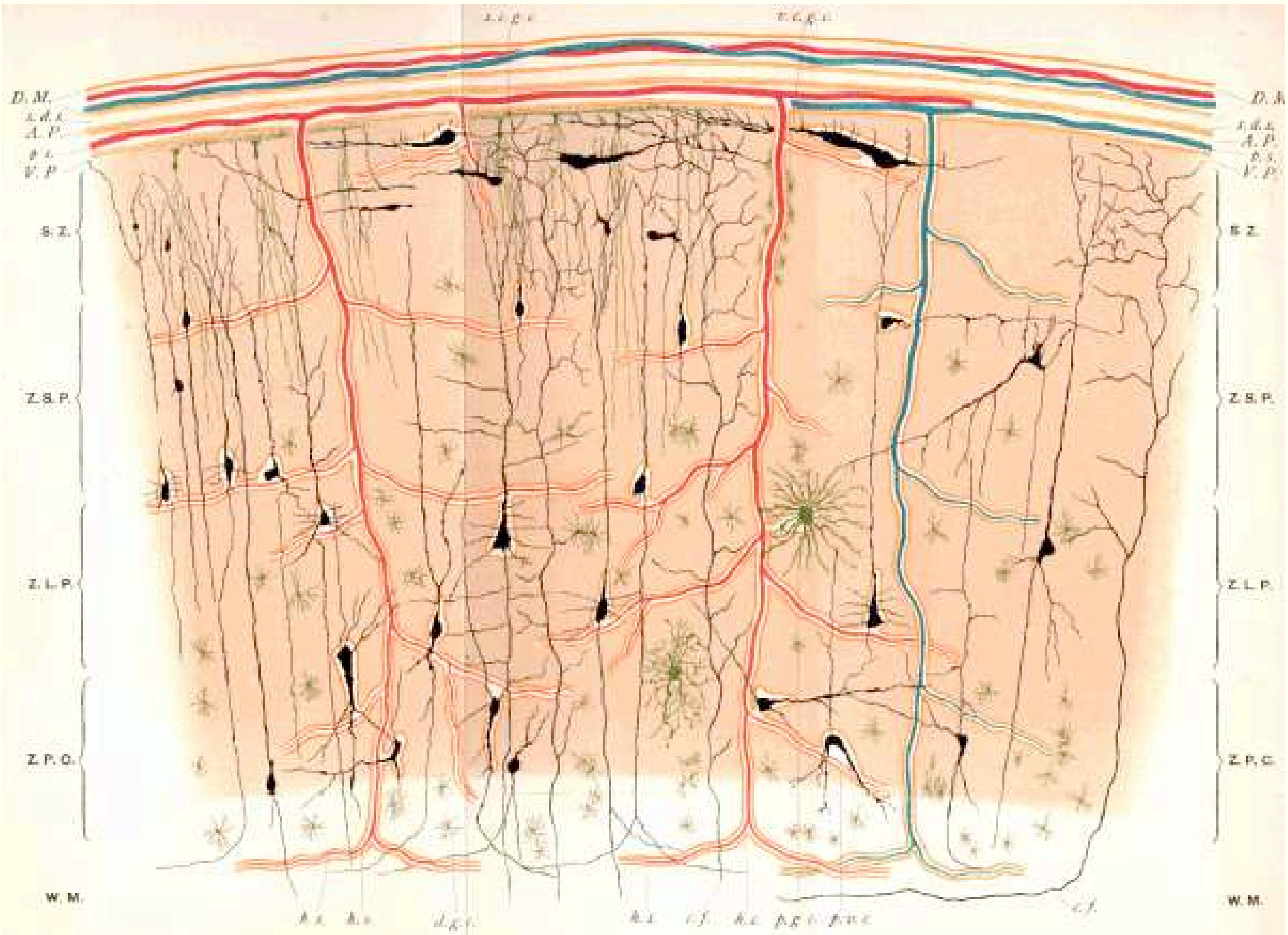
The spinal cord is the main nerve trunk, bringing all the sensory data of the body up to the brain and carrying the control signals back to the organs of the body. The medulla and the pons interconnect the brain and the spinal cord and cranial nerves (which carry information and control to and from the face and the surface of the head).

The midbrain ties together the visual and auditory systems with the nerves which control movement. The thalamus is the main relay station for the major sensory systems that project to the cerebral cortex, and as we shall see in the talk by Bernie Baars and Jim Newman the cortex projects back onto the thalamus to control what it receives from the sensory systems. The hypothalamus acts together with the pituitary gland as a master control system, releasing hormones which control the activities of other glands. The hormones then released by the endocrine glands feedback on the pituitary gland and hypothalamus regulating their activity. The hormones also probably have a great deal to contribute in the regulation of neurotransmitters and the overall modulation of brain activity. The limbic system seems to be mainly concerned with the complex analysis of odours and the appropriate responses to those data. The reticular system is what keeps us awake and aware of ther world, it is usually known as the reticular activating system.

Overlaying this whole complex is the **cerebral cortex** which

> "...makes humans what they are. Within the vast human cortex lies a critical part of the secret of human consciousness, our superb sensory capacities and sensitivities to the external world,

our motor skills, our aptitudes for reasoning and imagining and above all our unique language abilities." [Thompson, 1985, p22].

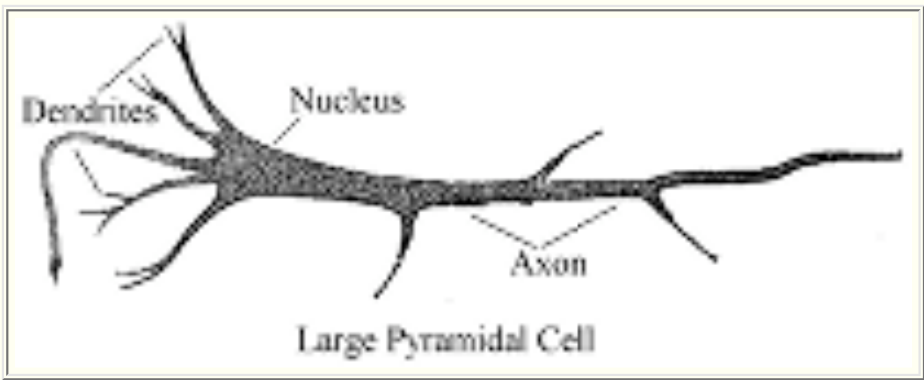


Golgi stained neural cells in the layers of the cortex (founded on plates by Ramon y Cajal, Retzius and Andriezen) showing large and small Pyramidal cells (black) with their vertical processes, Stellate cells (green), Horizontal cells (black, upper layer) whose processes transit across the cortex and afferent proceses from deeper in the brain (generally via the thalamus). [from J. Batty Tuke, *The Insanity of Over-Exertion of the Brain*, Edinburgh, 1894]

The Neuron

All information transfer in the brain and in the nervous system in general is mediated via **neurons**, in large bundles organised into pathways or channels. The neurons bring sensory data from both internal and external sensors about the state of the organs, the working of muscles, the perception of sound and vision, etc., which are all handled via bundles of nerves travelling into the brain (the **afferent** nerves). The control of muscles and one's general response to incoming stimuli are handled by bundles of nerves travelling out of the brain known as the **efferent** nerves.

A neuron is a single cell with quite specific architecture. Incoming information is transferred to the neuron through the synapses, small chemical detector bulbs on the end of a tree-like array of fibers, extending from the main cell body, called dendrites. The neuron is activated by impulses transferred to it, via the synapses, from the axons of many other neurons, and fires its own impulse response, through its own axon, when enough of the appropriate inputs have been activated. The details of nerve structure; myelin sheathing, membrane potential and transmission, etc.



Roger Penrose, in Shadows of the MInd has a nice description of the nerve transmission picture:

> ...the biological picture is of classical nerve signals travelling out from the central bulb (soma) of the neuron, along the very long fibre called an *axon*, this axon bifurcating into separate strands at various places. Each strand finally terminates at a *synapse* - the junction at which the signal is transferred, usually to a subsequent neuron, across a synaptic cleft. It is at this stage that the neurotransmiiter chemicals carry the message that the previous neuron has fired, by moving from one cell (neuron) to the next. This synaptic junction would often occur at the treelike *dendrite* of the next neuron, or else on its soma. [Penrose, 1994, p353]

Information processing takes place in the neurons and is a function of the numbers of connections of axons to dendrites via the synapses and whether or not the threshold for firing of that nerve is reached. The ease with which a synapse can operate is determined by the availability of neurotransmitter in the pre-synaptic bulb and the levels of neuromodulator molecules in the syanptic cleft as well as the excitatory or inhibitory nature of the synapse. All the excitatory and inhibitory synaptic junctions add to and subtract from the neuron's threshold trigger potential to determine whether it fires at any particular moment. There is some evidence to suggest that neuronal decision making takes place in the dendrites as well as in the averaging or thresholding of all synapses onto the neuron.

Two main areas dominated physiological discussion at the conference. One was the role of various parts of the visual system in contributing to perception and conscious awareness; and the other area related to the role of the thalamus as the hub of many aspects of brain processing. It is this latter material which I am going to concentrate on here. [For material on the visual system I suggest this link:
"Towards the Neuronal Substrate of Visual Consciousness" by Christof Koch.

and I also suggest Chapters 3-4 in Paul Churchland's book The Engine of Reason. The Seat of the Soul

The Thalamus

Bernie Baars and **Jim Newman** have developed a concept which for Baars is expressed in psychological terms as the **Global Workspace** or the working memory, and for Newman is expressed physiologically in their proposed neural corrrelate of working memory the **extended reticulo-thalamic activating system (eRTAS)** with its massive interconnections to and from the cortex.
Baars and Newman talk about the "Global Workspace"

I think that for me this was the most significant idea of the conference: this demonstration of the psychological: the working memory; being very closely coupled with the physiological: the eRTAS; as a really functional solution to the NCC problem. It is the extraordinary amount of interconnectivity between the thalamus and the cortex that allows the cortex, particularly the forebrain, to regulate the flow of information to itself: so that it doesn't get swamped, so that it can focus on what it needs to know about from moment to moment. This is a huge, self-regulatory feedback circuit which binds the processes of the brain, or at least those important to consciousness, together over time, within the time scale of our normal moment-to-moment operating frame.

[See the e-seminars led by James Newman for a series of papers and commentaries on the "Thalamo-cortical Foundations of Conscious Experience"]

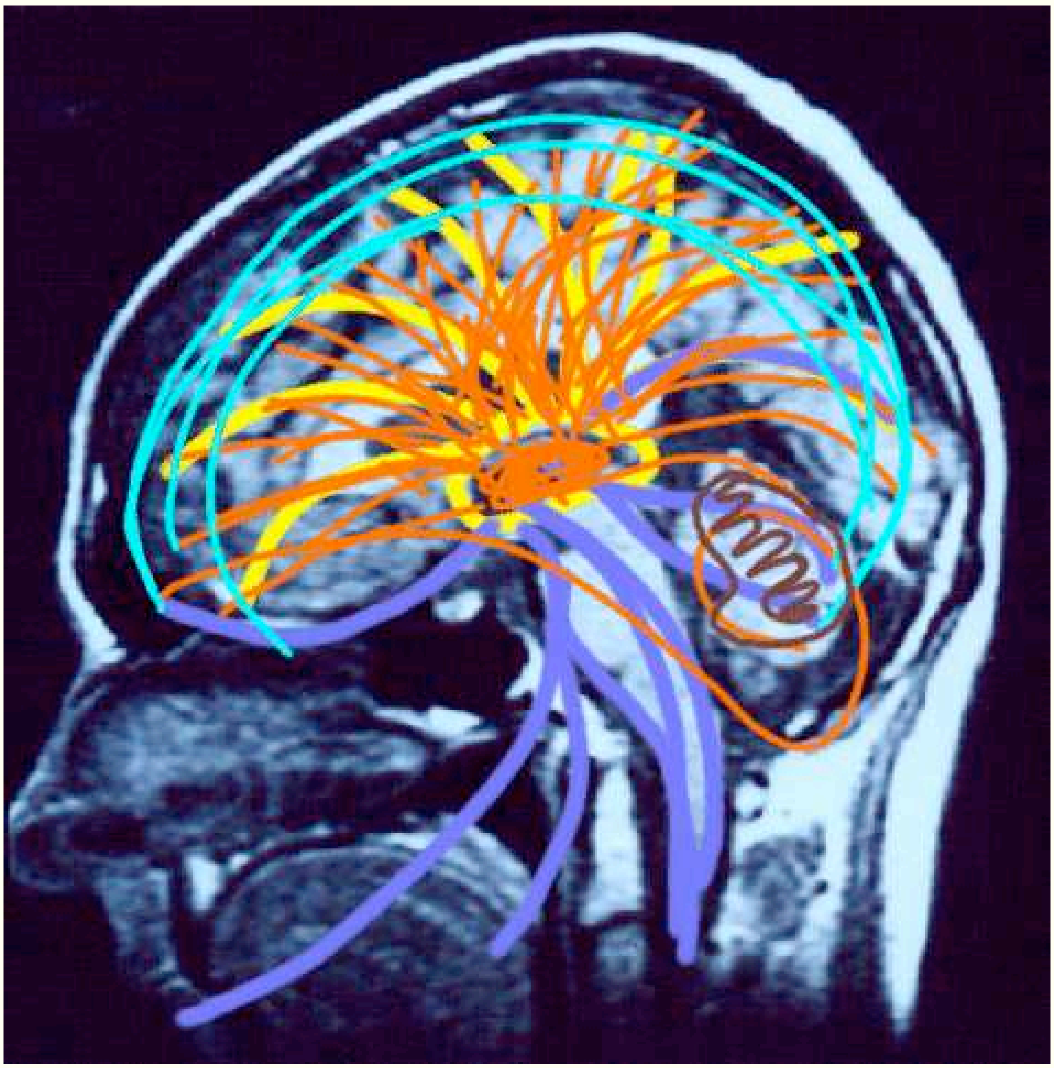
Now to return to **Joseph Bogen**. Bogen is on a search for the neural basis of "subjectivity". He suggests that any candidate for a neural correlate of consciousness must have widespread ditributivity, i.e. inward and outward connectivity. He refers to a subsection of the thalamus: the intralaminar nuclei as a site which satisfies his criteria as a primary site for subjectivity and he gives as primary evidence for this the fact that although quite widespread lesions in most parts of the brain do not cause loss of consciousness, merely modification of it, a very small lesion in the intralaminar nuclei will cause irrreversible loss of consciousness.

Rodolfo Llinas provided some detail on the physical shape of the interconnections through the thalamus as mentioned by Jim Newman. He argues that "the form and the way things are connected in space are very important" [Llinas' presentation to Tucson II] and makes the wonderful statement that:

"The brain is not a sausage, it's more like a well tuned musical instrument" [Llinas' presentation to Tucson II]

Llinas suggests that the view of the brain as a whole may give us an indication of how the brain works as a whole. He (as do Baars and Newman) describes the brain as being like the hub of a wheel. There is an array of radial projections carrying information patterns from all the afferent sensory systems upto the thalamus and projecting all the sensory pathways on up into the cortex. There are also a massive array of neural projections back from the cortex onto the thalamus. These project onto the array of gatelets of the intralaminar nuclei (see Bogen) and give the cortex control over what information it is getting.

Llinas has been using magneto-encephalography, (a vastly more sensitive brain activity sensing technology than electro-encephalography) to watch the brain functioning and has discovered bursts of 40Hz oscillation. These bursts are re-set when a sensory event occurs in an experimental situation and travel deep into the brain in a dialogue with the cortex. These continue while you are asleep. It is probable that dreams are the free-running of these oscillations through the thalamo-cortical system triggered only by the inherent noise of a biological system temporarily lacking the external sensory stimulus of being awake.



Neural Assemblies

Now we have all had experience of our degree consciousness varying over time, from being asleep through to being utterly and completely aware of the minutest details of things as when in some kind of emergency which requires our total attention. So consciousness changes in its degree over time. How can we describe this in physiological terms?

Susan Greenfield proposes a model of **Neuronal Assemblies** in the cortex which shrink and swell more or less on demand as the overall conscious situation requires.
See Susan Greenfield on "Neural Assemblies"

Alan Hobson has also developed a model of the ways in which consciousness changes over time. First he defines consciousness as

> "a graded integration of multiple cognitive functions yielding a unified representation of the world, our bodies and ouselves".

This integrated system and its unified representation also goes through a continuum of states as we go through our days and, in longer term, our years. So this model that Hobson has developed shows that:

> the **level** of consciousness changes as a function of activation;
> the **focus** of consciousness changes as a function of input/output gating; and
> the **form** (or perhaps the state) of consciousness changes as a function of modulatory neurotransmitter ratios.

He concludes that

> "Consciousness is the forebrain's representation of the world, our bodies and ourselves. It is always a construction whose level, focus and form depends upon the brain stem."

A speculative conclusion

It is my speculation that the neurophysiological evidence may be about to show up something quite significant in a distributed system which is conscious. The converging projections up to the thalamus from the sensory systems of the brain and the radiating projections from the thalamus up into the cortex are regulated by projections back from the cortex to the thalamus. This allows the cortex to have control over what systems' information it is being fed, thus allowing us to concentrate for example. (Such a difficult thing to do). This type of control is known as feedback control and is the basis of the metabolisms of all self-regulating (i.e. living) systems. The feedback can be reductive (or inhibitory or negative) and it can be expansive (or excitatory or positive) of any particular threads in the information flow.

Given that all processes take time to occur the fedback proportion of the output of the thalamo-cortical system will have been delayed with respect to the input information. In other words we have a short term memory, the previous state will be partly overlayed over the current state.

Also, given that we are in a condition of continuity (of continuous stimulus), even between waking and sleeping, there will always be new input being overlayed by previous input and so we live in that sense of continuity that we know as our temporal unity (one aspect of the unity of our consciousness). But this is a resonating system, it free-runs, it is an oscillating system. Llinas' 40Hz oscilations are one representation of the liveness of this resonating system. These are the most complex and the most organised resonances in one sub-system and will presumably be similar in other complexly organised layers of the brain. Like the visual system which keeps us in a generally seamless world.

References

Hubel, D. (1979) "The Brain" in Scientific American, September, 1979, pp38-47.

Newman, James (1997) "Putting the Puzzle Together: Part I: Toward a General Theory of the Neural Correlates of Consciousness." in *Journal of Consciousness Studies*, 4, No.1, 1997, pp47-66.

Penrose, R. (1994) Shadows of the Mind. Oxford.

Thompson, R F. (1985) The Brain. An Introduction to Neuroscience Freeman.

Nauta, W. & Feirtag, M. "The Organisation of the Brain" *Scientific American*, Sept. 1979, pp78-105

[See Poggio & Koch, SCI AM, May 1987]

Bernie Baars and James Newman

talk to Stephen Jones at Tucson II.

SJ: Bernie, would you like to talk about your concept of "Working Memory".

Bernie Baars: The background of this is that people have talked about consciousness and thought about it very seriously in very interesting and important ways for as long as we have recorded history, which is about 2500 years. There are Asian scriptures and meditation techniques and Plato and Aristotle and Socrates and so on and they all said very interesting important things about consciousness. Scientists have had a hard time dealing with it because they are dealing with the idea that inside each of our brains there is a separate reality. It's been problematic in a number of ways, partly because we don't know what to believe when people tell you about their experience. Now we've found over the last ten or twenty years that there are some things that people tell you when you ask them about their experience that are very consistent and the pattern of results has become so compelling and so strong that we can no longer avoid talking about consciousness as such.

I'll give you an example of that. If you repeat three or four words to yourself the way you might repeat a telephone number, and the words are "mother", "consciousness" and "science", let's say. If you repeat those words to yourself, what you find is that you're conscious of only one word at a time, but the other words will be unconscious at that particular instant and yet they will be immediately available. So you can always pull out "mother" even though you may not be thinking of that word at that particular time. Well that's interesting because that tells us that there's both conscious and unconscious aspects to this immediate memory task and today as it turns out we have PET scans where we can show that process taking place. The conscious component of our immediate working memory, which is the memory in which we talk to ourselves, basically, and it's the memory in which we interact with the world. It's about, oh, ten seconds long, its not very long at all, and its very limited in capacity.

Within that, consciousness seems to be involved with input: Putting stuff into that working memory that we have; with rehearsal and operations of various kinds, problem solving operations and also in output. The rest of it seems to be more or less unconscious. So its very inviting to take this very narrow "stage" of working memory and to think about it as a theatre. Whether there's a stage that's lit up or at least there's a spotlight that points at different parts of the theatre, pointing at different actors on the stage. There's an audience of 10,000 people or in the case of the brain, the audience has a 100 billion neurons, and then there's stuff happening behind the scenes. The Director's behind the scenes, and maybe the screen writer is behind the scenes and the people who make up the sets, and move them into place and so on. And these have tremendous influence about what happens on stage - but you can't see them, so they're unconscious, but they influence what is conscious.

So you have only three ideas basically, the stage, the audience, and behind stage, backstage, and together it turns out that these very simple ideas give us a fairly profound insight into everyday experiences like your perception of this particular event that you're seeing right now, your understanding through the language system, the tremendously complex language system in which you're taking in the words that you're hearing the words right now, and so on. So very profoundly complicated brain and mind mechanisms begin to make sense with this theatre metaphor.

So basically that's a simplification of a tremendously complex system. And simplification helps. We cannot understand things that are too complicated. But we don't want to lose sight of the complexity.

The whole effort in science is to use our own minds to understand the human mind. Well that's a bit of a problem because our own minds have this love for simplicity. We have a hard time dealing with more than 6 or 7 different elements, mental elements, whether they are words or numbers or ideas, at a particular time, so we need simplicity as human beings even in trying to understand something as complicated as the brain. And so the theatre metaphor is a way of simplifying things, that does not distort things, at least not in any obvious terrible way. And that's of course what we're looking for in science, we're looking for the cracks in reality, the places where we can separate out pieces of reality and come up with some simplicity out of this tremendous complexity. That's my little spiel on the theatre metaphor.

Working memory is an appendage to the global workspace. The global workspace is a very similar, a very simple kind of thing. It's basically the stage of the theatre. In order to deal with working memory you have to deal with the unconscious elements. When you're not thinking of "mother", (when) you're trying to rehearse "mother" and "consciousness" and "science" and so on... at that particular moment when you're not thinking about "mother" it's still available in some sort of memory and so you have to postulate the existence of a memory which goes beyond immediate consciousness.

JN: You can compare it to the wings? they're in the wings.

BB: They're in the wings. Or they're out of the attentional spotlight at that particular moment. The important thing is that we are dealing with a nervous system that is immense in complexity. And we know that it's highly specialised, so that the nervous system has 10,000 different audience members, some of them specialise in vision, some of them specialise in language, some specialise in emotion and so on, and you want to be able to reach all of those audience members at the same time, and some of those audience members we can think of as the working memory, holding things so that they're available to consciousness at any moment when we want it but not in consciousness.

The other thing of course that's so important in this whole scheme is some sort of notion of self. Because you have control over repeating those three words in your consciousness and working memory, and what does it mean to have control? Well it means for example that unexpected and unwanted words are not going to come in at any given time unless you want them to. In fact intrusive thoughts, intrusive words and so on characterise disorders of various kinds. Post-traumatic stress disorder, for example, one of the features that characterises post-traumatic experiences is this intrusion, i.e. sudden intrusions of feelings or thoughts that you don't want, and essentially there's a demand for consciousness. These very important things are happening in order to be able to adapt to the trauma that you may have experienced, and it intrudes on consciousness. Basically says may I come in, I must come in in order to deal with this very important problem adaptation. So short term memory and working memory are mostly under voluntary control but there are clearly times when something takes over that is so important that it has to come in.

James Newman: Such intrusive thoughts, those are rather rare and episodic in a person who has been traumatised, but what is particularly common in psychological trauma is something called "psychic numbing". This is this tremendous effort by consciousness to keep those painful feelings and memories from intruding and so the person's whole conscious psychology becomes one of avoiding anything that will cause them to re-experience the trauma. So that's just one example of how you can apply this Global Workspace metaphor to, for example, to psychopathology in this case or to traumatisation. But the other thing is that this model is also very useful in trying to understand how the brain actually instantiates consciousness. Of course the first thing that's essential to understand is that most processes in the brain are not conscious. As Bernie said, the brain consists in about a 100 billion neurons, a significant percentage of which are firing at the same time. And they're massively and reciprocally interconnected, so that it's hard to see how a single stream of consciousness with this very limited capacity of only seven, plus or minus 2, what we call "chunks" of information, can arise out of such a system. That's sort of the aspect that I've been working on with Bernie.

He actually first came up with a likely candidate for generating this stream when he was developing Global Workspace theory in the 1980's. He called it the "extended reticular thalamic activating system" and it is precisely an extension of the idea of the Reticular Activating System that was brilliantly elucidated by Magoun and Moruzzi, back in 1949. What these two neuroscientists discovered was that if you stimulate a portion of the very core of the brain stem you get an immediate arousal of the entire cortex, and the animal becomes alert and oriented to their environment. If they're sleeping they'll immediately come out of their sleeping state and start looking around, or if they're drowsy or quiescent they will suddenly become alert, they'll orient to something. Further research showed that this area in the brain stem was merely the beginning of an extended system that ascends, much like a fountain, all the way to the cortex and serves, in the most simple form I could put it, as the "searchlight" or "spotlight" on the theater of consciousness. It is what illumines the "play" of images that is taking place. Of course this is a very simple metaphor, just as the global workspace idea is a simple metaphor but from that you can elucidate what the spotlight is doing, and it turns out that really it's an immensely complex bank of spotlights. The thing that's so adaptive about these spotlights is that they can be controlled, in varying ways, by all of the systems and processors of the brain to ensure the spotlight is indeed illuminating what is most relevant to the person at that moment. Of course, the range of what is relevant to our consciousness can extend from the simplest crack of a stick in the woods causing us to immediately turn and look and see what is behind us; upto, at the highest level, our staying focussed on a very complex subject like I'm doing right now. I don't want to be distracted by the birds that are singing in the trees, I want to keep on our subject: the nature of consciousness.

This extended system begins in the reticular formation, in the core of the brain stem and continues to form the core of the brain as it goes up. The next "waystation" is the thalamus in the center of the brain. The thalamus turns out to be the communications hub of the brain. It's been known for many years that the thalamus relays all sensory information from the outside world. What is less appreciated is that the thalamus is really the main outside source of activation for the entire cortex. There are some other sources, but basically external activation of the cortex is a product of thalamic circuits.

One of the things that I wanted to look at when I was trying to instantiate this idea of the global workspace in the brain was, how does the thalamus create the "bottleneck" of narrowed information flow characteristic of moment-to-moment consciousness? In other words, how does the thalamus focus our attention in a selective way so that we only pay attention to exactly what we need to? Well, it turns out that the key to answering these questions is a rather peculiar little nucleus, or a pair of nuclei, called the reticular nucleus. Now, the thalamus is really *two* thalami and they're shaped like eggs. The reticular nucleus is something like the shell of these eggs. The curious thing about the reticular nucleus is that it consists of a vast array of neural "gatelets" that allow the thalamus to control the flow of information in the nervous system, and particularly between the thalamus and the cortex. The reason why this is so effective is that the cortex sends 10 times as many projections back down to the thalamus, and thus can regulate these gatelets in very sophisticated ways. Thus, we are able to consciously control the flow of information entering into our awareness or, in the model that we're talking about, the global workspace.

The most curious thing about this activation system is that when I say "we" control it, what "we" really means is that the vast audience of neural networks in our brains controls it. Essentially the control process is one of competition for access by these myriad networks to the global workspace. To use my example earlier, if a stick cracks behind me in the woods, all my awareness rapidly shifts to determine what it is behind me. In this case, a very primitive network of sensors specialised to pick up novel or potentially dangerous sounds takes over my awareness momentarily. Now, if everything that happened like that distracted me, my consciousness would not be focussed at all. But we have these cortical systems that then can inhibit that sort of orienting response so that I canin the present case, remain focussed on what I am talking about and not be distracted by the various sounds and sights going on around me.

As Bernie pointed out earlier, the mind is such a complex and multifaceted thing and our consciousness is so "simple-minded", that we need simplyfying metaphors for these complicated processes. A metaphor that I've used for this extended activation system is a "wagon wheel". If you visualise this, the hub of the wheel is the thalamus. All of the spokes converge on it. The wheel has got separate top- and bottom-halves. The bottom spokes carry sensory information to the thalamus. They also consist of inputs from this reticular formation that send projections to the reticular nucleus allowing the reticular formation to selectively open and close gates in response to novel or potentially dangerous sensory information.

So the lower spokes of this wheel consist of sensory pathways that funnel into the thalamus, with an exquisite set of projections onto the thalamic array of gates. This brainstem projection opens and closes the gates to orient us to relevant stimuli in the environment. The reason that this is possible, according to Arnold Schiebel, of the Brain Research Institute at UCLA, is that a "spatial envelope" exists in the core of the brainstem that serves to initially orientate our senses in the space surrounding our body. The odd thing about this is that this spatial envelope is an unconscious, and so its not that we're actually aware of what's going on in this brainstem area, but that what's going on in it can affect our attention, to orient us to what is important in terms of information coming in - new information or threatening information, or novel or dangerous information - coming in from the outer world.

Again, as I said earlier, if we only had that bottom-half of the wagon wheel we'd be constantly distracted by every sort of novel stimuli that came along, but the top part of this wagon wheel is essentially the cortex. It sends many more spokes back down to the thalamus after it has received all of that sensory information relayed by the thalamus, and it then modulates the flow of sensory information to itself. This is where the bottleneck originates, because you can only pay attention to so much at any particular time. And so the reticular nucleus, in a sense, becomes an essential bottleneck. The bottleneck is there so that we're not overwhelmed with all the sorts of information our brains process in parallel. We have a system in our heads that's generating thousands of bits of information per second, but we need to only pay attention to the seven or so most relevant bits of information. And that's what the stream of consciousness is: the most relevant pieces of information that we need to know from moment-to-moment.

The other marvellous thing about this system is that it integrates our experience into a unity. Paradoxically the complex processes of competition and co-operation mediated by thalamo-cortical circuits generate a global image, a unified image, in the workspace: the stream of consciousness. So we don't have separate sensations of a bird's chirp and the bird moving its wings in the trees. We see the entire object, indeed this process allows us, not only to see that object as a unified whole, but it also creates the background of our experience, so that the tree the bird is in forms the background of that conscious perception. And the bird that we're focussing on, including the chirp, movement and all the colours of its feathers, are all integrated into a common percept. We experience everything as an integrated whole. And this is, I think, the experiential basis of our sense of being an "I". This sense is not all there is to the self, certainly. It's only the most rudimentary beginnings of our self-image, but we have to have unified percepts in order to feel like like we're unified and the world is coherently unified.

Then what happens is that these percepts are in some marvellous way, synchronised temporally so that we have what William James described as a continuous stream of perception that's unitary in time. Out of that we have a foundational sense of being a single, coherent "mind". But it's really only the *conscious mind*. That's the irony, because it's somewhat of an illusion to think that what we have experienced consciously is all we are. Our consciousness is actually the tip of the iceberg of our mind.

Links and references

1. James Newman's e-seminars on **Thalamo-Cortical Systems**. Which includes a lot of links and papers, including Bernard Baars.
THALAMOCORTICAL FOUNDATIONS OF CONSCIOUS EXPERIENCE
led by James Newman, Colorado Neurological Institute

2. part of discussions in Psyche on Penrose's Shadows of the Mind
"Can Physics Provide a Theory of Consciousness?" by Bernard J. Baars. in Psyche, vol 2.

3. Baars, B.J. (1983) "Conscious contents provide the nervous system with coherent, global information". In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), Consciousness and self-regulation, 3, 45-76. New York: Plenum Press.

4. Baars, B.J. (1988) A cognitive theory of consciousness. Cambridge, UK : Cambridge University Press.

5. "What we really know about Consciousness", Bruce Bridgeman's (1996) "A thoroughly empirical approach to consciousness". Psyche 1.

6. Baars, B.J. (1997) In the Theater of Consciousness: The Workspace of the Mind. Oxford University Press.

7. Newman, J. (1994) "Consciousness Requires Global Activation", Commentary on "A Thoroughly Empirical Approach to Consciousness" by

B.J. Baars.

8. Newman, J. (1995) Review: Thalamic contributions to attention and consciousness. Consciousness and Cognition, 4:2, pp172-193.

9. Newman, J. and Baars, B.J. (1993) A Neural Attentional Model for Access to Consciousness: A Global Workspace Perspective. Concepts in Neuroscience, 4:2, pp255-290.

10. Newman, J. (1997) "Putting the Puzzle Together. Part I: Towards a General Theory of the Neural Correlates of Consciousness" in *Journal of Consciousness Studies*, 4, No.1, 1997, pp47-66.

To return to the discussion on the thalamus
To return to the discussion on organised systems

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

On Complexly Organised Systems and Consciousness

by Stephen Jones

John Searle: "Consciousness is not a Purple Glow"

Rodolfo Llinas: "The brain is not a sausage...it is more like a well tuned musical instrument"

The Emergence of Consciousness

The very complexity of the brain/body system may well be the basis for consciousness in itself. Steven Rose, in The Conscious Brain suggests that

> "consciousness has emerged as an inevitable consequence of one particular evolutionary strategy which has so far proved remarkably successful, that of the development of increasingly flexible and modifiable behavioural performance, achieved by increasing the size of the brain and the *complexity of the possible interactions of its components* (my italics)." [Rose, 1973]

In the study of the conscious brain organisation becomes the main factor. Just how is the brain organised? It has long been thought that the brain works as a hierarchy of controls. Hughlings Jackson in the 19th Century

> "saw a hierarchy of three levels, each of which controlled the one below...Evolution, he believed, progressively added higher

> levels of control" [Ferry, G. 1986, p41].

This position is at least partly derived from the three gross subdivisions of the brain itself.

> "The hindbrain or brain stem contains automatic mechanisms that regulate breathing, temperature and the circulation of the blood; the midbrain begins the processing of sensory information; and the forebrain is the seat of conscious awareness and voluntary decision making." [Ferry, op cit].

In modern neuroscience the hierarchical model begins to spread out, to flatten. Much of the brain's processing is done in parallel rather than serially from a top controller down. Various subsystems interacting through positive and negative feedback loops provide the control necessary to get things done accurately. It appears that the brain functions more through distributed networks of cells acting in concert than through localised functional areas.

Being in the world

We say "I am conscious" not "I have consciousness". Surely it is dynamic, continuous; a, so to speak, 'collection' of processings which one understands as being (for most people, most of the time) a single, unitary process. We even have a term for this singular process: my 'self'. To achieve this sense of unity or integrity this collection of processings is organised (i.e. self organised) over the growth and maturational process of individual becoming. We grow into consciousness of things about us and of our selves as an ongoing entity.

For the infant the culture (the enveloping 'agar agar' of language and family and socialisation) provides the nutrititive information in(wardly)forming the content of that organised thing, that collection of neurons, called brain. Inward and outward data-flow provides the stimulus for the exercise of neurons leaving some of the original massive oversupply for ongoing use and others to atrophy. The connectionism of the growing infant's brain is formed, 'wired down', by its envelope, its culture, its experience.

What we know of the world is a result of our being in the world. How we interpret what we are pesented with by the world is mediated by our history in the world, by what we consider to be 'natural' in the world, by what we have retained in memory.

What we think of as the 'natural world' is what we grew up with, the container, the environment, in which we developed and attained consciousness. The environment becomes invisible to us because it is like the air around us, so much at the basis of things, always with us, inseparable from us, our culture. In 1967, Marshall McLuhan, in The Medium is the Massage said

> "Environments are not passive wrappings, but are, rather, active processes which are invisible. The groundrules, pervasive structure, and overall patterns of environments elude easy perception" [McLuhan, 1967, p68]

In becoming we grow so accustomed to the information structure surrounding us that we don't even notice that it happened - the last thing a fish will discover is sea water. No matter how wierd and distorted the context within which an infant grows and is socialised, that context will still be the most 'natural' thing in the world, the most natural thing it knows. The social configuration and political system we live in is 'natural' until we are enabled to question it, until the environment is made visible.

Our personal story and the history of the world and our memory of it determine the way our interpetation processors in association areas of the cortex have been connected up by experience. Of course it is not a fixed result, we can, once we realise that we can, modify our interpretations on the basis of other knowledge, new knowledge, mulling over things and applied critical judgement. That is we can learn from our own reflection upon the contents of our referred or remembered consciousness. And this is the key that breaks the apparent determinism of what I have said above. We can reflect and feedback upon what we discover of the world and usethat to discover more, to come to grips with things and produce change and novelty. This is 'free will', this is the important consequence of Goedel's incompleteness theorem. We are essentially generative, there is always a new sentence to be said, a new tune to be sung. And this is perhaps the essential characteristic of conscious systems: that they are 'generative'.

It seems to me that one is a 'live' system living with an astonishingly complex flow of data and that one is in a sense, always at/on the knife edge of chaos. Like walking, which is a constant matter of falling and catching oneself, consciousness is a dataflow (the 'stream of consciousness' a la William James) maintained 'upright' by being continually/continuously caught from falling into chaos by the inhibitory neurotransmitter system (which governs the available noticeable datastream). If that collapses (too much LSD or succumbing to my schizophrenia) then chaos does ensue. Dataflow up and down the multiple feedback pathways of the brain regulated by the needs of the system itself: the need for information about what's going on out there? why am I feeling this pain? what does that red light mean? what is that patterned sound issuing from that person's mouth? etc. etc.

On Theories of Consciousness

What we're looking for in a viable theory of consciousness is something which accounts for the neurobiology and what has been shown within philosophy to be required for an awake and aware human being, but which also accounts for the fact that this is a continually developing/evolving system conferring a unity of "self" over a lifetime but which at the same time is never the same thing twice in a row, so to speak.

As I have said consciousness is a dynamic system and requires a theory of dynamic activity, in interaction with its environment. We cannot be a closed system without at least losing our humanity. One might suppose if an infant were to be abandoned at birth to complete isolation then (apart from the problem of its nutritional survival) it would possibly have nothing which we would recognise as consciousness. No ability to report on itself. Yes the capacity to start language acquisition is inherent but if it is not used, (stimulated by its surrounding culture) then it will atrophy and become eventually unrealisable and thus unusable. No recognition of others who shape and form us, no understanding of the emotion within, (perhaps even an atrophy of emotion). In short no consciousness.

So what should we say in the development of a theory which can account for our consciousness as we are discussing. We need to include a number of things:

> **1:** the tendency of a complex biological system to self-organisation
> **2:** hierarchical and heterarchical constructions
> **3:** generativity and the consequences of Goedel's Incompleteness Theorem
> **4:** some aspects of general systems theory - particularly open systems and self-regulation through feedback

So we will work through these requirements as they apply to both living and living/conscious entities.

Organisation

> "What distinguishes living from non-living matter is organisation. The operative term is the system, design or pattern, not the component parts, although these must be such as to enable the system to work." [Firsoff, 1967, p.7]

So what do we mean by 'organisation' ?
To distinguish what it is that we are referring to we should go back to one of the simplest possible systems of matter. The behaviour of a gas in a container is such that all the molecules rattle around more or less equidistant from each other and the gas settles down into thermal equilibrium, with the kinetic energy of the molecules of the gas being translated into some temperature which is dependent upon how many molecules of the gas were pushed into the container. Push more molecules of the gas into the container and the pressure goes up as does the temperature [Boyle's Law]. The molecules of the gas fall into thermodynamic equilibrium moving about randomly within the container. No one molecule of the gas is any more likely to be in any particular place in the container than any other. There is no particular **order** to the positions of the molecules of the gas. There is a minimum of information in the system, or a maximum of entropy. This is a **closed system**, very complex of description, to which the laws of thermodynamics apply, but lacking any structural caharacter which might be called order or organisation.

A completely opposite situation operates in living or **open systems**. There is a flow of energy and/or materials into and out of an open system, which maintains itself in a **steady state** which is far from equilibrium by chemical processes within (i.e. metabolism).

> **By necessity living systems are organised. The *particular relations* of any molecule or larger subunit (cells, organs and the like) are very important to the system and the system operates in a far from equilibrium condition high in information and low in entropy.**

In fact living systems are **neg-entropic** systems. Further, any group of similar open systems will show "equi-finality", that is they are able to reach their similar "final states" by diverse "metabolic" pathways whatever differences there are in the "initial conditions". For example this is what happens in cellular development.

So how do organised systems arise? The things we see as living and conscious are already complete organised entities, open-systems which are, so to speak, already "up and running". But how did this "bootstrap" occur? The old 'religious/vitalist' view that organisms come into being complete by the agency of some 'vital principle' (are put there by God) can be readily shown to be unnecessary. Complexes of molecular events can be shown to be "self-organising", able to set up and maintain all the pre-conditions and sub-conditions for a living open system.

Self-organising Systems

It has been shown by the Belgian chemist Ilya Prigogine that inorganic chemical systems can exist in highly non-equilibrium conditions and, under particular conditions depending on the particular chemical system being observed, can be self-organising.

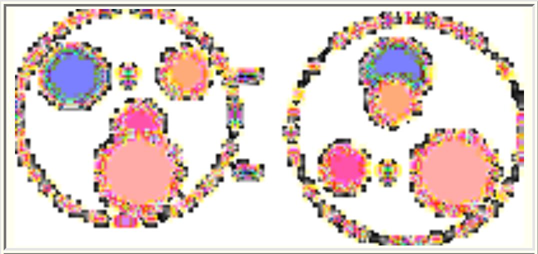
As we get more and more complex aggregations of chemical structures (as macro-molecules) we find that there is a tendency for these molecules to 'organise' themselves with respect to other molecules, still obeying the tendency to least energy consumption, in maintaining some sort of stable condition.

Stuart Kauffman in At Home In the Universe states

> "...that life is a natural property of complex chemical systems, that when the number of different kinds of molecules in a chemical soup passes a certain threshold, a self-sustaining network of reactions - an autocatalytic metabolism - will suddenly appear. Life emerged, I suggest, not simple, but complex and whole, and has remained complex and whole ever since...thanks to the simple, profound transformation of dead molecules into an organisation by which each molecule's formation is catalyzed by some other molecule in the organization." [Kauffman, 1995, p48]

Organic chemistry is the business of interactions between carbon and many other elements, most notably hydrogen, nitrogen, oxygen, phosphorus and a few metals. It is the organised interaction of these molecules that serve to act as living systems. Life is a function of, and finds utterly necessary for its sustainance, self-organising systems of molecules, which organise into the groupings of "metabolic" loops which we find in cells. Kauffman uses the term **catalytic closure** to indicate what brings 'life' to these loops. Kauffman again:

> "...when a large enough number of reactions are catalyzed in a chemical reaction system, a vast web of catalyzed reactions will suddenly crystallize. Such a web, it turns out, is almost certainly autocatalytic - almost certainly self-sustaining, alive." [Kauffman, 1995, p58]



An extremely simple autocatalytic reaction

An autocatalytic system works something like this: If one pair of molecules react to produce a third molecule, and this third molecule is useful as a catalyst in the reaction of some other pair of molecules, and this other reacting pair produce a molecule which is useful for catalysing the reaction of the first pair of molecules, then as long as supplies of all the reacting molecules last the system will continue as a self-sustaining loop.

These systems of molecules need feed-stock molecules which have to be brought in from outside and thus the system of auto-catalytic or metabolic reactions interacts with the environment. It is an **open system**, dependent upon its environment; such that if the environs can no longer supply any particular molecule needed for the reaction to continue, then the reaction stops.

If the complex of molecules is big enough as to be a cell then it will develop means for obtaining the necessary feed-stock molecules and so on. As the system becomes bigger it will become more organised towards its own sustainance. Life seems to have two characteristics here, the ability to obtain self-sustainance from the environment and the ability to replicate its system

Living systems operate in far from equilibrium conditions showing a maximising of information or a minimising of randomness or entropy. Organisms are systems which reduce entropy in a localised area.

> "An organism may be characterised as an enclosed, highly specialised chemical environment in selective interaction with its ambience, by means of which the features that ensure its specialised character are being constantly renewed. It is thus self-perpetuating and exists in a state of labile equilibrium. It is also capable of growth with a margin of adaptive variability, which allows selection to intervene as an evolutionary force." [Firsoff, 1967, p.7]

Dynamical Systems

In an article which appeared in New Scientist in 1987 Paul Davies speaks about non-linear complex systems:

> "The behaviour of nonlinear systems is enormously rich and diverse. When driven away from equilibrium, they are liable to leap abruptly and spontaneously into new, more complex or highly organised states. Alternatively, they may become chaotic. Often there are certain "singular points" where predictabiliy breaks down, the system becoming enormously sensitised to minute fluctuations. It is as if the system had a "free will" to choose between different paths of evolution, to explore new possibilities." [Davies, 1987, p43]

Living systems are non-linear dynamical systems which, although made out of perfectly materialistic physical stuff to which all the laws of classical and quantum physics apply, show emergent characteristics like self-organisation, as outlined above; complex metabolic self-regulation in cellular and organismic structures; as well as consciousness and cultural emergence. Such emergences as consciousness show well defined laws of dynamic behavior when looked at from certain perspectives such as cybernetics and psychology and psychotherapy. The work being done in exploring neural networks as models of the brain is an area where non-linear dynamics and self-organising systems are being used to great effect. To quote Paul Davies again:

> "We can model certain complex systems as an array of interconnected sites or nodes...[to] represent a brain with the nodes being neurones. The network is a dynamical system; each node can be in one of two states - 'on' or 'off'. Combined signals coming down interconnecting 'wires' determine the switching between the states. [the system is started in some state and evolved forward but this doesn't lead to impenetrable complication]. Remarkable things can occur. Coherent patterns of activity swirl around the net, organising themselves into stable cycles. Self-organisation may occur even from random input. Moreover the patterns are highly robust: severing a link in the net leaves them essentially undisturbed...these processes [may] relate to the neural activity of real brains". [Davies, 1987, p44]

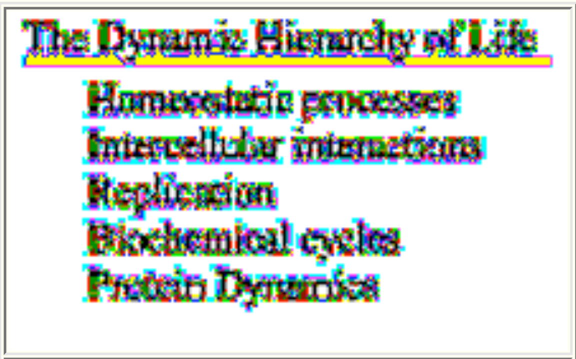
[See also Paul Davies' talk to Stephen Jones at Tucson II "Is a New Physics Necessary?"]

Living systems have always been very difficult to explain, especially while using reductionist procedures. In fact it is this very difficulty which led so many investigators to propose or accept "vitalist" explanation. The reductionist procedure of analysing the system into its component parts fails to see the system as an operating whole. It is totally necessary for an understanding of complex organised systems that they be seen as wholes because otherwise they cannot be seen at all.

Nevertheless, the reductionist procedure has produced a vast amount of useful data about the world and creatura living in it. We would be utterly unable to understand the world without it. The issue is simply that once you've broken a system down into its component parts you then have to gain an understanding of the relations of those component parts in producing an operating system. This is comparatively easy to do with non-living, engineering type, systems but not so easy with a living system.

Hierarchical Structures

At Tucson II **Alwyn Scott** author of Stairway to the Mind spoke on reductionism's failure to provide a full explanation of living systems and the value of looking at the hierarchies of organisation of biological systems and consciousness.



Slide from Scott's presentation

> "What we see here is a kind of a sketch of the hierarchical levels that are involved in the biological description of a living organism. There are many levels. There is an activity at the lower level: In the dynamics of individual proteins that are perhaps sitting in cell membranes in various places and mediated flows of ions in and out of cells and processing energy and things of that sort. At the highest level we have homeostatic mechanisms that are keeping our heart rates going evenly and our stomachs reasonably full and so forth. All of these kinds of things are going on in an organism. Furthermore each one of these levels and indeed more, constitute a branch of science." [from Scott's presentation to Tucson II]

These levels have an interdependence upon each other, the components of one level are made up of the whole entities of a lower level: the activities of groups of proteins determine the behaviour of a cell; the activities of a group of cells constitute the behaviour of an organ(ism); the interactions of organ(ism)s make up a body which in turn has to maintain a set of homeostatic processes in order to keep all the proteins and cells and organs functioning.

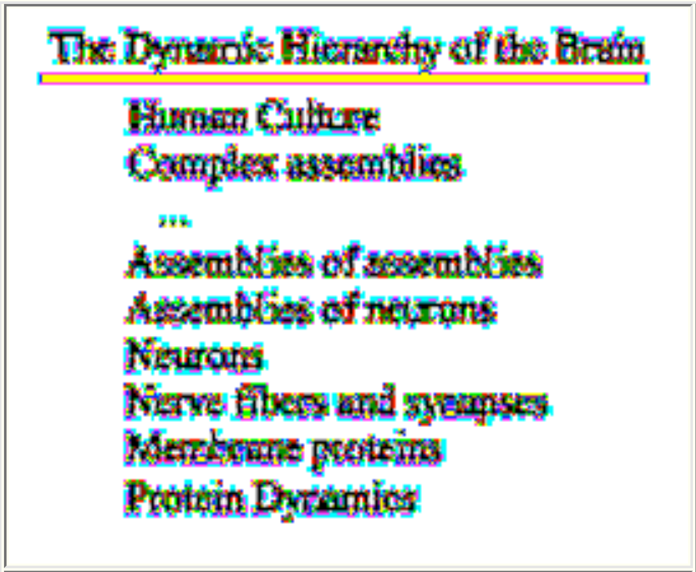
> At "each one of the levels, one could say that the time averages or space averages of variables at lower levels effect, more or less as parameters, what happens at that level. And at higher levels, essentially, variations and variables can be viewed as changing boundary conditions at lower levels." [from Scott's presentation to Tucson II]

There are myriads of "closed causal loops" [Scott] or feedback loops operating in and between all these levels which regulate the processes of each level to keep the whole system/organism going. Within an organism these are the processes of metabolic regulation: the available supply of necessary molecules will regulate some process and when that supply runs out then the organism might go seeking food to replenish the supply of that molecule. All the organs of a body are necessary for the full functioning of that body. If something is missing (through accident say) then other processes will be modified to compensate. There are cycles of cycles of inter-dependence in an organism all interacting to keep it going. The reductionist procedure cannot show these "hypercycles" [Manfred Eigen], they are more in the study of dynamical systems, or self-organising complexity.

Alwyn Scott again:

> "What is the physiological substrate of life? Is it the heart? Well if you remove the heart the organism will die. Is it, perhaps, the iron atoms in the haemoglobin molecules? If you remove all those, or change them to nickel atoms the organism will also die. Is it the lungs, or is it the microtubular cillia at the very basis of the lungs keeping the gases moving around? Either one of those. Does the question make any sense to ask for the substrate of life? Where there is a phenomenon that somehow emerges from all of these hierarchically oriented interacting dynamical levels." [from Scott's presentation to Tucson II]

He then went on to refer to similar configurations in the brain:



Slide from Scott's presentation

> "Once again we have a number of distinctly different levels of dynamic activity ranging all the way from the level corresponding to the dynamics of protein molecules that are embedded in nerve membranes in various places, all the way up to the level in which the brain itself is interacting dynamically within the culture in which it has developed. Within the culture in which the brain, the mind, the personality, had come of age and become acculturated. Once again, all of these other levels in between, all of these other dynamical levels each one interacting with levels above and levels below." [from Scott's presentation to Tucson II]

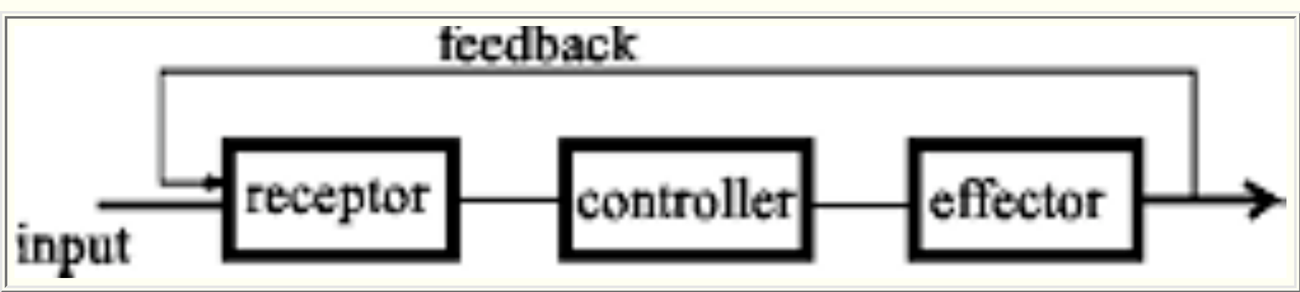
Assemblies of nerves, processing for example the low level geometric features of vision, get together with other assemblies, processing say the wavelength-determined aspects of light, and feed up to higher level assemblies which integrate the features into explicit object identifiers which then feed interpreters which were originally set up by cultural interactions onto the brain. So again we have myriads of loops of relations between levels. Only now these refer to informational stuff, phenomenal stuff, emotions and the activities of a person within a culture. All of these loops need to be teased out, but are not recognisable if removed from the context of their generation/operation. A reductionist approach ceases to be useful in this task.

The neural network is holistic. It is the pattern, not the state of any node, that is of interest. This pattern is not reducible to the activity of individual nodes, but is the "collective behaviour of the whole array". [Davies, 1987, p44] As Alwyn Scott said we cannot reduce mental processes to the electrochemistry of neurons, and thence to particle physics, we have to look at the sets of relations of system and sub-systems to find the basis for the phenomena we are looking at. It is a whole system which shows the phenomena of consciousness. A whole system operating via laws of self-organisation, self-regulation through feedback processes. A large scale complexly organised system operating in time and through time with relations of information to its environment as an open-system.

Feedback

Let's look briefly at the detail of feedback. The output of any sytem of effectors or processors when fedback into the input will have an effect which is determined by the relative value of the returned proportion and its delay relative to the input signal. For example, take the relatively simple system of an electronic amplifier: when a portion of the output of the amplifier is returned to an inverting input to the amplifier then this feedback controls the tendency of the amplifier to amplify, i.e. one is subtracting some of the signal from the overall gain of the amplifier, giving it some stability in its effect. This is inhibitory or negative feedback. If the feedback portion is returned to a non-inverting input the amplifier will turn into an oscillator, i.e. the feedback will add to the original signal and the gain will "go through the roof". This is the microphone in front of a speaker effect to which the term "feedback" is most commonly applied. This feedback effect is positive or excitatory.

Further, the output of an amplifier which is fedback to the input is necessarily delayed compared to the input. This is what is known as propagation delay and any circuitry, be it electronic or neural, will entail propagation delay through it. In an electronic system, our amplifier again, the amount of delay in the feedback can actually turn what would have been a negative feedback into a positive feedback, e.g. a sine wave delayed by half a cycle will be inverted with respect to the input sine wave and when this now inverted wave is subtracted from the input wave the subtraction of a negative signal results in an additive event which can easily become oscillation. This results in a situation in which the system may, very loosely, be said to go "live".



Feedback is used for the stabilisation of a system (the feedback is used to "damp" the behaviour of the system), for direction control in steering a system (the feedback provides a measure of the difference between the current heading of the system as against the desired heading) and for "homeostasis" or sustaining the metabolism of a biological system. We would perhaps be most familiar with feedback in ourselves (though it is a largely unconscious process) when we think about the process of picking something up: we continually adjust our aim towards what we are picking up by a measure of the difference between where we are now and where we need to be when the grasping should take place. [For more on feedback systems see the discussion of Ross Ashby's homeostat. Also see Norbert Wiener's Cybernetics (1948)]

Basically, feedback is used in the ongoing regulation of a dynamic system. In living systems there are a multitude of feedback pathways all carrying information about the (momentary) condition of some part of the system, be it metabolic or informational. In the brain there are massive arrays of feedback pathways (often called "descending pathways") from many stages (sub-systems) of sensory processing to their source stages,regulating what is actually being fed up to the current stage. This particularly applies to a system called the thalamo-cortical system. [see **James Newman**'s e-seminars on Thalamo-Cortical Systems.]

> To speculate briefly: **In a very complex system with a multitude of feedback loops relating to different neural pathways, operating with a variety of delays inherent such that the whole system will be in all kinds of states of inhibition of some signals and excitation of others, will inevitably have some pathways which are oscillating and reverberatory such that the whole system may be said to be "live".** [See my paper On Animation: The Illusion of Life. (1988)]

The Brain is an Organised System

Bernie Baars and **James Newman** have developed a concept which for Baars is expressed in psychological terms as the **Global Workspace** or the working memory, and for Newman is expressed physiologically in their proposed neural corrrelate of working memory the **extended reticulo-thalamic activating system** with its massive interconnections to and from the cortex.
see the Baars and Newman talk

What Baars and Newman and others (viz. John Taylor, Rodolfo Llinas) are uncovering is a physiological neural system which operates as a large-scale organised structure using open-system entrant data - perceptual data - and closed-loop feedback re-entrant data. Perceptual data from internal or external sources feeds up to the thalamus where it is allowed access or made available to the cortex by the cortex's own determination of its needs. This is handled by descending pathways of neurons and is essentially a feedback control system. And because of all the propagation delays involved there will be an element of recurrence of older information in the current state of the system - this is at least short-term memory, Baars' working memory.

> Again, to speculate: **What is it like to be inside this thalamo-cortical feedback system, which we, each of us, inevitably are? Is this what it feels like to be (conscious) ?**

Development of Consciousness

I contend that as an infant develops any impact on the body, limbs or muscles stimulates the afferent nerves. This stimulation initiates myelination of the nerve axon, and also asserts in the brain the existence of the fibre and its mapping into whichever cortical area it is involved with. Adjacent fibres carrying stimuli (signals) from adjacent areas on the sensing surface are similarly myelinated and mapped into the cortex. The infant will at first only sense inchoately. As more input occurs mappings will be consolidated and refined.

At the same time efferent nerves are carrying signals which initiate movements, the nerves are myelinated by their use and the muscles are stimulated to develop in their ability to respond. As muscles move they impact with external objects and stir internal proprioceptive sensors, thus returning signals to the appropriate sensory cortices which carry feedback on the muscle action. In the brain nerve processes from the sensory areas feed data to other areas of the brain including direct and higher level motor control areas.

When the infant hand impacts upon an object the grasping reflex leads to attempted interaction with the object which further stimulates the sensory systems. If grasping is not possible, say because of the size of the object, then other action will tend to take place to compensate in some way, by, for example, opening the hand more. The activity of reaching and touching stimulates the nerves in the arms to grow and myelinate and stimulates the differentiation and mapping of the nerves in the brain which handle the sense data feeding it into the control centres differentiating and mapping these nerves. This provides a substrate for finer control over the muscles.

A feedback loop of refining control by successive approximation results, our aim becomes better and better; and soon we are reaching, holding, pulling, crawling and so on better and better. Maturation of the brain takes place in direct relation with maturation of the body.

We can map onto the cortex the projections of the sensory organs. Information from specific senses maps onto specific areas, visual information projects to the rear of the brain in the visual cortex, auditory information projects into the auditory cortex, the body surface senses map onto the somato-sensory cortex as a representation of the sensory importance of various areas forming a map known as the sensory homonculus. [see Penfield]

If we generalise this process over both time and modalities we can develop an idea of the brain being connected up synaptically on the basis of external experience: the impact of the environment, culture, family, etc.etc.; internal experience: proprioceptive and kinesthetic responses; as well as reflective activity: mulling things over, wondering about things; and generative activity.

References

Davies, P. (1987) "The Creative Cosmos" in New Scientist, 17 Dec, 1987.pp41-44.

Ferry, G. (1986) "The Egalitarian Brain", in New Scientist, Jan.9, 1986, p41.

Firsoff, V.A. (1967) Life, Mind and Galaxies, Oliver & Boyd

Kauffman, S. (1995) At Home In the Universe, Oxford.

McLuhan, M. (1967) The Medium is the Massage, Random House

Nicolis, G. & Prigogine, I. (1989) Exploring Complexity, Freeman.

Prigogine, I. & Stengers, I. (1984) Order out of Chaos, Heinemann.

Rose, S. (1973) The Conscious Brain, Weidenfeld and Nicholson.

Scott, A. (1996) Stairway to the Mind.

von Bertalanffy, L. (1968) General System Theory, Brazillier.

Wiener, N. (1948) Cybernetics: or Control and Communication in the Animal and the Machine. MIT Press.

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

The Philosophy behind Ordinary Consciousness

Part 1: Pre-20th Century

by Stephen Jones.

Some history of the philosophy of mind:

Descartes to William James leading to the current setup of the debate.

The first thing we notice when we talk about consciousness is that we, as humans, come in two conditions, either awake or not (viz. asleep, knocked out, in a coma, or dead), either conscious or unconscious. Leaving the various states of not being conscious aside, we are asking, in these papers, what is it about humans and probably other organisms that constitutes being conscious. How is it that we are conscious? Is it some sort of externally applied thing or is it somehow a function of being an organism? [Note: I will tend to use the term "organism" to stand for a complexly organised biological entity which exists as, and only as, a whole.]

Within the western world up until the late Renaissance and the extraordinarily productive period of the 17th century, the human was simply a product of God, a unitary being, the pinnacle of a hierarchy of being ordained by God and nourished by the Church.

In the background of all this stands the classical Greek civilisation. For Aristotle

> "It is the soul which guides unconsciously the processes of growth and reproduction, in both plants and animals; and which in animals is responsible for sense-impressions, for memory and for desires, which pass over into actions." [Wilkie, 1958, p23].

Thus the soul was probably more a description of a type of organised (perhaps organising) entity which we now call the self and was not seen as something separable which continued on alone after bodily death. The relationship between the soul and the body was more a relation of form to matter than that the soul was in some intrinsic way separable from the body.

That other word which nowadays is so often a synonym for soul, i.e. spirit (or spirits, plural) was more of a vapourous fluid which the body breathed and internally distilled to give nourishment and heat. The idea of "consciousness" *per se* doesn't seem to exist, but its nearest analogy is in the idea of the 'intellect' which covers our abilities to differentiate and to deal with abstractions. It's here in this ability to abstract (using the 'active intellect') that Aristotle seems to refer to a part of the soul which is somehow divine or ideal.

It had more or less been established that the brain was the seat of consciousness by the end of the 16th century. For example the English physician Robert Burton in his The Anatomy of Melancholy describes the state of knowledge ....

> "The brain itself is divided into two parts, the fore and hinder part; the fore part is much bigger than the other, which is called the little brain in respect of it. This fore part hath many concavities distinguished by certain ventricles, which are the receptacles of the spirits, brought hither by the arteries of the heart, and are there refined to a more heavenly nature, to perform the actions of the soul. Of these ventricles there are three - right, left, and middle. The right and left answer to their site and beget animal spirits; if they be in any way hurt, sense and motion ceaseth. These ventricles, moreover, are held to be the seat of the common sense. The middle ventricle is a

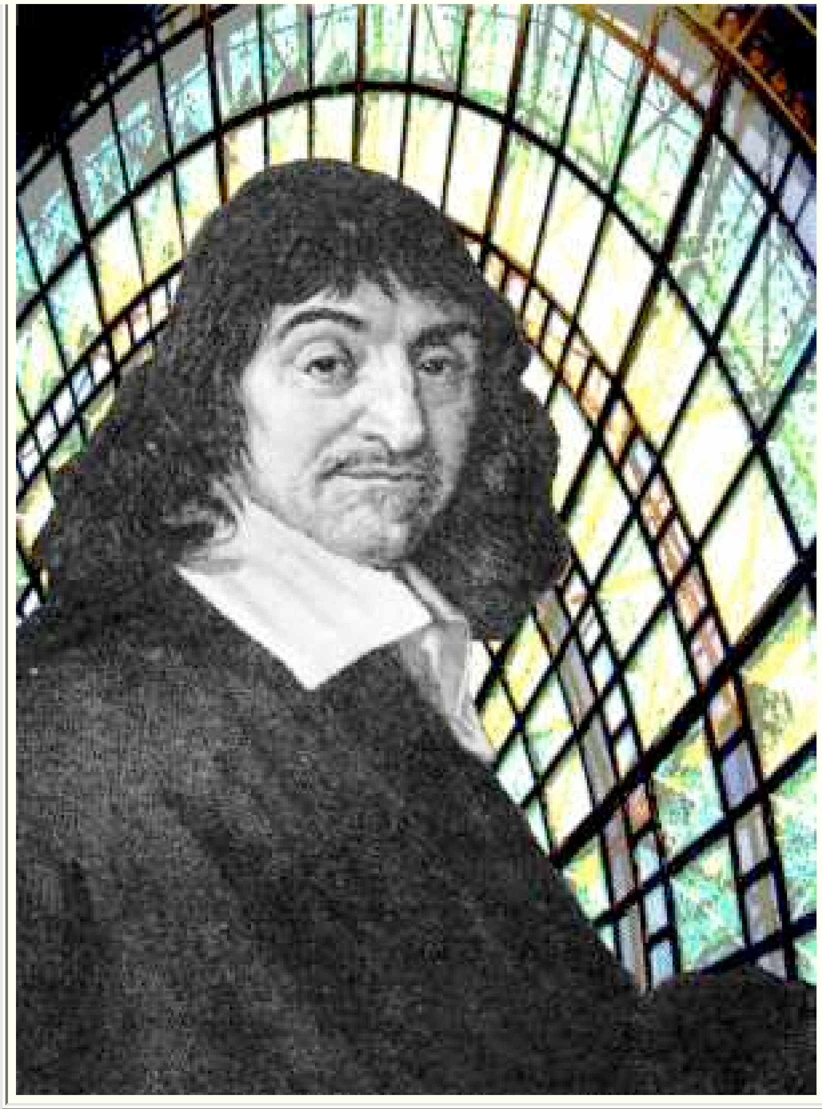
> common concourse and cavity of them both, and hath two passages - the one to receive pituita, and the other extends itself to the fourth creek; in this they place imagination and cogitation, and so the three ventricles of the fore part of the brain are used. The fourth creek behind the head is common to the cerebral or little brain, and marrow of the back bone, the last and most solid of all the rest, which receives the animal spirits from the other ventricles, and conveys them to the marrow in the back, and is the place where they say the memory is seated." [Burton, p97]

We can see in the above the prevailing paradigm of the Galenical humours coupled with a, by now Church modified, Aristotelian soul as being the guide for thinking about what was being exposed by anatomy in the newly emerging science of the 17th century.

Rene Descartes

In the 17th century, with the rise of a mechanistic description of the solar system and the animal body, among other things, the French philosopher **Rene Descartes** developed a philosophical method for exploring questions of the 'rational soul', the self and the body; such things as sensation and perception, thought and ideas, etc.

He did this by employing a process of doubting



Rene Descartes comes to Los Angeles (with apologies to Frans Hals)

everything which he could not find an incontrovertable reason for believing. Ultimately, using a thought experiment based on the proposal that there could be a being which was capable of deceiving him on every possible matter, he came to the conclusion that there was only one thing the truth or existence of which could not be doubted and this was himself as a disembodied thinking thing. For if the deceiver were able to deceive him about everything to the point of shaking his belief in himself then he, as an entity being deceived, still existed. So that whether he was or was not being deceived about this there was still something that might be being deceived.

From this Descartes produced the famous *Cogito ergo sum*. "I think therefore I am". Having established that there was a thinking thing which he knew as himself, he then had to discover where this thing lay. Now, given that everything he knew about the body had material existence and could largely be explained by mechanistic reasoning (even though this would take a further three centuries) he was left with one issue which was inexplicable in these mechanistic terms, this deceivable entity, this thinking thing (the *res cogitans*). What could it possibly be and where could it possibly reside? So Descartes, given the prevailing paradigm and being a good subject of the Roman Church and having heard of the fate of Galileo and others before him, assumed that this thinking thing was made of some ineffable, non-extended stuff and resided without (outside of) the physical realm (probably as a kind of aura, a purple glow). But how was it to control the body for which it was responsible, what was the mechanism of the tie between the two? Descartes' proposal for this was the Pineal gland which is a single gland lying in the center of the head, the task for which was to convert the immaterial thoughts and desires of the soul to material acts (the opening and closing of valves within Descartes' view of bodily mechanism) and vice versa the converting of bodily sensations and feelings to the immaterial form needed by this soul.

One thing Descartes did usefully establish was the recognition that the sensations transmitted to the consciousness from the outside world were not in fact images or

> "...faithful copies of the objects. Nothing can come from external objects to our mind by the medium of the senses, he said, *except certain corporeal movements; but neither these movements themselves nor the figures arising from them, are concieved by us such as they are in the organs of sense*." [Descartes: Dioptrics in Riese, W. 1958, p133]

Descartes, though a great scientist of his time and well versed in the mechanical analogy for the description of the brain and nerves (he illustrated the action of the nerves by using the analogy of the hydraulic systems of automata then in great favour for entertainments in the pleasure gardens of the kings and princes of Europe) was unable to accomodate the *res cogitans*, the "thinking thing", within the housing of the body and having gotten himself to a point where his radical **doubting** thought experiment appeared to leave him with no option, neatly severed the mind from the body, an act from which we are still recovering. Thus was **Dualism** engendered to become the source of the Mind/Body problem which has vexed the minds of philosophers till today.

It seems to me that Descartes' *thinking thing* is pretty much identical with what we know as consciousness today. So it is Aristotle's 'intellect' and Descartes' 'cogito' which are our first real inclusions for the concept of consciousness.

Descartes more or less isolated the *res cogitans*, the mind or consciousness from the brain/body. He estabished the idea of the mind or consciousness, but what of its contents?

John Locke

This is the matter of where Thoughts and Ideas come from? The English empirical philosopher John Locke examined the idea of the mind and demonstrated that there is no intrinsic, *a priori* knowledge built into the mind at birth.

In his An Essay Concerning Humane Understanding, first published in 1689, Locke says of ideas:

> Every Man being conscious to himself, That he thinks, and that which his Mind is apply'd about

frontispiece portrait of John Locke

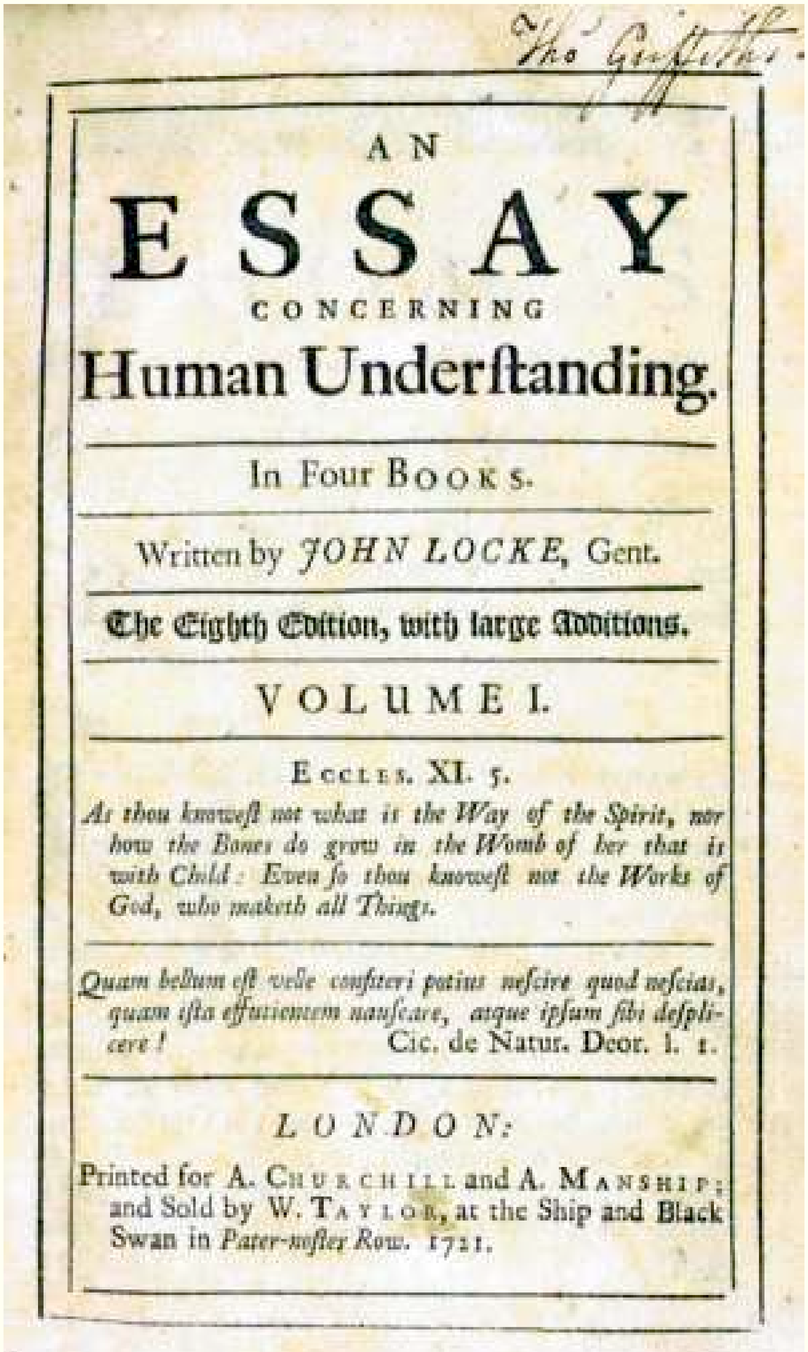
whilst Thinking, being the *Ideas* that are there, 'tis past doubt, that Men have in their Mind several *Ideas*, such as are those expressed by the Words, *Whiteness, Hardness, Sweetness, Thinking, Motion, Man, Elephant, Army, Drunkenness*, and others: It is in the first Place then to be enquired, How he comes by them ? [Locke, 1721, p67]

He notes that received Doctrine is that ideas are innate, and then refers to his prior demonstrating that the concept of innate ideas doesn't work; for, he asks, why then does an infant not arrive fully knowing the world?, and why are there then idiots who are unable to know such things as right and wrong?, and why is it that all people of the world do not have the same ideas? Having shown that innate ideas don't exist he then proceeds to analyse where ideas come from.

> Let us then suppose the Mind to be, as we say, white Paper, void of all Characters, without any *Ideas*; How comes it to be furnished ?...To this I answer, in one word, from *Experience*...Our Observation

imploy'd either about *External sensible Objects; or about the Internal Operations of our Minds, perceived and reflected by our selves*...These Two are the Fountains of Knowledge, from whence all the *Ideas* we have, or can naturally have, do spring. [Locke, 1721, p67]

These two sources he calls **Sensation** for ideas from external sources and **Reflection** for ideas from internal activity of the mind. Sensations are derived from the senses and


Tho^s Griffiths

AN
ESSAY
CONCERNING
Human Underſtanding.

In Four BOOKS.

Written by *JOHN LOCKE*, Gent.

The Eighth Edition, with large Additions.

VOLUME I.

ECCLES. XI. 5.
As thou knoweſt not what is the Way of the Spirit, nor how the Bones do grow in the Womb of her that is with Child: Even ſo thou knoweſt not the Works of God, who maketh all Things.

Quam bellum eſt velle confiteri potius neſcire quod neſcias, quam iſta effutientem nauſeare, atque ipſum ſibi diſplicere! Cic. de Natur. Deor. l. 1.

LONDON:
Printed for A. CHURCHILL and A. MANSHIP; and Sold by W. TAYLOR, at the Ship and Black Swan in *Pater-noſter Row*. 1721.

Title page of the 8th edition of Locke's primary work (from the copy in Stephen Jones' library)

have **qualities** such as yellow, white, heat, cold, etc. Whence the modern philosophical jargon word **qualia**. Reflection is

> ...that Notice which the Mind takes of its own Operations, and the Manner of them, by reason whereof, there come to be *Ideas* of these Operations in the Understanding. [Locke, 1721, p68]

Locke asserts that these are the sole sources of ideas in our minds, and that experience of qualities is necessary for their knowing:

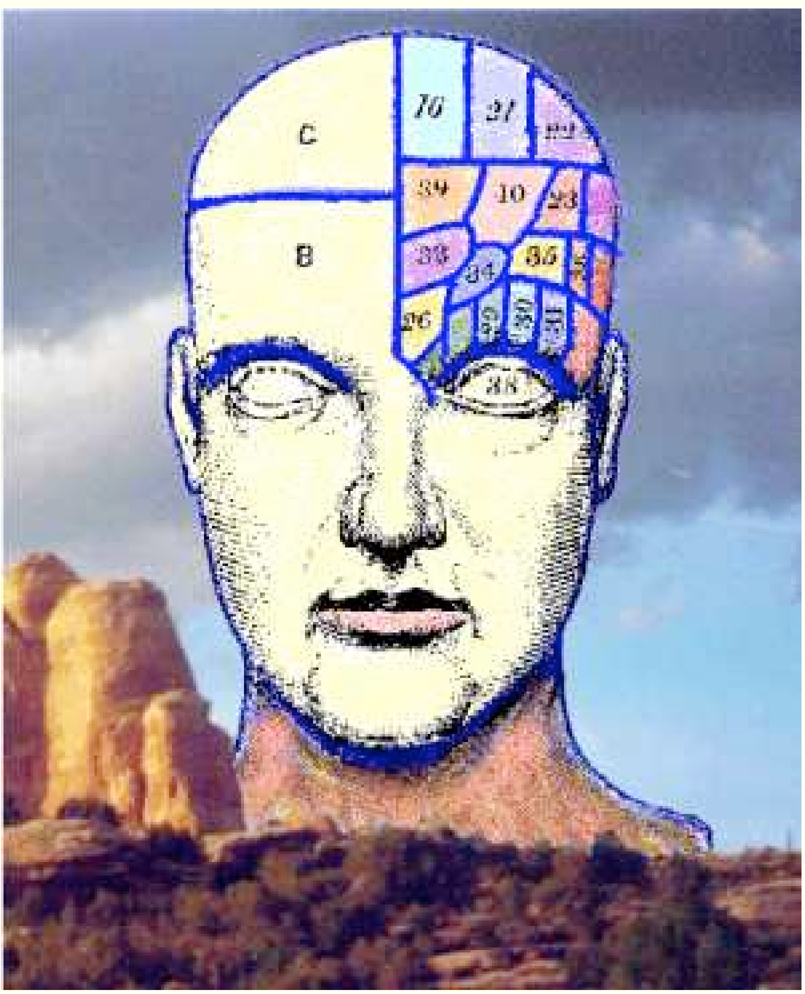
> ...if a Child were kept in a Place, where he never saw any other but Black and White, till he were a Man, he would have no more *Ideas* of Scarlet or Green, than he that from his Childhood never tasted an Oyster, or a Pine-Apple, has of those particular Relishes. [Locke, 1721, p70]

He suggests that thought is a function of the mind:

> ...the Perception of *Ideas* being...to the Soul, what motion is to the body, not its Essence, but one of its Operations. [Locke, 1721, p70]

In summary: Locke introduced the idea that a newborn infant's mind is *tabula rasa* or a blank slate and that experience is necessary for the building up of the contents of the mind. These contents have two sources: the Sensations as being from experience and having perceived qualities; and the Reflections of our minds upon their contents. He then proceeds to engage in the analysis of the kinds of Ideas we may have, be they simple or complex. But we won't pursue him along this line.

Franz Gall



Franz Gall, in the 18th century, began the process of localisation of function in the brain. The faculties were considered innate and the function of particular organs of the brain. He maintained "that the brain was composed of as many organs as the individual had faculties, tendiencies and feelings." [Ackerknecht, 1958, p150]. It was this approach which led Gall's assistant Spurzheim to later develop the ludicrousness of Phrenology. Nevertheless Gall discovered a great deal about the anatomy of the brain, placing the main faculties in the cortex, established the idea of nerve pathways and established the division between the grey matter as the matrix of the nerves and the white matter as the conductor function.

And what of the modern view of consciousness? During the 19th century evidence accumulated through study of the consequences of the excision of differentiable parts of the brain in animals and through the investigation of what amounted to natural experiments in people caused by disease and accident that the brain could continue to operate despite the loss of various parts of its substance. This slowly led to a view that the mind dwelt in the whole of the brain not in a particular anatomical location and that consciousness was one function of the brain.

James Mill

In the 19th century James Mill in his Analysis of the Phenomena of the Human Mind (1869) argued that:

> **"Having a sensation, and having a feeling, are not two things. The thing is one, the names only are two. I am pricked by a pin. The sensation is one; but I may call it sensation, or a feeling, or a pain, as I please... And (similarly) to say I am conscious of a feeling, is merely to say that I feel it. To have a feeling is to be conscious; and to be conscious is to have a feeling. To be conscious of the prick of the pin, is merely to have the sensation... To feel an idea, and to be conscious of that feeling, are not two things; the feeling and the consciousness are but two names for the same thing. In the very word feeling all that is implied in the word Consciousness is involved." [Mill, 1869, p224]**

William James

William James in his Text Book of Psychology, 1892:

> "**The immediate condition of a state of consciousness is an activity of some sort in the cerebral hemispheres**... One has only to consider how quickly consciousness may be abolished by a blow on the head...by a full dose of alcohol...to see how at the mercy of bodily happenings our spirit is... Destruction of certain definite portions of the cerebral hemispheres involves losses of memory and of acquired motor faculty of quite determinate sort...Taking all such facts together, the simple and radical conception dawns upon the mind that mental action may be uniformly and absolutely a function of brain-action, varying as the latter varies, and being to the brain-action as effect to cause." [James, 1892, pp5-6]

James pointed out
1: that we have a sense of a personal consciousness, that it is ours, not something that we share. [This may have had interesting consequences in philosophy in that this attitude tends to ignore the idea that consciousness (or the self) is a construction, a result of a process of socialisation/acculturation, and thus producing the "problem of other minds"]
2: that our states of consciousness are always changing
3: that we have a sense of temporal continuity in consciousness (my past can only ever be my past) which leads to a conception of a "**stream of consciousness**", a function of memory, and
4: that it is selective of what it pays attention to.
We may carry a number of thoughts "in mind" but we attend to them one at a time. Further, there exists a "fringe" of properties attached to that thought which are (analogous to) its relations to the world.

> "**States of consciousness themselves are not verifiable facts**....Everyone assumes that we have direct introspective acquaintance with our thinking activity as such, with our consciusness as something inward and contrasted with the outer objects which it knows. Yet...whenever I try to become sensible of my thinking activity as such, what I catch is some bodily fact...It seems as if consciousness as an inner activity were rather a *postulate* than a sensibly given fact, the postulate...of a knower as correlative to all this known."[James, 1892, p467]

References:

Ackerknecht, E.H. (1958) "Contributions of Gall and the phrenologists to knowledge of brain function". in Poynter, F.N.L. (ed.) The Brain and its Functions. Blackwell Scientific, Oxford, 1958

Burton, R. The Anatomy of Melancholy London, 16??

James, W. (1892) Textbook of Psychology

Locke, J. (1721) An Essay Concerning Humane Understanding. London

Riese, W. (1958), "Descartes's Ideas of Brain Function", in Poynter, F.N.L. (ed.) The Brain and its Functions. Blackwell Scientific, Oxford, 1958

Wilkie, J.S. "Body and Soul in Aristotelian tradition", in Poynter, F.N.L. (ed.) The Brain and its Functions. Blackwell Scientific, Oxford, 1958

Continue to Part.2: The 20th Century or

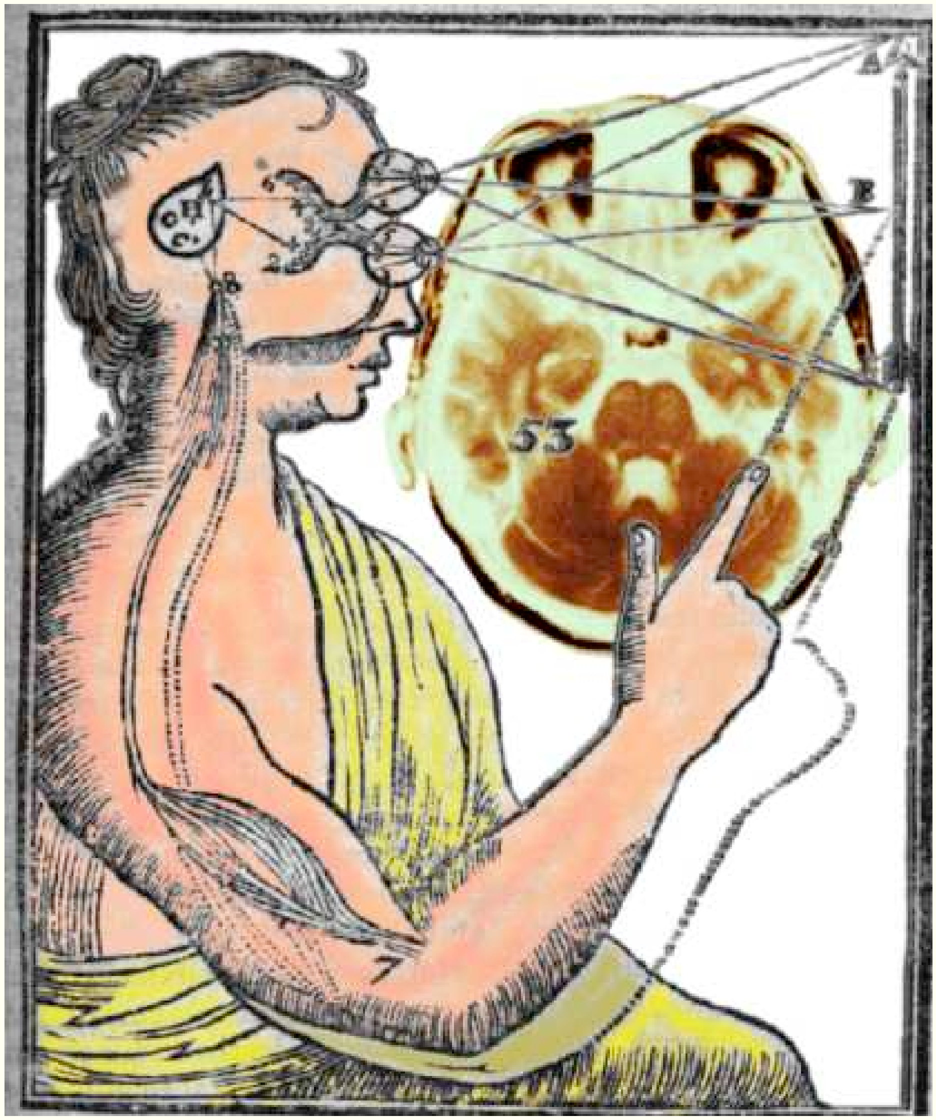
 **return to Chapters index**	 **return to Philosophy index**	 **return to overall index**	 **return to Introduction and textual index**

The Philosophy behind Ordinary Consciousness

Part 2: The 20th Century

by Stephen Jones.

What is Consciousness?



Just what is the definition of consciousness? In simple terms it is the subjective, my self, the I, that seems to have some existence over and above the body. We view it as the body's driver, controller and also reason for being. Consciousness is an essentially private matter, I know that I am conscious and I presume that you are but it is very difficult to prove (this is what is known as the problem of "other minds"). This privacy of consciousness then makes it very difficult to define and describe in an externally consistent manner, consistent with other peoples' descriptions of their Consciousnesses.

At the Tucson conference **John Searle**, had this to say about the definition of consciousness:

> **The Commonsense definition of consciousness:**
> **- Subjective qualitative states of awareness, sentience or feeling.**
> **- A biological phenomenon that is intrinsic to certain biological systems.**
> [slide from Searle's presentation.]

Consciousness is **subjective** and there lies the "$64,000 question". **How do we reconcile this subjective phenomenal business with the objective physical stuff requisite for its continuance?**

Well, the first thing to do is to figure out what consciousness does for us and what to include, from the repertoire of things that organisms do, in consciousness.

In his presentation to the conference **Paul Churchland** showed a slide listing some of the things we might consider as aspects of consciousness.

Some Salient Aspects of Consciousness **1) Short-term memory (with decay)** **2) Directable attention** **3) Multi-valent comprehension through "mulling"** **4) Independence from sensory inputs (e.g. in daydreaming)** **5) Disappearance in sleep** **6) Mundane or prototypical character of dreamed scenarios and sequences** **7) Unity across the senses** [slide from Churchland's presentation]

Nowadays we think of awareness (i.e. being awake and aware) as one necessary aspect of consciousness. The ability to perceive and to respond to those perceptions has long been regarded as the primary indicator of consciousness. This active response is considered as being over and above the mere 'irritability' of the senses and the body's normal deep level responsiveness (or reflexivenss) to sensations which may or may not then have conscious impact.

To get matters going about what consciousness does for us **Robert Kirk** of the Department of Philosophy, Nottingham University, has proposed what he considers to be a basic set of requirements for the attribution of 'consciousness' to an entity. This is his **"Basic Package"**, a set of behaviours that all conscious beings have but which is lacking in non-conscious creatures and inanimate objects. The Basic Package consists in the capacity of a creature to gather and use information for itself in the modification of its activities and behaviours in dealing with the world.

text of Robert Kirk's talk on "The Basic Package"

The information gathered by the conscious system for its own use is the content of our conscious minds. It is to this that the term **qualia** (the qualities of a thing, e.g. the sense of its colour, (its 'accidents' in the older language)) is usually applied (we will deal with qualia shortly). We know of our consciousness by the stuff that we perceive and know of the world; the phenomena of the world. All that we know as individuals is contained in our brains in some manner. The study of this stuff of consciousness is generally known as Phenomenology.

Consciousness and the Brain

Now, consciousness is but one function of the nervous system but it is definitely the most mysterious. We have found the physiological correlates of sensations in the visual pathways (Hubel & Weisel; Crick and Koch) and the audtiory and other sensory systems. We have elucidated the somato-sensory cortex (Penfield) and the organ that triggers our awakeness in the reticular activating system (Magoun and Moruzzi) and as we see in the discussion of neuro-physiology we now know a great deal about a possible integrating mechanism which may provide a sort of working memory system and which is thus a neural correlate of the activity of consciousnesss as an ongoing process, in the thalamo-cortical structures (Baars and Newman; Taylor; Bogen). But do all these largely discrete 'Neural Correlates of Consciousness' (the NCC) provide us with the integration that is necessary when we are confronted with the unitary nature of consciousness?

So, what is the relationship of the physiology to consciousness? On the one hand we have this brain in its body, and I would contend that we should really know it as this **intelligent or conscious body**. We have the anatomy of the brain, the neuronal pathways and their physiology, the actions of neurons and synaptic transmission. We have the growing detailed knowledge of the particular information processing pathways involved in vision, hearing and arousal (i.e. being awake), speech and planning. We know some of the complex chemistry of neurotransmitters and neuromodulators; the molecules that respectively mediate inter-nerve transmission (i.e. the interconnections of nerves at the synapse) and modulate the ease of such transmission and interconnection (i.e. affect the thresholds to be reached for the firing of the next nerve in the pathway). Much of the chemistry of neurotransmitters and neuromodulators is directly related to emotional states and affect.

We know quite a lot about the physiology of the brain, but, on the other hand, we still seem unable to tie the mind into this brain. A range of questions on the relation of the physiology to the mind arise:

> 1: Is consciousness simply a function of the brain physiology? This is the materialist or physicalist position.
> 2: Is it something which somehow exists independently of the brain having no real causal relation to the acts of the brain (i.e. as some kind of epiphenomenon)?
> 3: Does it exist independently of the brain but directly apply or effect its desires and acts upon the brain? This is "Dualism", the legacy of Descartes, and is probably most akin to the religious view. Or
> 4: Is consciousness in some way a field of what... force? quantum physical wave/particles? some extra-physical stuff (to which the modern use of the word 'spirit' is applied)?

And further about the "nature" or the "stuff" of consciousness:

> 1: What is this mind made of, this conscious thinking thing that I know of at least in my self as 'my self'?
> 2: How did this happen? How did it evolve?

> 3: Do we need a new or an extended physics? a new kind of "field" or "force" to explain it? [See the article on On Quantum Physics and Ordinary Consciousness] Or perhaps,
> 4: Is consciousness a necessary function of a highly complexly organised system of physical material? [See the article on On Complexly Organised Systems and Consciousness. See also my paper "On a hypothesis for how the Brain is Consciousness"]

And if this latter is possible,

> 1: What then of the possibilities for developing a non-carbon-based, intelligent or conscious artificial device or system?
> 2: That is, for example, is a conscious silicon computer possible?
> [See the article on Neural Networks and the Computational Brain. See also my paper "On the possibilities for a Conscious Machine"]

We know that consciousness cannot exist without the brain, at least in the living, physical world. So what is the relationship between the brain and its consciousness? To reinvoke Robert Kirk's question: How is it that an object can be a subject? This is what the conference at Tucson was really trying to elucidate.

Chalmers and the "Hard Problem"

Which brings us to the question which formed the basis of most of the discussion within the philosophical aspects of the conference and spread itself far and wide into the physiological discussions, and into which the physicists dived to provide their versions of a solution. This is the matter of the **Hard Problem** as described by **David Chalmers**: Given all that we know about the physiology of the brain and its processes (i.e. the neural correlates of consciousness and the unconscious processes) do these generate consciousness or is there a need for some other agency? This was probably the main question of the conference.

In Chalmers' video talk he discusses his division of the question of how it is that we are conscious into the hard and the easy problems. He argues that delineation of the anatomy and physiology of the brain, the description of, say, the visual system or the systems of speech, the physiological pathways of pain, etc, no matter how difficult to carry out are all soluble and therefore of the class of "Easy" problems. But, given all the physiology and so on, Chalmers claims that this still does not explain how it is that we have a subjective view of the world and ourselves. So the Hard problem is: Where does this subjectivity come from? How does all the physiology produce subjectivity? Chalmers speculates that there may be two aspects to information, a physical aspect and a phenomenal aspect.

Chalmers discusses his view in David Chalmers' talk on "The Hard Problem" and for his central paper on the Hard Problem see Facing Up to the Problem of Consciousness

Chalmers' way of positing the problem of whence arises consciousness is an essentially Cartesian act of separating off from the physical data the, perhaps, single thing which we find most difficult to reconcile within it, namely, the mental, phenomenal world of consciousness.

Responses to Chalmers

During the conference there was a great deal of debate over the implications of the Hard Problem as formulated by Chalmers. To start with: is it a useful thing to do to separate the question up in this way? **Patricia Churchland**, Professor of Philosophy at the University of California, San Diego, and co-author, with T. J. Sejnowski, of The Computational Brain, asked this question in her presentation.

First she disputed the idea that the problem of consciousness is somehow well defined and really does stand separate from the rest of the so-called "easy" problems. By separating out the problem of consciousness from all the "easy" problems, it makes it possible to conceive of a situation in which

> "there could indeed be a theory of all of those things and still we wouldn't know what it was for the... for the what? for the light to go on?" [Churchland, 1996a].

Perhaps short-term memory or awareness might "fuzzy-up that boundary". She indicates doubts about the role of consciousness in a number of what are very intelligent human processes, such as the ability to maintain an upright stance or "intelligent moving of the eyes. It is in some broad sense purposeful, but lot's of us are unaware of the specific intelligent movements that we make." [Churchland, 1996a].

> "So when David Chalmers suggests that in a way for those problems we can see how a solution can go, but for this problem we haven't the slightest idea, I'm inclined to disagree. I think there are many profound puzzles about the nature of brain function that we have yet to solve and thinking of consciousness as the lone enigma here will, I suggest, lead us to look for kinds of solutions that may not be all that useful." [Churchland, 1996a].

So, is the division between the "hard" and the "easy" problems a good or even useful division?

> "The next thing that worries me about this is that it suggests that the hard problem, we can already see, is going to have to have a real *humdinger* of a solution. That it's going to have to be really radical. That it's going to have to come from somewhere really deep like *quantum mechanics*. That it can't just be a matter of a complex dynamical system doing its thing. Yes there are emergent properties such as rhythmicity and emergent properties such as memory retrieval but this is an emergent property like unto no other and we can already see that you're

> going to have to have a radical solution. Well I can't actually see that!" [Churchland, 1996a].

For Chalmers to argue that because consciousness is a very mysterious thing that it must then be unexplainable is to confuse the difficulty of working on the problem with an explanation of the problem:

> "...the mysteriousness of a problem (and I grant you, how the brain works, in general, is mysterious) is not a factor of the problem. It isn't a metaphysical feature of the world, it's an epistemological fact about us. It's about where we are in current science, it's about what we can and can't, at this stage of our knowledge, understand. It's about what we, given the rest of our beliefs, can and cannot imagine. It isn't a property of the problem itself." [Churchland, 1996a].

And finally the distinction between the "hard" and "easy" problems tends to mean that those who are working on neurophysiological problems may be discouraged from ever suggesting that some particular results they have may provide some insight on the problem of consciousness. Which makes the division rather counter-productive.

> "Notwithstanding that, I recognise that there may be very useful things also about the distinction. It's just that on balance I feel very uneasy about it." [Churchland, 1996a].

For Patricia Churchland's full paper from the Tucson presentation see The Hornswoggle Problem

John Searle also made a brief comment on the Hard Problem. To wit: we know consciousness happens; so it must be like many of the other hard problems in science, awaiting further discovery. Thus he too raises the point that Pat Churchland does that it may simply be that there is more work to be done within the directions we have now. My contention will be that this work applies particularly within the understanding of complexly organised dynamical systems.

Qualia or the Contents of Consciousness

We then get into difficult questions about the contents of consciousness, which are usually known in the jargon as **qualia**. Qualia are such things as the colour of a rose or a feeling of pain or the sound of a train whistle. Any and all of the myriad perceived aspects of the world which our brains provide our consciousness. A great deal of debate went on as to whether the idea of qualia is useful and just what is the relation of qualia to consciousness.

Daniel Dennett professor of Philosophy at Tufts University and author of Consciousness Explained had some very important things to say about qualia.

Dennett says he is trying to rid the discussion of the notion of qualia because the term has been used in so many different ways that it is impossible to get a consistent usage for it. He would have us simply use the phrase "the contents of consciousness" instead. The effect of the term qualia is to induce a kind of dualised view of brain processes which is precisely what allows consciousness to be hived off from the physiological and turned into such things as Descartes' *res cogitans* or Chalmers' "Hard Problem".

Dennett spoke of "The Myth of Double Transduction" in theories of consciousness. He argues that the concept that qualia are in some sense (perhaps physically) independent of the processes of the brain is to badly misunderstand what is going on in the physiology. The fact that we interpret what the brain presents as being somehow *other than* the patterns of data that the brain presents is a function of our psycho-social histories, the evolution of our *memes*: the interpretive forms and structures which we receive from our socialisation, the process of our growth into language-using *human* beings. To return to Dennett:

> **"Since you are nothing over and above the various subagencies and processes in your nervous system that compose you, the following sort of question is always a trap: "exactly when did I (as opposed to various parts of the brain) become informed (aware, conscious) of some event."** [slide from Dennett's presentation at Tucson II]

Which leads to the next problem that this dualising produces: just what is this "I" to whom these qualia are presented? As Dennett says, it is an empirical fact that there is no little man (homunculus) in the brain (in the Cartesian theatre) to whom qualia are presented. If there were we would simply have to start all over again with understanding how this homunculus works. Qualia are a linguistic attitude about the contents of consciousness. He then re-presented a comment from Michael Levin:

> **"Essentialists may think, and think that materialists think, that conscious states are appearances of brain processes... But materialism is not a double-aspect theory. It does not construe mental states as how neural processes appear. Appear to whom? Materialism construes mental states as appearances, not of the nervous system, but of the world impinging on the nervous system"** [Michael Levin, "Tortuous Dualism" Journal of Philosophy, 92, p323]

Dennett's argument is that what we are is not something extra, a "self", but

> "**what you are in fact, just is all this organisation of competitive activities between processes in your brain and your body**." [from Dennett's presentation]

Dennett showed some video of a demonstration of an effect in which a change in qualia is undetectable. If the change in the qualia, e.g. a change in the colour of someone's pullover, happens during the essentially ballistic period of a saccade of the eyes from one point of attention to another, then the qualia change will be undetectable to most people, implying to a large extent that a memory of the previous qualia state is not being stored. Dennett seems to imply that this means that the things we call qualia don't exist, nothing is stored in, so to speak, the "real-time" of consciousness, or at least of conscious vision. Or does this mean that there is no visual consciousness, as distinct from some more general consciousness, and that it is only what is fed up to general consciousness (Bernie Baars, "working memory") that somehow become qualia? So Dennett says he wants to drop the term qualia and simply say the "contents of consciousness".

John Searle comments on qualia such: "conscious states are qualitative states, if you carve off qualia you don't have anything left." [Searle's presentation at Tucson II]. With which, my guess is, if he said that as "... if you carve off the contents of consciousness..." Dennett would agree. The brain can't help but produce these contents and it is largely a matter of what we say about them, how we re-present them to others, which sets up our belief structures about what they are and their (the contents') relation to the brain. (and so on around the loop... the bootstrap of consciousness).

In Conclusion

Chalmers suggests that information may have a double aspect. That it somehow has not only the normal physical presence, but also has some non-physical presence which allows information to exist independently of any embodiment. But in a physical world we would be hard put to find any kind of information that is not embodied, even the most tenuous ideas still require embodiment in at least one brain to exist and be reported in any way. So information is stored in brain states and other physical states, like magnetic particle orientation on a hard disk or printing ink on paper.

It is a cultural artefact that we see any information we have as being somehow not embodied (it may be damned difficult to find all the details of the brain states that are the current information state of the brain, but anything we know must have this embodied condition). Descartes' separation of the mind from the body has largely allowed a condition of thought to arise which makes such a distinction and thus reinforces the apparent non-physical condition of information. Now I recognise that ideas are very infectious and travel from brain to brain with great ease and similar epidemiology to the medical variety of infectious diseases. But this does not in any way suggest that ideas somehow have existence independent of the brains that carry them. The wonderful thing about brains is their very plasticity and susceptibility to ideas and new information. And this is the point: information can be accomodated in brains but it is as the 'shape' of the information (the patterns of data), the organisation of that information. We don't just have some microwave-like resonant cavity in the brain which is the information antenna or store, all the information is stored in patterns of data, (patterns of activation: a la P.M. Churchland). So the idea that ideas have independent existence is a culturally generated idea from within which we look at the meta-data of what the data of information are. We have been told that information is non-physical therefore all our percptions about information are that it is non-physical (which is obviously circular). Yes information is organised, but it is organised by and in our brains, and its continued existence depends upon those brains. The point being that the phenomenal aspect of information, i.e. the contents of consciousness, is another kind of representation within the brain of the patterns of data running through the brain. That ever mysterious "subjectivity" is simply the brain's way of being inside, being made of, its information structure. "I am the dataflow". There is no hard problem, though there are heaps of very difficult problems. The hardest problem is coming to grips with the idea that the flow of data through this highly complex system, with all its reverberent activity is all we need to be conscious.

References:

Chalmers, D. The Conscious Mind Oxford, 1996

Chalmers, D. "Minds, Machines and Mathematics" Psyche 2:1, 1995

Chalmers, D. "The Puzzle of Conscious Experience" Scientific American 273: pp80-86, 1995

Churchland, P.M. The Engine of Reason, the Seat of the Soul MIT Press, 1995,

Churchland, P.S. Neurophilosophy MIT Press, 1986,

Churchland, P.S. and Llinas, R. The Mind-Brain Continuum MIT Press, 1996.

Churchland, P.S. and Sejnowski, T. J. The Computational Brain MIT Press, 1992.

Dennett, D. Brainstorms MIT Press, 1978

Dennett, D. The Intentional Stance MIT Press, 1987

Dennett, D. Consciousness Explained Little, Brown, 1991

Kirk, R. Raw Feeling: A Philosophical Account of the Essence of Consciousness. Oxford, 1994

Searle, J. The Rediscovery of the Mind MIT Press, 1992

Searle, J. "Minds, Brains, and Programs." In: D. R. Hofstadter and D. C. Dennett, The Mind's I: Fantasies and Reflections on Self and Soul Penguin, 1982.

 **return to Chapters index**	 **return to Philosophy index**	 **return to overall index**	 **return to Introduction and textual index**

Robert Kirk

talks to Stephen Jones at Tucson II.

SJ: So if you'd like to outline the basic package stuff and then we can extend it from there.

Robert Kirk: Right, well, what I'm interested in is the fact that we and other animals perceive the world consciously. We see the colours and shapes of things, we smell, we're sensitive to all aspects of the world. And there's something it's like for us to see the world, hear and smell and taste things. As there is, many of us think, for cats and dogs, horses, cattle, chickens, and no doubt lots of other animals too. Now, this is a fact but there's something puzzling about it, because we and the other animals are physical systems, really. Complex physical organisms and there are lots of other physical things around us, tables and chairs, and things, but we don't think that they perceive the world.

Now we can do a lot of things differently from inanimate things like tables and chairs. Is it just the fact that we behave differently from these other things that makes us conscious? Well I don't think that's quite right, I think that just behaving in the right way is not enough. I think that the internal processes inside the organism make a difference. So they have to be the right kind of internal processes. And now the question is: What is it about those internal processes that makes us subjects of conscious experience? That makes us have experiences of the world, visual experiences, auditory experiences and the rest? So what I want to do is give a framework in terms of which we can understand this, and I want to do this as simply as possible.

I just want to make it possible for us to understand how it can be that an object can be a subject. That something, a structure of neurones and other bits of matter or perhaps a structure made out of bits of silicon and wires and so on, can at the same time be something that it is like something to be.

There are two stages in my approach to this. The first stage is to set out what I call the basic package which is a set of capacities that an organism has, either all of, or none of. Having explained that, which I think is reasonably straitforward, I then want to go to the second stage, and the second stage is explaining what else is needed beyond having the basic package in order to be a subject of conscious perceptual experience, in order to be something that is like something to be.

Now, having the basic package is a matter of being able to collect information. One thing about perceiving the world is clear, that when we see or hear the world, smell, whatever it is, we collect information. Now the information has to be for us. Think of a camera. A camera collects information, but the camera can't use information. It can't do anything with it. So the camera although it collects information does not collect information which is for it. But a dog does collect information which is for it, a dog might hear its food being put out and at the same time see a cat. What's it going to do? The dog has to make up its mind what it's going to do, and it has information about the cat and about his food being put out and it has to decide, so there's a big difference between the dog and the camera which just collects information that is for us.

So we want an explanation of what it is for something to collect information which is for it. Now my thought is that we have a number of related concepts that we know how to use reasonably well and it turns out that if any of these concepts applies to something, all of them apply, so its a package, an unbreakable package of concepts. For instance, acquiring information which is for the system, is impossible unless the system can use it. Using the information requires the system to be able to apply the information in its activity, so its got to be able to initiate and control its activity on the basis of the information it collects. It's also got to be able to asses the information. It's got to be able to assess the situation it's in because unless it can assess the situation it could not act in one way rather than another. So in a sense it has to decide. Now cats and dogs are not brilliant thinkers but there's something like assessment of the situation going on with a cat or a dog. And there's something like decision making going on too. So these concepts, collecting information, assessing the situation, interpreting the information, deciding what to do, all of these concepts form a package which something either has all of or not. That's what I call the basic package.

So that's the first stage of my account.

Now I think that it's reasonably straightforward to decide, when you're faced with something, whether or not it has the basic package. I think it's helped by the fact that it's all or nothing. You have to check through the various components of the basic package. Does this thing you're faced with have all these abilities together? Is there something going on in it which counts as its assessing its situation? Is there something which counts as its deciding what to do? If there isn't it doesn't have the basic package, so its not a candidate. A thermostat doesn't assess its situation, it doesn't decide what to do, its automatic, so it doesn't have the basic package, for example.

So that's the first stage.

Now the second stage is this. I think it's pretty clear that something can have the basic package but not be a subject of conscious perceptual experience. Because I think it's clear that something can acquire the information, and the information can be in a way available for it, because it's stored in its memory, but before it can use the information it has to call it up from memory. It has to do something. This is the situation with the case of blind sight, and its also the situation with subliminal perception. The information does not immediately affect the organism's processes of assessment and decisionmaking. So I think this is the key thing that we have to focus on. That just having the basic package is not sufficient. In addition, incoming information, some of the incoming, perceptual information has got to affect the organism's assessment processes in a direct way and it has to affect these processes in such a way that the organism is instantly able to do things that it otherwise would not be able to do.

So when I'm looking around this room I see, for example, Stephen's red shirt and I see the grey green chairs, I hear certain sounds and so on, and all of this is information which is not just coming in and being stored somehow, but it is information which is coming in and is directly affecting my abilities. It directly affects what I can do. It affects my assessment of my situation by affecting what I can do if I choose. It doesn't force me to do anything, but it enables me to choose to do these things. For example I can point to the collar of that shirt, I can point to the legs of that chair. I can do all these kinds of things as a result of the information directly affecting my assessment processes. So that in a nutshell is the idea of the basic package. First of all we use a set of concepts that we have a reasonably good grip of. We notice that they form a package, something either has all of them or none. We then notice that those are not sufficient for something to be a conscious subject. We consider what extra is required and that is that the information directly affect the organism's processes of assessment, and then I say (this is something that needs further argument, but this is my idea): then no more is needed. That is all is needed in order for a thing to be a subject of conscious perception. So there's something it's like for something which has the basic package and is also collecting information which is directly active in the way that I've explained.

So my thought is that this solves, at least this is the basis for making real progress, with the solution of the very difficult problem of how it is, how it can be that an object can at the same time be a subject.

To return to the philosophy discussion

 **return to Philosophy index**	 **return to overall index**	 **return to Introduction and textual index**

Paul Davies

talks to Stephen Jones at Tucson II.

SJ: You would have a fairly impartial view of what's going on at this conference in terms of the conjunction, if that's the correct word, if there is a conjunction in fact, between the question of whence consciousness arises, and the proposals about quantum physics that are being made, and perhaps the necessity for a new physics or a new layer in physics.

Paul Davies: Yes, I suppose uncommitted is the best way of saying that. For a long time I've wondered whether we need to import something new, not in terms of something necessarily mystical, but some new physics or new aspect of physics in order to understand consciousness clearly. The one thing I think we're all agreed on is that consciousness requires complexity. That's the thing about the brain which is most striking, its incredible complexity. And I really have no sympathy for the point of view that's being propounded here which is that consciousness infuses everything. I refer to the so called panpsychic models or panpsychic ideas, where atoms are just a tiny bit conscious and you put lots of atoms together and they're a lot conscious. It seems to me that that just won't work. I believe that only in systems sufficiently complex will consciousness emerge and flourish. Now whether complexity alone is enough is what I am not sure about.

In regard to the quantum aspects of this, which are exciting and intriguing, let me make a few points. The first is, that of course, a link between consciousness and quantum physics has been there right from the start, in the sense that quantum mechanics tells us about our knowledge of the world, as opposed to how the world is. That was made very clear by Niels Bohr, one of the founders of the subject. I think we all agree that there is a problem about the so-called measurement or observation of physical quantities at the quantum level, so if we're making a measurement of an electron or an atom or something, we want to try to understand that in terms of the wave function and the so-called collapse of the wave function. This seems to have something to do with the observer, something to do with consciousness.

Now recently some physicists have tried to get rid of the conscious observer in this measurement process. In my opinion their attempts to do that, whilst they may seem superficially successful, are a move in the wrong direction. It seems to me that quantum physics gives us the one chance that we know of within physics to take into account the observer. To try to write the observer out of the picture seems to me to be missing an opportunity. The problem about theories of physics which abolish the observer is that they make no distinction between physical processes in which an observation is taking place, and any other type of physical processes. If quantum mechanics is really a theory of the whole world as it purports to be, then it had better tell us what it is that distinguishes one system, namely an observer, from other systems.

So my feeling then, is that we ought to be able to understand, within the framework of quantum mechanics, how an observer fits in. But that is not the same thing as saying that quantum mechanics is going to play a crucial role in the brain. There I'm more sceptical. I've heard what Hameroff and Penrose (1) have to say about the famous microtubules. It is clearly the case that, at some level, quantum mechanics is going to make a difference in the brain. But I still consider the major obstacle to their work to be the fact that the brain is at the temperature of 310K (37deg C) or about. We know that a quantum system will interact with its environment and if this environment is at a reasonably high temperature, it's a noisy environment. It will then have the effect of, to use the jargon, *decohering* the wave function. This is to say that the crucial phase relationships in the quantum wave, are rapidly destroyed by this noisy environment. Unless some sort of mechanism can be found to protect quantum coherence from that environmental noise, then I think we must remain sceptical. Having said that of course, we do know of a case, namely high temperature superconductivity, where this decohering thermal noise is surprisingly suppressed at the relatively elevated temperature of, I don't know what the record is, but something up to about 100 degrees Kelvin. But the thought of protecting decoherence right the way up to blood temperature, it's very hard to see how that would happen. So that's one main obstacle.

The other is that I think it's yet to be demonstrated that the microtubules, or any quantum effects that might occur within them, can propagate outside of an individual cell. Whilst I have some sympathy with the notion that there may be some information processing going on at the sub-cellular level which could perhaps explain why single-celled organisms like bacteria can perform such clever tricks, what I have yet to hear is any attempt to argue that these quantum coherence effects via the microtubules could spread out in the brain across distances of the order of say centimetres. Unless we can find an effect like that then I don't think this type of theory would be totally credible.

I must admit I agree with Daniel Dennett that we've got to abolish all notions of a so-called Cartesian theatre, of a little man in side my head who watches the movie and is in a sense the seat of the soul or the seat of the self, 'cause it gets us nowhere. Why? Because we then have to take the little man apart and try and understand what's inside him. I don't think we can localise consciousness in the brain down to the level of systems so small that we would normally think quantum mechanics would be important. In other words, whatever it is that's giving rise to consciousness in the brain is clearly distributed, and I would say distributed over distances of the order of centimetres. So if quantum mechanics is going to be important we have to find some way of understanding how quantum effects can propagate outside of individual neurons and across millions or even billions of neurons. That's a pretty tall order. Nothing I've heard at this conference so far has addressed that issue.

SJ: You did bring up a point that I feel is quite important to explore, one that hasn't been explored readily yet, which is complexity. And complexity to me implies organisation, and then organisation implies certain kinds of principles of organisation, using say the cybernetic terminology, and I was wondering whether you have anything to say about that kind of stuff?

PD: Yes well I do. So to pick up once again on the subject of complexity. It's very important to distinguish between two quite different types of complexity: disorganised and organised complexity. Consider a gas in thermodynamic equilibrium, with all the molecules rushing around chaotically, well that's pretty complex. If you wanted to write down the microscopic state of that system there'd be an awful lot of numbers to record. But we can distinguish that type of chaotic complexity from, say, the complexity of a bacterium, which is highly organised complexity.

Attempts to capture the notion of organisation in complexity go back quite a few years now. I'm particularly impressed by the work of people like Charles Bennett and Gregory Chaitin (2) at IBM, who've introduced the notion of *depth*. It's an attempt to quantify organised complexity by looking at the amount of information processing that has gone on, or the time required to process that information. Consider something like a bacterium, where we see the record of its information processing in its DNA, a precious molecule that has been fashioned over billions of years in a very large number of evolutionary steps. There's a lot of information processing required to achieve a bacterium.

Now, I think complexity, and specifically *organised complexity*, is the key to consciousness. I also think it's the key to understanding the measurement process in quantum mechanics. The reason is this. Until now we've tended to treat complexity in a qualitative manner. That is to say we recognise when something is complex, but we tend to think it's just complication. In other words, the reductionist spirit of physics has tended to make us look at the individual components and if we see that the whole system is complex, well, we may remark on that but we don't think that it is physically terribly relevant. But seeing as we now have ways of quantifying complexity I wonder whether these quantities might enter into the laws of physics in a manner that might be called fundamental, rather than just incidental. In other words, these principles of organisation or principles of complexity, which we would place alongside the laws of thermodynamics in describing bulk matter, might be every bit as fundamental as the basic laws of quantum mechanics. And I wonder if there is a link between the two.

I've often toyed with the idea that maybe the so called collapse of the wave function, or the concretisation of quantum potentialities, that occurs when we make a definite observation of the world, occurs when the quantum system - the total system that we're considering - is sufficiently complex. We know that a human observer can do this, so presumably the human brain is sufficiently complex. I'm sure the brain of a cat (to take the famous Schroedinger cat example) would be sufficiently complex to collapse the wave function. It's possible that even a computer would be sufficiently complex.

A problem that's long interested me is the origin of the universe and how the universe might come into existence as a result of a quantum process. We face the major problem of how it was that a quasi-classical world emerged from the quantum origins, the quantum foam, or whatever metaphor one wants to use about the very early universe. I think one can still play this game because in its early stages the universe was actually very simple. If we allow for just a few degrees of freedom to be quantised we can describe the coming into being of the universe in this quantum mechanical manner, but as the universe evolved it became more and more complex. I conjecture that it was this emerging complexification that brought about the quasi-classicality of the universe. This is an idea that, as far as I know, has not been developed. The only other person I know who's commented on it is Tony Leggett (3), who has clearly had similar ideas. I would like to see an attack on this problem.

To conclude, I think that the emergence of consciousness is associated with sufficient organised complexity, and that what is sometimes misleadingly called the collapse of the wave function or the concretisation of potentialities in quantum mechanics is also associated with sufficient complexity. I believe this is a more fruitful avenue for finding the bridge between quantum mechanics and consciousness: it's through complexity and not at the simple microtubule level which Penrose and Hameroff are pursuing. But that's just pure conjecture.

References

1: Penrose, Roger. Shadows of the Mind. Oxford University Press, 1995

2: Chaitin, Gregory: Algorithmic Information Theory. Cambridge University Press.

3: Leggett. A.J. The Problems of Physics. Oxford University Press

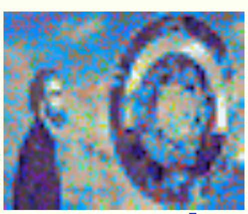
Links:

Paul Davies home page

Review by Ilya Prigogine of About Time: Einstein's Unfinished Revolution by Paul Davies

---

To return to the discussion of organisation and complexity.

 **return to the quantum discussion**	 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

On Animation: The Illusion of Life

by Stephen Jones.

This is the body of a paper presented to **The Illusion of Life** a Conference on Animation; held in Sydney, July, 1993. It presents, perhaps in too poetic a fashion, the beginnings of an hypothesis for the emergence of consciousness in us.

Where does the term Animation come from? Why does animation have meanings both in film terms and in the idea of something being alive? Its origins are from the Greek "anima", and it was used to refer to the "animal soul", and later, in the 16th century, "animal spirit". I shall explain...

In the effort to understand living things the Greeks, particularly Aristotle, postulated three souls: the vegetal soul, the basis of plant life, which in animals and humans was to be found in the liver; the animal soul, seated in the heart and, coupled with the vegetal soul, the basis of life; and the rational soul, housed in the brain and that which distinguished humans from animals.

Several spirits provided nourishment for and produced movement in these souls. It was the function of the lungs to gather pneuma from the air, whence it was delivered to the brain and mixed with the vital spirit from the heart to form 'animal spirit' for redistribution through hollow nerves to 'animate' the body. Galen wrote this theory up and his writings held sway until in the 17th century much of his ideas provided the framework within which new anatomical knowledge was interpreted. For example, Descartes wrote:

> "The cavities of the brain are central reservoirs ... animal spirits enter these cavities. They pass into the pores of its substance and

> from these pores into the nerves. The nerves may be compared to the tubes of a waterworks; breathing or other actions depend on the flow of animal spirits into the nerves. The rational soul takes place of the engineer, living in that part of of the reservoir that connects all of the various tubes. These spirits are like the wind. When they flow into a muscle they cause it to become stiff and harden, just as the air in a balloon makes it hard."

Descartes, in his radical doubting thought experiment reduced life to a mechanical model. He attempted to reach a point of 'certain', i.e. indubitable, knowledge and in the *Discourse on Method* he threw out all knowledge derived from the senses. Finally the only thing he could not doubt was that he was a 'thinking thing', a '*res cogitans*' currently housed in a mechanical system or body animated by hydraulics. The hydraulic fluid was the 'animal spirit' which provided the motivational force for all mobile living things.

Descartes used an analogy based on the technologically most advanced concepts of the day. Within the world of public entertainment at the time the latest rage were the "automata", animated mechanical models which emulated the activity of people and animals by hydraulic means. And it was this mechanical "animation" which gave the appearance of life.

Descartes split mind and consciousness from the body because he could not rely on the information of the senses and his devotion to the Roman church would not allow him to reduce his soul (or mind) to a material thing, a '*res exsitans*'. The soul was connected into the body through the pineal gland, a singular organ at the base of the brain. Singular in the sense that unlike the rest of the organs of the brain it is not twinned.

As regards sensations, in this case those of the eye, Descartes suggested that all the nerves of the eye went to the pineal body at the base of the brain to be relayed to the soul. But he denies that what we see are tiny pictures formed within our head.

> "(For) we have to consider that thought may be induced by many things besides pictures - e.g. by signs and words, which in no way represent the things signified." [Dioptrics, Discourse IV].

Though, as Descartes showed by experiment with the eye of a newly dead ox, the eye focuses a tiny inverted image of the scene onto the back of the eye, onto the retina, it is from here that encoding and processing begins. I shall give a rough description of the visual processing system as elucidated by neurophysiological work of the last several decades.

The neural cells of the retina consist in rods and cones overlayed by a middle layer of interneurons and an outer layer of ganglion cells whose axons make up the fibres of the optic nerve. The rods detect changes in brightness and the cones detect different wavelengths of light, i.e. colour. The concentration of cones to rods is much greater at the fovea (which is the centre of focus) and decrerases towards the edges of the retinal field. The rods are able to detect down to the single photon (or packet of light energy).

Once an array of retinal cells has been stimulated by photons, those neural cells so stimulated propagate an 'action potential', as the signal is called, to the layers of interneurons which then feed to the retinal ganglion cells. Some of these interneurons do motion detection using the propagation delays through the nerve-cell axons. The ganglion cells then gather up all the light pattern information processed in the retina and send it through the optic nerve to the visual cortex. In the optic nerve right and left sides of the retinal data are mapped together such that the right side of both retinas is sent, via the lateral geniculate nuclei deep in the centre of the brain to the left visual cortex and the left side of both retinas is sent (also via the lateral geniculate nuclei) to the right visual cortex. In the lateral geniculate nuclei the processing pathways separate into a colour sensitive pathway and a brightness contrast sensitive pathway.

In the visual cortex three kinds of processing seem to take place, each in several stages. Certain cells are selective for colour or brightness. Other cells are selective for orientation. And in other cells various aspects of feature detection take place. Starting with low level edge detection performed by cells sensitive to bright-to-dark transitions of particular orientations, these are then integrated into more complex features through connections into further layers of cells. So we have here a hierarchy of detected features.

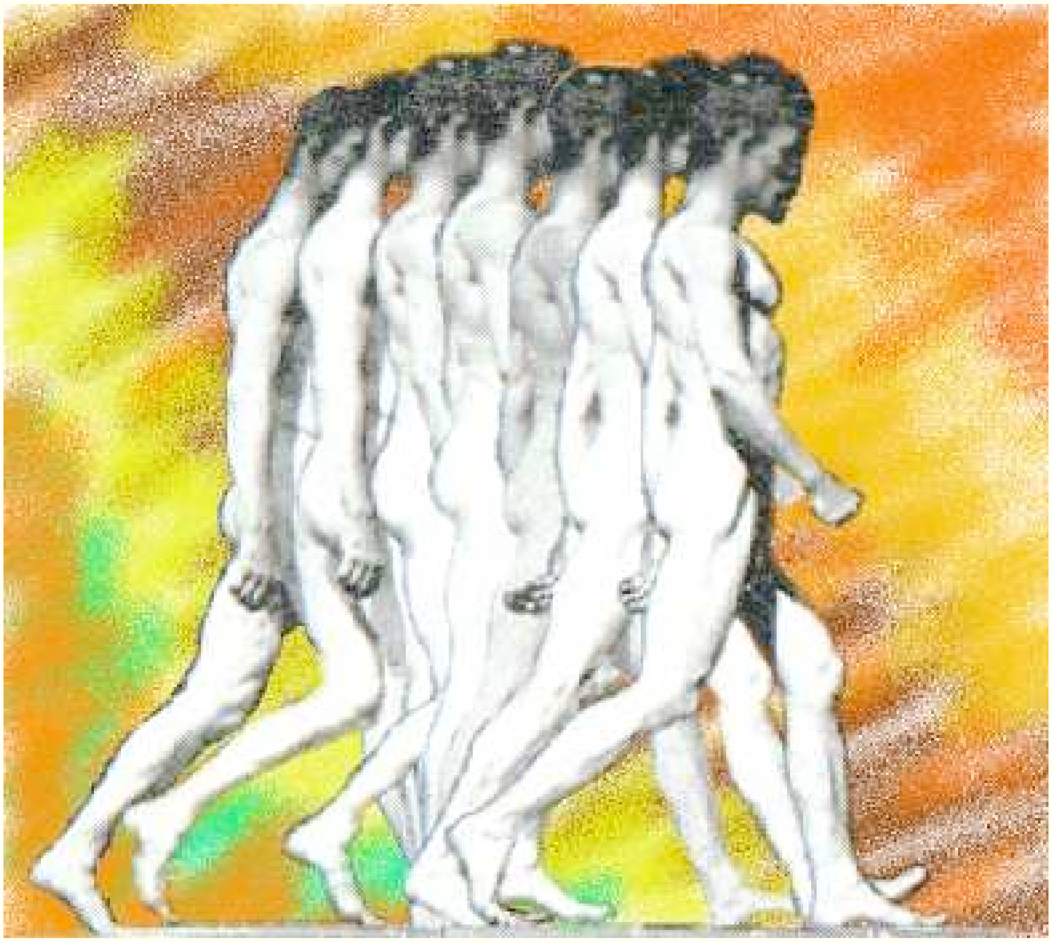
Projections from the visual cortex also go to medial temporal regions which appear to be involved in detection of movement and in stereopsis (which is one aspect of depth perception) possibly correlating or integrating that information with other information about movement derived from other sensory pathways.

The brain cannot be carrying images as representation but can only be carrying coded data about images for analysis by the various detection and recognition processors in it. This is an ongoing processing, never finished, always becoming, acquiring more examples, the current frame overlaying the previous as they process through the brain via different pathways experiencing different propagation delays providing the means for comparison, the current frame with the previous. In this we find various tie-ins with memory processing.

Of course the brain doesn't deal with coded image data in frames of film or video as we talk about it, unless it is being presented with film or video. Nevertheless, at the finest level, the quantum mechanical, all things which might cause sensation come in discrete packets, or quanta, so that we might say that at the finest level the universe is digital. It is the processes of sensation that give the illusion of an analogue world. It takes time for things to happen in the brain, so the differences in propagation delay through different processing pathways smooth out the gaps between quantum events as sensed.

All the pathways in the brain are massively interconnected, not just hierarchically as levels of integration, but also horizontally. It has been suggested that every neuron in the brain, all 10 billion or so, may only be half a dozen neurons away from every other neuron, in interconnection terms. The combination of highly organisd, highly complex processing systems and subsystems, with this massive interconnectedness is probably enough to in-and-of-itself generate consciousness. And even more especially so when coupled to the social/linguistic processes of the multiplicity of generative entities providing the apparent environment which each individual encounters. Within an individual system of sensations and their reverberations, consciousness is generated through the whorls of feedforward and the regulation of feedback, and within a social system, language and its reverberations as culture and so on all feedback into the apparent individual elements of the system intertwining them and interconnecting them, creating continuously a single unitary social fabric which directs, and is a consequence of, the becoming of consciousness. If we can only sense, can we know anything else? Language is the means solely of propagating sensations through groups allowing and governing the interprtation of sensation and shaping the consciousness. There can be nothing which cannot be talked or written about. There can only be that which is sensed and which might be talked or written about.

Our knowledge of the world is entirely a knowledge of sensation, of the encoded signs of sensory data established at the sensor be it eye or ear or skin. The processing systems of the sensors and the brain, the feature detectors and the integrators, the comparators and the models for comparison all derive their capacities from the exercise of the sensory system in the first place and its moulding into sociality during the whole process of development and maturation.



What I think is going on is something like this. All these highly organised, highly interconnected pathways of neural tissue form into networks which are then trained by the data flowing through them so that they become pathways of interpretation. All this data is patterned and as more and more patterned data flows through the networks they develop form around the information. The different pathways and splittings-off of this patterned data carry out particular processing transforms on the data, all the time preserving the relational consistency of the pattern. Data flows through various processing elements. In the visual system data is split off into colour information pathways, brightness and contrast information, edge orientation and depth detection pathways, and motion detection pathways using both positional and temporal differentiation. The resulting patterns having been analysed for these data are then sent to higher interpretive processing and image recognition. Somewhere along the way they are integrated with auditory and other sense data and the unitary conscious knowing of the world and ourselves ensues.

All the organisation, interconnectedness, time-delayed reverberation and, probably, re-integration provide this continuous but always regenerating and generative system of perception and ideas which by its own process either is or generates consciousness. We are talking about highly complexly organised systems in continuous flux which show a centripetally organised structure feeding into itself almost as an oscillatory system which by its own activity may be said to go 'live'. Very many complexly interlinked feedback loops which put the system in oscillation (or on the very edge of chaos) which is then the liveness, the presence in the continual present, of consciousness.

Indivduality is a function of our language and the framework of our culture and its in-forming of ourselves. It is an illusion because the sensation and in-form-ation forms part of the processor (software builds hardware, so to speak). Though it is regarded within our language system as being outside us it is really part of us. We are open systems without real boundaries between us and the world. A unitary entity of information and flesh, ecology and relations between things and the sensing in-formed thing (almost a processing element in the mind of Gaia, though gone mad with the psychosis of vastly conflicting desires) folding back onto itself generating controlled far-from-equilibrium fluid structures always in becoming.

The system is generative partly because it is an open system and partly because it is so complexly interconnected with all sorts of levels of feedback relations, that we can't help but always see things in the light of things known and seen previously, and vice-versa, see what is current in the light of what is new. It is a natural function of the brain to be generative. language is generatiuve, creativity is a generative process. Animation is produced from generative processes both of concepts and visual manifestations. Culture is a construction made out of individuals, history and ideas.

---

return to Top Index

David Chalmers

talks to Stephen Jones at Tucson II.

SJ: Perhaps you could outline how you come to the idea of the "Hard Problem" and what it means to you?

David Chalmers: I got into this field to try and understand the problem of how a physical system like a brain could also be a conscious being with subjective experience. A lot of people recently have started thinking about consciousness in many different ways - from philosophy, neuroscience, psychology and elsewhere. But when you take a closer look at what people are doing in these fields you find out that they are addressing a number of different questions in the vicinity, not all of which are the deep mystery.

So quite often when you go along to a conference like this Tucson conference, you find out "well, there's a bunch of results on how the brain integrates a bunch of information, for example, and brings it to bear in the control of behaviour", you think "OK well, interesting results" and then you think: "Well, isn't the mystery still there? Why is it that all this processing should give rise to a subjective view of the world? a subjective inner life?" So, in trying to get at this I made a distinction between the easy problems of consciousness and the hard problem.

The easy problems are the kind of problems that neuroscience and psychology can get at sort of straightforwardly. How is it that a brain can discriminate information from the world? How is it that it can bring it together in the brain and integrate it? How is it that the brain or a human being can verbally report their mental states? How is it that we bring information to bear in controlling our action? Now, of course, the easy problems aren't trivial problems. They're going to take a long time to solve, and they'll require a lot of intelligence, creativity and hard work. But we're gradually getting at these questions using the methods of neuroscience and psychology. It's slow work, but there is a clear sense that we have a research program there. We know roughly which direction to move in to get a result.

The trouble is, it seems that even as we work on all these questions, the deep mystery still remains. Despite all the complex brain functioning that we're finding out about, it still remains mysterious. Why is it that all this functioning should give rise to a subjective inner life, an experienced subjective world. I have such a world, I have subjective experience, I presume that you do and that most people do as well. Well, I don't know about others for sure but I'm certain about me. The hard problem is: *why is it that physical processing in the brain, no matter how sophisticated, should give rise to any subjective inner life at all, why couldn't that have all gone on in the dark?* That's the real mystery.

SJ: So why couldn't this be in some way just the normal functioning of the brain?

DC: There's no question there's an close link between what's going on in the brain and what's happening in conscious experience. They're at least very tightly correlated. I wouldn't resist the claim that the brain somehow gives rise to or produces conscious experience. The question is how does that happen? And does an explanation of what's going on in the brain, itself explain why consciousness arises?

Now for most natural phenomena that need explaining, we get a physical explanation. To explain the gene, say, we get an explanation involving DNA. We discover how DNA and a bunch of complex molecules store information and transmit it from one generation to the next and then we say we've explained the gene - that's what Watson and Crick started to do and that research program has been pushed forward. In that sense we don't say we just have a correlation - "wow, where you have DNA you have a gene" - there you have an explanation. You've explained everything because all you have to do in explaining the gene is explain how hereditary characteristics get passed from one generation to the next. That's a problem about how functions get performed, and we solve that sort of problem by specifying mechanisms.

Same for all the easy problems I talked about. How is it that the brain discriminates information? How is it that we make reports? Those are questions about how the brain performs certain functions, how it does something. To solve that kind of problem what you need to do is specify a mechanism that performs the function, and then I think you've answered those questions. But what makes the hard problem different, almost unique, perhaps, is that it doesn't seem to be that kind of question. You can specify the mechanism, you can tell how all the functions in the vicinity are performed and there's still this further question: *Why is it that that mechanism and the performance of those functions gives rise to experience?* So, while one may have a correlation there, there doesn't seem to be explanation, precisely because what needs to be explained here isn't something about what the brain does but something, in a sense, about what the mind is.

So, given that the brain does give rise to subjective experience (and I believe it does, I think there's a very tight link there), it seems that there's something fundamental and primitive about that link. The way I put it sometimes is this: think about God creating the world, metaphorically. It seems that it is within God's powers, logically, when creating the world, to create creatures which are physically just like me, even have brains just like me, but have no subjective experience. That doesn't seem to me a conceptual contradiction. They are the famous philosopher's zombies.

Now, we aren't such zombies, we are in fact conscious beings. But the very fact that we are conscious beings indicates that there's something in us, over and above the basic physical components. Maybe there are some sort of basic mental components in there as well.

SJ: How can we move from correlation to explanation?

DC: We may have to recognise that there are just correlations between what's going on in the brain and what's going on in consciousness. But if we do that, we don't just want to stick with correlations like "okay, well when you get a brain like X you get a conscious state like Y". That would be a big tangle of messy correlations. I mean, in physics you may start with a bunch of correlations describing phenomena in mechanics or thermodynamics, and you can say well, when this gas goes into this state then it produces such and such behaviour, i.e. pressure and temperature work like this. But then you go for the underlying laws, the underlying principles that explain all this, and you go for the ultimate, the simple, fundamental principles of physics which underlie and explain all these high level correlations.

So if we have to do something like this with consciousness, bring something else fundamental into our theory, then what we want is a theory of how it is that that fundamental component and the underlying physical processes are related. We don't want it to be a mess of correlations, we want it to be this simple, fundamental theory, a basic set of fundamental laws. Physicists sometimes say they want a set of laws that are so simple that they're part of the fundamental furniture of the universe, they're so simple you can write them on the front of a T-shirt, and, in a sense, that's the goal of a fundamental theory of consciousness too. We want a bunch of what I call psychophysical principles because they're principles that connect physical processing with the psychological or the mental. These psychophysical principles should ultimately be as fundamental, as simple as the kind of principles we find in fundamental physics. Because I think it may well be that those principles themselves are part of the fundamental furniture of the world.

Then the question is "What are those principles?" That's of course the question for which nobody yet has an answer - that's the research program, in a sense, on the hard problem as I conceive it - to try and find the simple underlying principles connecting physical processing to consciousness, such that when you apply those laws in familiar cases like ours, to my brain for example, you predict that you're going to get the kind of conscious experience that I have, and if I apply it to a system like you it will give the kind of conscious experience that you have, and so on.

For this psychophysical framework, I have a couple of very speculative ideas about this involving aspects of information. The idea is that information may have a physical aspect and a phenomenal aspect. Information is embodied physically in the brain in such and such a way, but at the same time, it has an experiential, or phenomenal side. Information has this double-sided nature and our conscious lives are in a sense the mental side of this information which is also embodied physically. And maybe we can take that very speculative principle and develop it into a systematic simple framework which will then be a fundamental psychophysical theory. On the other hand that may be completely wrong. We may need to go in another direction completely. I think there is nevertheless a research program there and one of the things I'm going to be interested to see over the next few decades, century, whatever, is whether this hope for a fundamental theory might eventually pan out.

SJ: Do you think this dual aspect to information is going to require a new physics?

DC: One doesn't necessarily need to bring in new physics to explain the physical side of information. It's something that's there, it's the product of the mass, the charges and the forces that are out there already. And so we take that information that is already implicit within physical theory and then we add this extra component - the phenomenal component - and we say: "Hey, If I bring in the hypothesis that information has this two sided nature..." that may then give us what we need to then bring consciousness into our picture of the natural world.

References and Links for David Chalmers

Dave Chalmers' Home Page

The Conscious Mind by David Chalmers. His new book. Well worth the read.

Facing Up to the Problem of Consciousness The central paper on the Hard Problem

Absent Qualia, Fading Qualia, Dancing Qualia In which Chalmers "argues that functional organization fully determines conscious experience".

Availability: The Cognitive Basis of Experience? in which Chalmers "argues that the cognitive correlate of conscious experience is direct availability for global control".

David Chalmers' Annotated Bibliography Possibly the best bibliography in the business.

<table>
<tr><td colspan="3">To return to the philosophy discussion</td></tr>
<tr>
<td>
return to Philosophy index</td>
<td>
return to overall index</td>
<td>
return to Introduction and textual index</td>
</tr>
</table>

The Hornswoggle Problem.

by Patricia Smith Churchland.
Philosophy, University of California San Diego, Salk Institute
8/12/96

Abstract:

Beginning with Thomas Nagel, various philosophers have proposed setting conscious experience apart from all other problems of the mind as "the most difficult problem." When critically examined, the basis for this proposal reveals itself to be unconvincing and counter-productive. Use of our current ignorance as a premise to determine what we can never discover is one common logical flaw. Use of "I-cannot-imagine" arguments is a related flaw. When not much is known about a domain of phenomena, our inability to imagine a mechanism is a rather uninteresting psychological fact about us, not an interesting metaphysical fact about the world. Rather than worrying too much about the meta-problem of whether or not consciousness is uniquely hard, I propose we get on with the task of seeing how far we get when we address neurobiologically the problems of mental phenomena.

I: Introduction

Conceptualizing a problem so we can ask the right questions and design revealing experiments is crucial to discovering a satisfactory solution to the problem. Asking where animal spirits are concocted, for example, turns out not to be the right question to ask about the heart. When Harvey asked instead, "how much blood does the heart pump in an hour?", he conceptualized the problem of heart function very differently. The reconceptualization was pivotal in coming to understand that the heart is really a pump for circulating blood; there are no animal spirits to concoct. My strategy here, therefore, is to take the label, "The Hard Problem" in a constructive spirit -- as an attempt to provide a useful conceptualization concerning the very nature of consciousness that could help steer us in the direction of a solution. My remarks will focus mainly on whether in fact anything positive is to be gained from the "Hard Problem" characterization, or whether that conceptualization is counterproductive.

I cannot hope to do full justice to the task in short compass, especially as the contemporary characterization of the problem of consciousness as *the* intractable problem has a rather large literature surrounding it. The watershed articulation of consciousness as "the most difficult problem" is Thomas Nagel's classic paper "What is it like to be a bat?" (1974) In his opening remarks, Nagel comes straight to the point: "Consciousness is what makes the mind-body problem really intractable." Delineating a contrast between the problem of consciousness and all other mind-body problems, Nagel asserts: "While an account of the physical basis of mind must explain many things, this [conscious experience] appears to be the most difficult." Following Nagel's lead, many other philosophers, including Frank Jackson, Saul Kripke, Colin McGinn, John Searle, and most recently, David Chalmers, have extended and developed Nagel's basic idea that consciousness is not tractable neuroscientifically.

Although I agree that consciousness is, certainly, a difficult problem, difficulty *per se* does not distinguish it from oodles of other neuroscientific problems. Such as how the brains of homeotherms keep a constant internal temperature despite varying external conditions. Such as the brain basis for schizophrenia and autism. Such as why we dream and sleep. Supposedly, something sets consciousness apart from all other macro-function brain riddles such that it stands alone as **The Hard Problem**. As I have tried to probe precisely what that is, I find my reservations multiplying.

II: Carving Up the Problem Space

The-Hard-Problem label invites us to adopt a principled empirical division between consciousness (The Hard Problem) and problems on the "Easy" (or perhaps hard but not Hard?) side of the ledger. The latter presumably encompass problems such as the nature of short-term memory, long-term memory, autobiographical memory, the nature of representation, the nature of sensory-motor integration, top-down effects in perception -- not to mention such capacities as attention, depth perception, intelligent eye movement, skill acquisition, planning, decision-making, and so forth. On the other side of the ledger, all on its own, stands consciousness -- a uniquely Hard Problem.

My lead-off reservation arises from this question: what is the rationale for drawing the division exactly there? Dividing off consciousness from all of the so-called "easy problems" listed above implies that we could understand all those phenomena and still not know what it was for ..... what? The "qualia-light" to go on?? Is that an insightful conceptualization? What exactly is the evidence that we could explain all the "Easy" phenomena and still not understand the neural mechanisms for consciousness? (Call this the "left-out" hypothesis.) That someone can imagine the possibility is not evidence for the real possibility. It is only evidence that somebody or other believes it to be a possibility. That, on its own, is not especially interesting. Imaginary evidence, needless to say, is not as interesting as real evidence, and what needs to be produced is some real evidence.

The left-out hypothesis -- that consciousness would still be a mystery, even if we could explain all the Easy problems -- is dubious on another count: it begs the question against those theories that are exploring the possibility that functions such as attention and short-term memory are crucial elements in the consciousness. (See especially Crick 1994, P. M. Churchland 1995) The rationale sustaining this approach stems from observations such as: that awake persons can be unaware of stimuli to which they are not paying attention, but can become aware of those stimuli when attention shifts. There is a vast psychological literature, and a nontrivial neuroscientific literature, on this topic. Some of it powerfully suggests that attention and awareness are pretty closely connected. The approach might of course be wrong, for it is an empirical conjecture. But if it is wrong, it is wrong because of the facts, not because of an arm-chair definition. The trouble with the Hard-Problem characterization is that on the strength of a proprietary definition, it rejects them as wrong. I do find that unappealing, since the nature of consciousness is an empirical problem, not a problem that can be untangled by semantic gerrymandering.

What drives the left-out hypothesis? Essentially, a thought-experiment, which roughly goes as follows: we can conceive of a person, like us in all the aforementioned Easy-to-explain capacities (attention, short term memory etc.), but lacking qualia. This person would be exactly like us, save that he would be a Zombie -- an anaqualiac, one might say. Since the scenario is conceivable, it is possible, and since it is possible, then whatever consciousness is, it is explanatorily independent of those activities. (Something akin to this was argued by Saul Kripke in the 1970's.)

I take this argument to be a demonstration of the feebleness of thought-experiments. Saying something is possible does not thereby guarantee it is a possibility, so how do we know the anaqualiac idea is really possible? To insist that it must be is simply to beg the question at issue. As Francis Crick has observed, it might be like saying that one can imagine a possible world where gasses do not get hot, even though their constituent molecules are moving at high velocity. As an argument against the empirical identification of temperature with mean molecular KE, the thermodynamic thought-experiment is feebleness itself.

Is the problem on the "Hard" side of the ledger sufficiently well-defined to sustain the division as a fundamental empirical principle? Although it is easy enough to agree about the presence of qualia in certain prototypical cases, such as the pain felt after a brick has fallen on a bare foot, or the blueness of the sky on a sunny summer afternoon, things are less clear-cut once we move beyond the favored prototypes. Some of our perceptual capacities are rather subtle, as, for example, positional sense is often claimed to be. Some philosophers, e.g. Elizabeth Anscombe, have actually opined that we can know the position of our limbs without any "limb-position" qualia. As for me, I am inclined to say I do have qualitative experiences of where my limbs are -- it feels different to have my fingers clenched than unclenched, even when they are not visible. The disagreement itself, however, betokens the lack of consensus once cases are at some remove from the central prototypes.

Vestibular system qualia are yet another non prototypical case. Is there something "vestibular-y" it feels like to have my head moving? To know which way is up? Whatever the answer here, at least the answer is not glaringly obvious. Do eye movements have eye-movement qualia? Some maybe do, and some maybe do not. Are there "introspective qualia", or is introspection just paying attention to perceptual qualia and talking to yourself? Ditto, plus or minus a bit, for self-awareness. Thoughts are also a bit problematic in the qualia department. Some of my thoughts seem to me to be a bit like talking to myself and hence like auditory imagery but some just come out of my mouth as I am talking to someone or affect decisions without ever surfacing as a bit of inner dialogue. None of this is to deny the pizzazz of qualia in the prototypical cases. Rather, the point is just that prototypical cases give us only a starting point for further investigation, and nothing like a full characterization of the class to which they belong.

My suspicion with respect to The Hard Problem strategy is that it seems to take the class of conscious experiences to be much better defined than it is. The point is, if you are careful to restrict your focus to the prototypical cases, you can easily be hornswoggled into assuming the class is well-defined. As soon as you broaden your horizons, troublesome questions about fuzzy boundaries, about the connections between attention, short term memory and awareness, are present in full, what-do-we-do-with-that glory.

Are the Easy Problems known to be easier than The Hard Problem? Is the Hard/Easy division grounded in fact? To begin with, it is important to acknowledge that for none of the so-called "easy" problems, do we have an understanding of their solution. (See the partial list on p. 2) It is just false that we have anything approximating a comprehensive theory of sensori-motor control or attention or short-term memory or long-term memory. Consider one example. A signature is recognizably the same whether signed with the dominant or non-dominant hand, with the foot, with the mouth or with the pen strapped to the shoulder. How is "my signature" represented in the nervous system? How can completely different muscle sets be invoked to do the task, even when the skill was not acquired using those muscles? We do not understand the general nature of motor representation.

Notice that it is not merely that we are lacking details, albeit important details. The fact is, we are lacking important conceptual/theoretical ideas about how the nervous system performs fundamental functions -- such as time management, such as motor control, such as learning, such as information retrieval. We do not understand the role of back projections, or the degree to which processing is organized hierarchically. These are genuine puzzles, and it is unwise to 'molehill' them in order to 'mountain' up the problem of consciousness. Although quite a lot is known at the cellular level, the fact remains that how real neural networks work and how their output properties depend on cellular properties still abounds with nontrivial mysteries. Naturally I do not wish to minimize the progress that has been made in neuroscience, but it is prudent to have a cautious assessment of what we really do not yet understand.

Carving the explanatory space of mind-brain phenomena along the Hard and the Easy line, as Chalmers proposes, poses the danger of inventing an explanatory chasm where there really exists just a broad field of ignorance. It reminds me of the division, deep to medieval physicists, between sublunary physics (motion of things below the level of the moon) and superlunary physics (motion of things above the level of the moon). The conviction was that sublunary physics was tractable, and is essentially based on Aristotelian physics. Heavy things fall because they have gravity, and fall to their Natural Place, namely the earth, which is the center of the universe. Things like smoke have levity, and consequently they rise, up being their Natural Place. Everything in the sublunary realm has a Natural Place, and that is the key to explaining the behavior of sublunary objects. Superlunary events, by contrast, we can neither explain nor understand, but in any case, they have neither the gravity nor levity typical of sublunary things.

This old division was not without merit, and it did entail that events such as planetary motion and meteors were considered unexplainable in terrestrial terms, but probably were Divinely governed. Although I do not know that Chalmers' Easy/Hard distinction will prove ultimately as misdirected as the Sublunary/Superlunary distinction, neither do I know it is any more sound. What I do suspect, however, is that it is much too early in the science of nervous systems to command much credence.

One danger inherent in embracing the distinction as a principled empirical distinction is that it provokes the intuition that only a real humdinger of a solution will suit The Hard Problem. Thus the idea seems to go as follows: the answer, if it comes at all, is going to have to come from somewhere Really Deep -- like quantum mechanics, or -- Wow -- perhaps it requires a whole new physics. As the lone enigma, consciousness surely cannot be just a matter of a complex dynamical system doing its thing. Yes, there are emergent properties from nervous systems such as co-ordinated movement as when an owl catches a mouse, but consciousness must be an emergent property like unto no other. After all, it is The Hard Problem! Consequently, it will require a very deep, very radical solution. That much is evident sheerly from the hardness of The Hard Problem.

I confess I cannot actually see that. I do not know anything like enough to see how to solve either the problem of sensori-motor control or the problem of consciousness. I certainly cannot see enough to know what one problem will, and the other will not, require a Humdinger solution.

III: Using Ignorance as a Premise

In general, what substantive conclusions can be drawn when science has not advanced very far on a problem? Not much. One of the basic skills we teach our philosophy students is how to recognize and diagnose the range of nonformal fallacies that can undermine an ostensibly appealing argument: what it is to beg the question, what a non sequitur is, and so on. A prominent item in the fallacy roster is *argumentum ad ignorantiam* -- argument from ignorance. The canonical version of this fallacy uses ignorance as the key premise from which a substantive conclusion is drawn. The canonical version looks like this:

> We really do not understand much about a phenomenon P.
> (Science is largely ignorant about the nature of P.)
> Therefore: we do know that:
> (1) P can never be explained
> or
> (2) Nothing science could ever discover would deepen our understanding of P.
> or

(3) P can never be explained in terms of properties of kind S.

In its canonical version, the argument is obviously a fallacy: none of the tendered conclusions follow, not even a little bit. Surrounded with rhetorical flourish, much brow furrowing and hand-wringing, however, versions of this argument can hornswoggle the unwary. From the fact that we do not know something, nothing very interesting follows -- we just don't know. Nevertheless, the temptation to suspect that our ignorance is telling us something positive, something deep, something metaphysical or even radical, is ever-present. Perhaps we like to put our ignorance in a positive light, supposing that but for the *Profundity* of the phenomenon, we would have knowledge. But there are many reasons for not knowing, and the specialness of the phenomenon is, quite regularly, not the real reason. I am currently ignorant of what caused an unusual rapping noise in the woods last night. Can I conclude it must be something special, something unimaginable, something.... alien ... other-worldly? Evidently not. For all I can tell now, it might merely have been a raccoon gnawing on the compost bin. Lack of evidence for something is just that: lack of evidence. It is not positive evidence for something else, let alone something of a humdingerish sort. That conclusion is not very glamorous perhaps, but when ignorance is a premise, that is about all you can grind out of it.

Now if neuroscience had progressed as far on the problems of brain function as molecular biology has progressed on transmission of hereditary traits, then of course we would be in a different position. But it has not. The only thing you can conclude from the fact that attention is mysterious, or sensorimotor integration is mysterious, or that consciousness is mysterious, is that we do not understand the mechanisms.

Moreover, the mysteriousness of a problem is not a fact about the problem, it is not a metaphysical feature of the universe -- it is an epistemological fact about us. It is about where we are in current science, it is about what we can and cannot understand, it is about what, given the rest of our understanding, we can and cannot imagine. It is not a property of the problem itself.

It is sometimes assumed that there can be a valid transition from "we cannot now explain" to "we can never explain", so long as we have the help of a subsidiary premise, namely, "I cannot imagine how we could ever explain..." . But it does not help, and this transition remains a straight-up application of argument from ignorance. Adding "I cannot imagine explaining P" merely adds a psychological fact about the speaker, from which again, nothing significant follows about the nature of the phenomenon in question. Whether we can or cannot imagine a phenomenon being explained in a certain way is a psychological fact about us, not an objective fact about the nature of the phenomenon itself. To repeat, it is an epistemological fact -- about what, given our current knowledge, we can and cannot understand. It is not a metaphysical fact about the nature of the reality of the universe.

Typical of vitalists generally, my high school biology teacher argued for vitalism thus: I cannot imagine how you could get living things out of dead molecules. Out of bits of proteins, fats, sugars -- how could life itself emerge? He thought it was obvious from the sheer mysteriousness of the matter that it could have no solution in biology or chemistry. He assumed he could tell that it would require a Humdinger solution. Typical of lone survivors, a passenger of a crashed plane will say: I cannot imagine how I alone could have survived the crash, when all other passengers died instantly. Therefore God must have plucked me from the jaws of death.

Given that neuroscience is still very much in its early stages, it is actually not a very interesting fact that someone or other cannot imagine a certain kind of explanation of some brain phenomenon. Aristotle could not imagine how a complex organism could come from a fertilized egg. That of course was a fact about Aristotle, not a fact about embryogenesis. Given the early days of science (500 BC), it is no surprise that he could not imagine what it took many scientists hundreds of years to discover. I cannot imagine how ravens can solve a multi-step problem in one trial, or how temporal integration is achieved, or how thermoregulation is managed. But this is a (not very interesting) psychological fact about me. One could, of course, use various rhetorical devices to make it seem like an interesting fact about me, perhaps by emphasizing that it is a really really hard problem, but if we are going to be sensible about this, it is clear that my inability to imagine how thermoregulation works is *au fond*, pretty boring.

The "I-cannot-imagine" gambit suffers in another way. Being able to imagine an explanation for P is a highly open-ended and under-specified business. Given the poverty of delimiting conditions of the operation, you can pretty much rig the conclusion to go whichever way your heart desires. Logically, however, that flexibility is the kiss of death.

Suppose someone claims that she can imagine the mechanisms for sensorimotor integration in the human brain but cannot imagine the mechanisms for consciousness. What exactly does this difference amount to? Can she imagine the former in detail? No, because the details are not known. What is it, precisely, that she can imagine? Suppose she answers that in a very general way she imagines that sensory neurons interact with interneurons that interact with motor neurons, and via these interactions, sensorimotor integration is achieved. Now if that is all "being able to imagine" takes, one might as well say one can imagine the mechanisms underlying consciousness. Thus: "The interneurons do it." The point is this: if you want to contrast being able to imagine brain mechanisms for attention, short term memory, planning etc., with being unable to imagine mechanisms for consciousness, you have to do more that say you can imagine neurons doing one but cannot imagine neurons doing the other. Otherwise one simply begs the question.

To fill out the point, consider several telling examples from the history of science. Before the turn of the twentieth century, people thought that the problem of the precession of the perihelion of Mercury was essentially trivial. It was annoying, but ultimately, it would sort itself out as more data came in. With the advantage of hindsight, we can see that assessing this as an easy problem was quite wrong -- it took the Einsteinian revolution in physics to solve the problem of the precession of the perihelion of Mercury. By contrast, a really hard problem was thought to be the composition of the stars. How could a sample ever be obtained? With the advent of spectral analysis, that turned out to be a readily solvable problem. When heated, the elements turn out to have a kind of fingerprint, easily seen when light emitted from a source is passed through a prism.

Consider now a biological example. Before 1953, many people believed, on rather good grounds actually, that in order to address the copying problem (transmission of traits from parents to offspring), you would first have to solve the problem of how proteins fold. The former was deemed a much harder problem than the latter, and many scientists believed it was foolhardy to attack the copying problem directly. As we all know now, the basic answer to the copying problem lay in the base-pairing of DNA, and it was solved first. Humbling it is to realize that the problem of protein folding (secondary and tertiary) is still not solved. That, given the lot we now know, does seem to be a hard problem.

What is the point of these stories? They reinforce the message of the argument from ignorance: from the vantage point of ignorance, it is often very difficult to tell which problem is harder, which will fall first, what problem will turn out to be more tractable than some other. Consequently our judgments about relative difficulty or ultimate tractability should be appropriately qualified and tentative. Guesswork has a useful place, of course, but let's distinguish between blind guesswork and educated guesswork, and between guesswork and confirmed fact. The philosophical lesson I learned from my biology teacher is this: when not much is known about a topic, don't take terribly seriously someone else's heartfelt conviction about what problems are scientifically tractable. Learn the science, do the science, and see what happens.

References:

Churchland, Paul M. (1995). The engine of reason; the seat of the soul. Cambridge, MA: MIT Press.
Crick, Francis (1994). The Astonishing Hypothesis. New York: Scribner and sons.
Jackson, Frank (1982). "Epiphenomenal qualia". Philosophical Quarterly. Vol. 32.
Nagel, Thomas (1974). "What is it like to be a bat?". Philosophical Review . Vol. 83.

Biography

Patricia Smith Churchland is a Professor of Philosophy at the University of California, San Diego, and an adjunct Professor at the Salk Institute.

She is author of Neurophilosophy (MIT Press 1986), co-author with T. J. Sejnowski of The Computational Brain (MIT 1992) and co-editor with Rodolfo Llinas of The Mind-Brain Continuum (MIT Press 1996).

She and her husband, Paul M. Churchland are the focus of a recent book, The Churchlands and Their Critics (ed. R. M. McCauley; Blackwells, 1996).

She has been president of the American Philosophical Association (Pacific Division) and the Society for Philosophy and Psychology, and won a MacArthur Fellowship in 1991.

Notes:

This paper is based on a talk I presented at the Tuscon II meeting on consciousness in April 1996. Many thanks are owed to the organizers of the meeting, and thanks also to Paul Churchland, David Rosenthal, Rodolfo Llinas, Michael Stack, Dan Dennett, Ilya Farber and Joe Ramsay for advice and ideas.

As I lacked time in my talk at Tucson to address the "Mary" problem, a problem first formulated by Frank Jackson in 1982, let me make several brief remarks about it here. In sum, Jackson's idea was that there could exist someone, call her Mary, who knew everything there was to know about how the brain works but still did not know what it was to see the color green (suppose she lacked "green cones", to put it crudely.) This possibility Jackson took to show that qualia are therefore not explainable by science. The main problem with the argument is that to experience green qualia, certain wiring has to be in place in Mary's brain, and certain patterns of activity have to obtain and since, by Jackson's own hypothesis, she does not have that wiring, then presumably the relevant activity patterns in visual cortex are not caused and she does not experience green. Who would expect her visual cortex -- V4, say -- would be set ahumming just by virtue of her propositional (linguistic) knowledge about activity patterns in V4? Not me, anyhow. She can have propositional knowledge via other channels, of course, including the knowledge of what her own brain lacks vis a vis green qualia. Nothing whatever follows about whether science can or cannot explain qualia.

Links

Pat Churchland's UCSD Faculty page

MIT Press notice of The Computational Brain by Patricia S. Churchland and Terrence J. Sejnowski.

Bruce Bridgeman's review of The Computational Brain by P. S. Churchland and T. J. Sejnowski in PSYCHE: an interdisciplinary journal of research on consciousness 1(3), December 1993.
From Bridgeman's review: "The broad goal of this book, expressed at the start, is "to understand how neurons give rise to a mental life." A mental reductionism is assumed in this seductively simple formulation. Indeed, the book represents reductionism at its best, as the authors guide the reader through the many intermediate levels that link neurons with mental life."

This paper appears on this site by permission of P.S.Churchland.

<table>
<tr><td colspan="3">To return to the philosophy discussion</td></tr>
<tr><td>
return to Philosophy index</td><td>
return to overall index</td><td>
return to Introduction and textual index</td></tr>
</table>

The 17th century view of the Mind

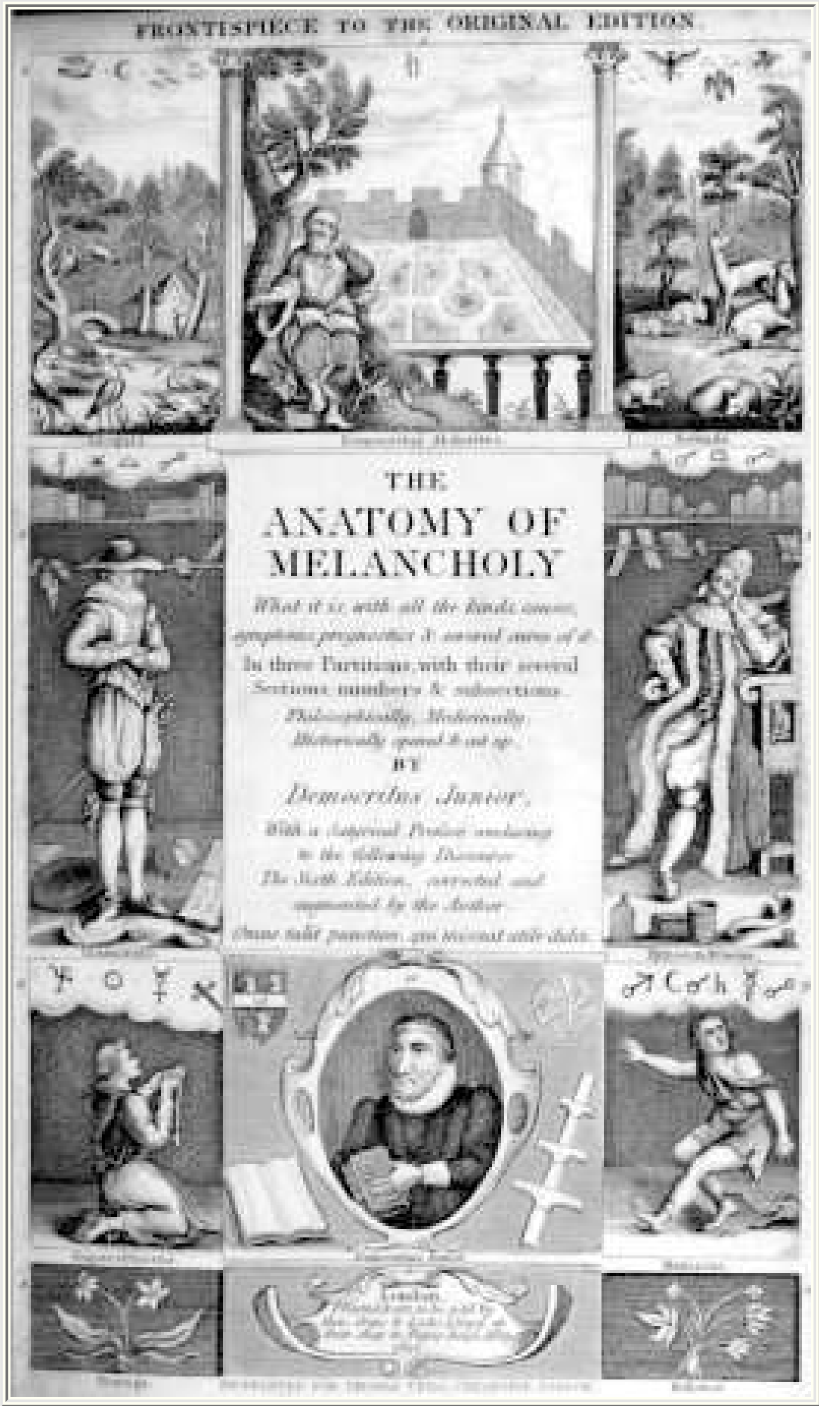
Here follows some notes on the shift from Galen's view of the mind and its body to the development of the mechanistic view.

Part of **The Brain Project** by Stephen Jones.

Galen's Humours

Galen's view of human anatomy became the framework for all further consideration of the body and its brain for the next 1500 years. Investigative inquiry into the anatomy didn't begin until Nicholas Copernicus challenged the prevailing Church backed view of the world as the centre of the universe by showing that the earth and the planets moved around the sun; and William Harvey demonstrated that the blood was pumped in circulation around the body. But the concept of "pneuma" still held sway in any discussion of the brain. Rene Descartes wrote, in the mid 17th century, in reference to the ventricles:

> "The cavities of the brain are central reservoirs...animal spirits enter these cavities. They pass into the pores of its substance and from these pores into the nerves. The nerves may be compared to the tubes of a waterworks; breathing or other actions depend on the flow of animal spirits into the nerves. The rational soul (the pineal) takes place of the engineer, living in that part of the reservoir that connects all of the various tubes. These spirits are like the wind. When they flow into a muscle they cause it to become stiff and harden, just as air in a balloon makes it hard." [Bergland, p61]



Title page from Robert Burton's *The Anatomy of Melancholy*, 1651. [from the 1849 edition in Stephen Jones' library]

Robert Burton in The Anatomy of Melancholy (first published in London in 1652) represents the "humours" view based on Galen. His book is possibly the first major treatise on a psychological problem, namely depression, ever published.

Burton summarises the state of anatomy with discussion of the humours. The four humors were:

blood [sanguine] a hot, sweet, temperate humour whose office is to nourish the whole body,to give it strength and colour.
pituita [phlegm] a cold and moist humour,

> his office is to nourish and moisten the members of the body.
> **choler [yellow bile]** hot, dry, bitter, helps the natural heat and senses, and serves to the expelling of excrements.
> **melancholy [black bile]** cold, dry, thick, black, and sour.

He also adopted the Aristotelian views on the nature of "life" referring to spirits:

> "Of these spirits there be three kinds, according

to the three principle parts, brain, heart, liver; natural, vital, animal. The natural are begotten in the liver and thence dispersed through the veins, to perform those natural actions. The vital spirits are made in the heart of the natural, which by the arteries are transported to all the other parts: if the spirits cease, then life ceaseth, as in a syncope or swooning. The animal spirits formed of the vital, brought up to the brain, and diffused by the nerves, to the subordinate members, give sense and motion to them all." [Burton, p94] and the soul (or the *anima*) which was divided

> "into three principle faculties - vegetal, sensitive, and rational, which make three distinctive kinds of living creatures - vegetal plants, sensible beasts, and rational men. How these three principle faculties are distinguished and connected...is beyond human capacity,... The inferior may be alone, but the superior cannot subsist without the other; so sensible includes vegetal, rational both; which are contained in it (saith Aristotle) as a tringle in a quadrangle." [Burton, p98]

He then goes on to describe the brain as a device for distilling the animal spirits:

> "...the brain...is a soft, marrowish, and white substance, engendered of the purest part of seeds and spirits, included by many skins, and seated within the skull or brain pan; and it is the most noble organ under heaven, the dwelling-house and seat of the soul, the habitation of wisdom, memory, judgement, reason and in which man is most like unto God; and therefore nature hath covered it with a skull of hard bone, and two skins or membranes, whereof the one is called *dura mater*, or meninx, the other *pia mater*. The dura mater is next to the skull, above the other, which includes and protects the brain. When this is taken away, the pia mater is to be seen, a thin membrane, the next and immediate cover of the brain, and not covering only, but entering into it.
> The brain itself is divided into two parts, the fore and hinder part; the fore part is much bigger than the other, which is called the little brain in respect of it. This fore part hath many concavities distinguished by certain ventricles, which are the receptacles of the spirits, brought hither by the arteries of the heart, and are there refined to a more heavenly nature, to perform the actions of the soul. Of these ventricles there are three - right, left, and middle. The right and left answer to their site and beget animal spirits; if they be in any way hurt, sense and motion ceaseth. These ventricles, moreover, are held to be the seat of the common sense. The middle ventricle is a common concourse and cavity of them both, and hath two passages - the one to receive pituita, and the other extends itself to the fourth creek; in this they place imagination and cogitation, and so the three ventricles of the fore part of the brain are used. The fourth creek behind the head is common to the cerebral or little brain,

and marrow of the back bone, the last and most solid of all the rest, which receives the animal spirits from the other ventricles, and conveys them to the marrow in the back, and is the place where they say the memory is seated." [Burton, p97]

And so in referring to the cause of disease and paricularly mental dis-ease he says:

> "...as the body works upon the mind by his bad humours, troubling the spirits, sending gross fumes into the brain, and so per consequens the faculties of it, with fear, sorrow, &c., which are ordinary symptoms of this disease [melancholy]: so on the other

side, the mind most effectualy works upon the body, producing by his passions and perturbations miraculous alterations, as melancholy, despair, cruel diseases, and sometimes death itself." [Burton, p164]

....................forgotten quite
All former scenes of dear delight,
Connubial love .... parental joy ......
No sympathies like these his soul employ;
But all is dark within ....
[Penrose. from the frontispiece to Burton's *The Anatomy of Melencholy*, 1849 edition]

To revue Burton: The rational soul was seated in the brain, and received sensations and controlled movement, via the action of the fluid 'animal spirit'.

The emergence of the mechanistic view.

It took a very long time and much valiant work (vide: Nicholas Copernicus and Giordano Bruno) to begin the liberation of science from the overarching control of the mystico/religious framework and the political needs of the Roman Church. This change started to really happen at the end of the 16th century with the appearance of a new attitude to the observation of what actually happens, followed up by a desire to experiment on and test what is being observed. But at this early stage the mystical framework still greatly influenced theory.

In A Short History of Science, Charles Singer notes Kepler's mystical adherence to the Pythagorean/Platonic solids and to the idea "that the arrangement of the world and its parts must correspond with some abstract conception of the beautiful and the harmonious" [Singer, 1941, p200].

Referring to Kepler's first approximation of his theory of the orbits of the planets, Singer says:

> "That Kepler sought so persistently for a simple mathematical scheme of the material world, and that, having found one, he regarded it as fitting his scheme of the moral world, suggests certain reflections on the workings of the mind itself. Whatever reality may be, we seem to be so made that we aspire towards an interpretation of the universe that shall hold together in a complete and reasonable scheme. The fact that we thus aspire does not in the least prove that such a scheme corresponds to reality. Nevertheless, all great religions attempt to provide such an interpretation. All become skillfully 'rationalised'.[Singer, 1941, p203]

It looks awfully like Singer adopts a vitalistic view of science here: that the motivation of science is to find a unified view of the "world" and that this in some way is a "natural" function of the mind. Yet this has considerable political consequence...

> "It is because science disturbs part of this already carefully rationalised field that religion resents its intrusion. The mind recoils from a dualistic universe, and rationalised religion usually seeks to minimise even such remnants of dualism as the conception of a spirit of evil. It is easy for us now to regard the opponents of Galileo and Kepler as purblind fools. Base motive

certainly prompted some of the opposition; but in essence the opposition expresses the reluctance of the human mind to adopt any teaching which disturbs it unitary conceptions. A reasoned view of the universe, physical and moral, had grown up during the Middle Ages. It would have been indeed a marvel if this had been relinquished without a struggle, for faith is not necessarily accompanied by either wisdom or learning or foresight." [Singer, 1941, p203]

The 17th century was a most remarkable period in its extraordinary fecundity of quite revolutionary ideas. That the earth travelled around the sun was only now being established. **Copernicus** had really only found that the Ptolemaic system of the Medieval period had too many anomalies (the epicycles) to allow it to stand against observation any longer. It took **Kepler** and Tycho Brahe to get the really useful data that allowed **Galileo** to finally publish (much to his trouble) his great synthesis **Dialogues on the Two Great Systems of the World**

Galileo conceived the world as reducible to measurement and mechanical principle. He was first to exploit the telescope and also instituted the use of telescopes and microscopes of high craftmanship as tools of investigation. That the heavens were vast and complex with a multiplicity of worlds was now mirrored in the startling multiplicity of matter and life in the microscopic world.

Francis Bacon in **The Proficiencie and Advancement of Learning** and **Rene Descartes** in his **Discours on Methode** laid down the principles of experimental science which we still follow. Firstly one should gather all the facts that are relevant to the matter being investigated. This selection of relevance is based on the work of one's predecessors with which one is familiar through study. Having gathered all the facts one forms them into an Hypothesis which links all the facts together. Then one

Title page of the *Dialogues on The Two Great Systems of the World* by Galileo Galilei (3rd edition, 1641) [from Stephen Jones' library] tests the hypothesis by experiment, modifying the hypothesis as required by the results of its testing. It is this which finally allows the development of a Theory.

With Galileo's development of the science of mechanics came the attempts by the biologists to explain the animal body as a machine. It becomes apparent to the experimental philosophers of the 17thC. that one might hunt out principles of a mechanical nature which applied alike to the motions of the heavens as they did to the earth and to living things. The world view of science becomes increasingly mechanistic. For example, in 1615, William Harvey discovered the process of the circulation of the blood and thus that the heart is a pump. The mechanical model of the heart as a pump stands as an early version of the process of using a working mechanical model to form a clearer picture of some part of the animal body. The classical microscopists, Malpighi and Leeuwenhoek, "discovered the corpuscles of the blood, the secretory functons of 'glands', and the fibrillary character of muscles, thus helping to complete details of the animal machine." (Singer, 1941, p243).

The Rise of Anatomy

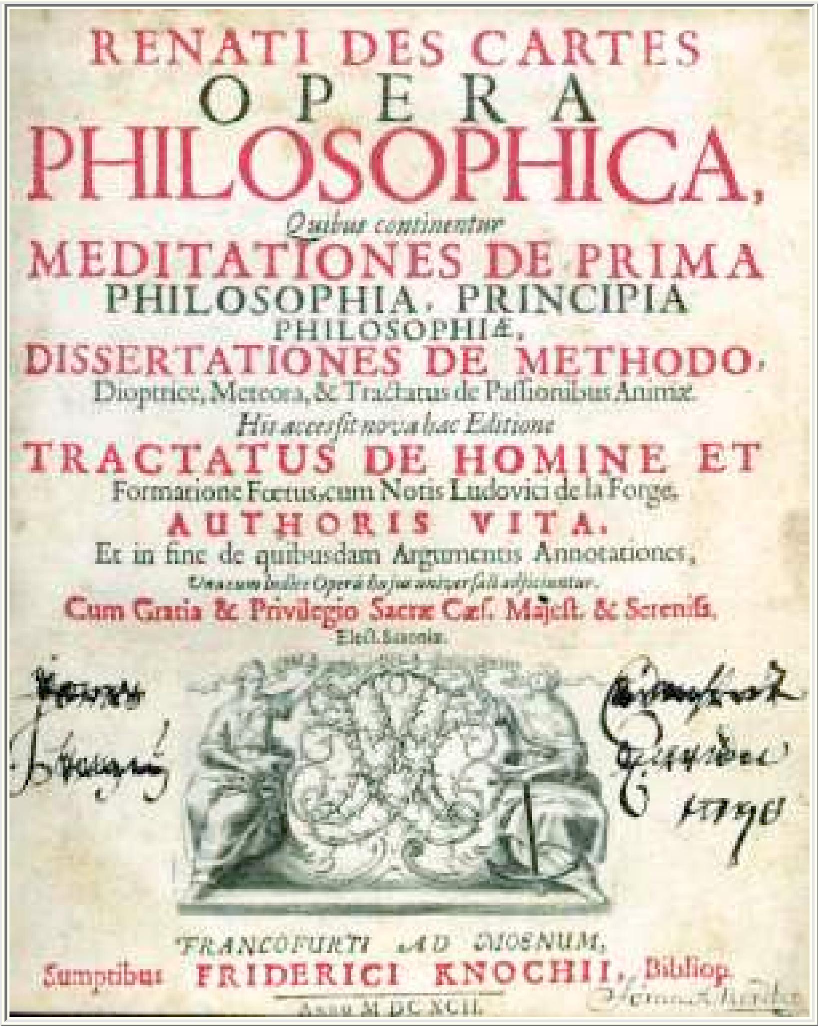
Uptil Descartes the rational soul was intimately housed in the brain. The humours which supported the activities of the various souls running the person could be seen and their pathways mapped (to a limited extent, given the difficulties in carrying out anatomical investigation, in obtaining bodies, imposed by the Roman church). The vegetal soul is in the liver, the animal soul is in the heart and the rational soul in the brain.

The role of Authority in teaching could not allow the questioning of handed down wisdom, especially as that wisdom was held by the Roman Church. During the darkness of the middle ages, the Church was the sole repository in Europe of the books and knowledge emmanating from the Greeks and the Romans. The Arab world had kept up a continuing spirit of inquiry through the middle ages but this material did not become available in Europe until it filtered out through the Moorish colonisation of Spain. Any re-appearance of information was controlled by the Church, they had control of the books and the institutions of learning, which, immediately before the Rennaisance where confined to the monastries.

They also carried the ideological power, to maintain the position of authority of the Church, with the Pope as God's representative on earth, essentially bestowing upon him the supreme right of decision making.

As with the clerical hierarchy so was there a hierarchy of social relations and a hierachy within the person and their body. The rational soul was available only to humans. The animal and vegetal souls, available to animals as well, were enough to deal with the bodily needs, both long term and everyday. The head was given a superior value through its position on top of the body and so it must be the seat of the rational soul. Further in what anatomical work was done, the main arteries carried the 'sanguine' to the head, and it was there that the vital spirit, the 'pneuma' was distilled out of the blood and distributed through the body by the nerves. So as the Pope was the head of the Church, and the man was the head of the household, the skull housed the brain which must be the head of the body.

Descartes reduced the humoural description of the body/brain with its variety of souls to a mechanical/hydraulic model. He used the most celebrated technological achievements of his time as his analogy. The great waterworks of fountains and water driven clocks and automatons, the showpieces of men of power, provided Descartes with models for describing how the brain operated the muscles and the general description of nerve process. But where now is the soul? Descartes demonstrated


RENATI DES CARTES
OPERA
PHILOSOPHICA,
Quibus continentur
MEDITATIONES DE PRIMA
PHILOSOPHIA, PRINCIPIA
PHILOSOPHIÆ,
DISSERTATIONES DE METHODO,
Dioptrice, Meteora, & Tractatus de Passionibus Animæ.
His accessit nova hac Editione
TRACTATUS DE HOMINE ET
Formatione Fœtus, cum Notis Ludovici de la Forge.
AUTHORIS VITA.
Et in fine de quibusdam Argumentis Annotationes,
Una cum Indice Operâ hujus universali adjiciuntur.
Cum Gratia & Privilegio Sacræ Cæs. Majest. & Serenis. Elect. Saxoniæ.
FRANCOFURTI AD MOENUM,
Sumptibus FRIDERICI KNOCHII, Bibliop.
Anno M DC XCII.

Title page from Descartes' *Opera Omnia* (Collected Works) 1692 [from Stephen Jones' library]

philosophically that we needed the capacity to keep some sort of 'reason-able' continuity, and the Church ideology demanded some sort of spiritual man which would be able to have continuity after bodily death to keep its carrot-and-stick control over the lives of its subjects and the source of its cash-flow. Thus a purely mechanical model of the human would not do. So Descartes divided the soul or the mind, the thinking thing, from the body and established Dualism as a way of thinking.

By a process of radically doubting everything of which he could not be absolutely certain, all sensation, movement, bodies, physical things were rendered unreliable. Finally only 'I' could be said to exist, I the thinking thing. All else is perceived only by a process of understanding, mediated by the mind.

So there is that about which Descartes is certain, i.e. the thinking thing, and there is everything else. He has separated the mind from all the world of sensations and physical things.

It could be argued that all Descartes really did was to separate the phenomenal from the physical. This had two consequences: for the physical, biological scientist it allowed ever more detailed and effective analysis of problem of elucidating what it is that allow living systems to work, but for the philospher it so utterly misdirected the agenda for understanding the phenomenal, the mind, that we still have not completely escaped its effect. Dualism still rides with us and the religious view still has enormous sway over the physical/biological sciences.

> Descartes ruptured "the traditional stair of life ranging upward step by step to man. Science since Descartes has repaired the stair and finds it more significant than before. It marks the way that man has climbed. And it is a stair of mind as well as body, and it is without break, man's mind nothing more than the topmost rung continuous with related degrees below." [Sherrington, 1940, p186]

In a sense it is the ongoing closure of the gap, opened up by Descartes, between body and mind which has become the characteristic of the development of neuroscience ever since. The increasing localisation of function and the increasing visual and conceptual magnification of the means of exploring the brain, show us more and more that the fine structure of the processes of the brain, the chemistry, interneuronal linking and organisation, can account more and more for the operations carried on.

On the role of modelling

When we make a tool we project ourselves onto the world. We create something which fits a mental model of the tool to achieve some goal, from turning over a large piece of rock with a stick used as a lever, to creating a mechanism in metaphor with which we can manipulate and represent our idea (eg. our idea of ourselves).

We seem to want to be able to explain the world in terms simpler than the operations of the world, i.e. reductionism. The models we use will in general be the latest or the most acceptable depending on how conservative we are. We need laws, spiritual or temporal to fix our relationship with the world and nature and God, if we consider the latter to be necessary. With the rise of a mechanistic description of the workings of inanimate nature, new models of how animate nature might work can be generated and thus the models of the animal as a machine.

> "A machine being a man-made contrivance, to call a living organ a machine implies that it is mechanism humanly intelligible. The whole man being organs the implication is that the whole man is mechanism humanly intelligible." [Sherrington, 1940, p.186].

Perhaps here lies the key to the mechanistic modelling, it is the urge to understand and the opportunity offered by modelling which drives the whole process. The spiritual/religious explanation denies the option of actually understanding the processes of nature while the mechanistic starts with the view that nature can be understood.

References:

Bergland, R. The Fabric of the Mind. 1985

Burton, R. The Anatomy of Melancholy. 1811

Sherrington, Sir C. Man on his Nature. 1940

Singer, C. A Short History of Science. 1941

Recent papers by Stephen Jones

1: Notes and Suggestions towards an Hypothesis of Consciousness.

A discussion of the operational definition of consciousness, criteria for ascribing consciousness to an entity, and a suggestion for a possible neuro-anatomy of consciousness, (based on the work of James Newman, et alia).

2: What would a Conscious Machine want to do for Itself?

If the criteria for ascribing consciousness applied to possible artificial consciousnesses, what then could we expect that device to want to do with itself?

3: A note on a possible physiology of subjectivity, and some comments on what a conscious machine might want to do for itself.

A concatenation of the above two papers, somewhat shorter, presented at *Consciousness Reframed*, 1997.

 return to Chapters index	 return to overall index	 return to Introduction and textual index

Notes and Suggestions towards A Theory of Consciousness

Part of **The Brain Project** by Stephen Jones.

Abstract:

I open by considering briefly some of the characteristics of ordinary consciousness which seem to have gained consensus acceptance among those studying the problem of how the physiological systems we are attain subjectivity. I then look at the Thalamo-Cortical system so ably described by Newman and offer some observations on the roles of feedback processes, propagation delays, resonant systems and Llinas' (et alia) 40Hz oscillations. These factors, I argue, provide a way of locking together the phenomenal, subjective world and the physiological substrate on which consciousness runs. I then discuss the role of culture in forming brains and consciousnesses, and close with the suggestion that there is no "Hard Problem": that the phenomenal and physiological representations are two sides of the same coin.

I want to open by briefly covering some factors of our being which we usually include as being aspects of consciousness.

1. What is Consciousness ?

To know the world and to act independently within it and to do things for oneself. Reflection within on the patterns within as well as reflection off the world by interacting with others and by projecting into the world for puposes of generating feedback.

1.1. Consciousness is the "**I**", the "**self**" that we all know, from which we view the world and interact with it, that sense of (to quote John Searle) "subjective qualitative states of awareness, sentience or feeling"(1). In the 17thCentury Rene Descartes showed that no matter to what extent our senses might be deceived there would still remain a something which could be called "myself" even if it were utterly deceived as to the existence of any one or any thing else. I suppose his "*cogito ergo sum*" might have been better put if he'd said I am deceived therefore I am. William James (2a) (in the late 19thCentury) put it as having a sense of a personal consciousness that is ours, not something that we share [though in passing, how we consider this against the matter of the social construction of our world view remains to be discussed].

1.2. Everything we know is a function of **experience**, either through sense perception or reflection upon that experience. The mind or the "I" is born empty of knowledge of the world. As John Locke (3) described it we are born "*tabula rasa*" (or a 'blank slate'). It is only by our experience of the world that we gain ideas of it. There are no "innate ideas".

> "Let us then suppose the Mind to be, as we say, white Paper, void of all characters, without any Ideas; How comes it to be furnished ?....To this I answer, in one word, from Experience... Our observation imployed either about External sensible objects; or about the Internal Operations of our Minds, perceived and reflected by our selves... These two are the Fountains of Knowledge from whence all the ideas we have, or can naturally have, do spring. [Locke. *A Treatise of Humane Knowledge*, 1721, p67]

It is the data of sensation that have the qualities or "qualia" which are the stuff of our subjectivity, that which we know.

1.3. Consciousness is a function of the state of our central nervous system, i.e. the **physiology is the substrate upon which consciousness runs**. To quote William James in his Text Book of Psychology, 1892 (2b):

> "The immediate condition of a state of consciousness is an activity of some sort in the cerebral hemispheres... One has only to consider how quickly consciousness may be abolished by a blow on the head...[or] by a full dose of alcohol...to see how at the mercy of bodily happenings our spirit is... Destruction of certain definite portions of the cerebral hemispheres involves losses of memory and of acquired motor faculty of quite determinate sort...Taking all such facts together, the simple and radical conception dawns upon the mind that mental action may be uniformly and absolutely a function of brain-action, varying as the latter varies, and being to the brain-action as effect to cause." [James, 1892, pp5-6]

1.4. Our state of consciousness:

> **a/** is always changing as we are exposed to continually novel sensations.
> **b/** is selective of what it pays attention to, and
> **c/** provides a sense of temporal continuity which the normal day-to-day changes of sleep and wakefulness, as well as abnormal changes such as unconsciousness, do not interrupt.

1.5. Paul Churchland (author of "The Engine of Reason; The Seat of the Soul") (4) has described some salient aspects of consciousness, which we would need to explain in a theory of consciousness.

> **a/** Short-term memory and its decay.
> **b/** Directable attention, or conscious control over what we attend to and what we do.
> **c/** Multi-valent comprehension through "mulling" or reflection.
> **d/** Independence from sensory input in say, daydreaming.
> **e/** The disappearance of consciousness during sleep.

f/ Unity across the senses and unity over time.

1.6. Robert Kirk (5) (of Nottingham University) has developed a concept of what he calls the "**basic package**" of capacities that an organism must have for it to be conscious. This is first about being able to collect information which is specifically intended for the organism's own use, and then about its capacity to decide what to do with that information. There is a package of related activities which apply to this collecting information.

a/ The organism must be able to use the information.
b/ The organism must be able to initiate and control its activity on the basis of the information it collects.
c/ The organism must be able to assess the information for its usefulness or interpret it.
d/ The organism must be able to assess the situation it is in so that it can decide how to respond or whether it should respond.

1.7. Our ability to respond to **novelty** in active and constructive ways is considered to be somehow over and above the mere 'irritability' of the senses and the body's reflexivenss to sensations which may or may not then have conscious impact. The plant which follows the sun across the daily sky is not able to refuse to do that. It has no opportunity to decide whether or not it will not respond to the tropism of the sun. It can only do it, there is no capacity to act independently of this tropism.

1.8. So ultimately the problem for a theory of consciousness is to explain how it is that the physical system that we are, this bundle of cells and organs and nerves, can have **subjectivity**, can behave independently, can do things for itself and can respond to input in a way that takes account of whim and ideology rather than simply reacting to current conditions. This is the question that David Chalmers (6) asks in his formulation of the "**Hard Problem**":

> "Why is it that physical processing in the brain, no matter how sophisticated, should give rise to any subjective inner life at all? Why couldn't that have all gone on in the dark? That's the real mystery." [Chalmers: Brain Project interview]

1.9. Two views of the relationship of the brain to the mind or consciousness seem to have developed since Descartes. The first is the "dualistic" view that consciousness somehow exists independently of the physical world and body. This position derives from Descartes but is perhaps more akin to the modern religious view of the "soul". The second is the physicalist view which says that in some manner consciousness is a direct function of the physiology. This view was espoused early on by de la Mettrie in his very radical 17th Century work "*Man a Machine*" (7).

My argument is essentially that it is not necessary to dualise the mind away from the brain, but that normal physiological functioning of the brain includes, within the array of its processes, that which we consider to be consciousness. That is, that consciousness is an *inescapable* result of an immensely complex but highly organised information processing and representation system.

2. How we know.

2.1. The multiple ways of talking about consciousness are **representations**. So we have the diverse array of philosophical representations, the physiological representation, the phenomenological representation, the neural net representation, etc., which I argue are all simply different ways of talking about the same thing. Representation is transformation of information from one modality to another. Within a physiological discussion (representation) of sensation; colours, sounds, feelings and other qualia are phenomenological representations of those sensations coming in through the eyes and ears and proprioceptive senses. Each layer of information processing in the brain is a transform of the input representation into a new representation. The very act of seeing is in fact a mediated process through a series of layers of representations of which we are largely unconscious into a series of representations which are bound (usually inextricably) with sense representations from the other modalities; hearing, touch, etc. and of which we *are* in some way aware or conscious. **What we know is entirely mediated by the sense organs and the brain. What we know of the world is inference.** So we have two areas of study, on the one hand that which physical and biological scientists do, and on the other, that which philosophers and psychologists and cognitive scientists do. For simplicity I will call these the physiology and the phenomenology of consciousness.

2.2. Phenomenology is the study of what we actually know of the world; **qualia,** the stuff, **the contents of consciousness**, rather than the study of what we infer about the world based on that phenomenology, which is the business of the physical sciences. How we know it is one aspect of the phenomenology. We perceive and report our perceptions. We imagine and reflect on the contents of the phenomenology. We report and interact through language which we gain through the culture as we grow up. Culture and language mould our consciousness and render us human. But, ultimately, we see colours, hear sounds, feel emotions all through the use of the brain. **Consciousness, whatever its relationship to the physiology, requires that physiology as the substrate on which it runs.**

2.3. So it is the study of the physiology of the brain that is our first task in developing a theory of consciousness. Whatever we may think about the phenomenology, whether we take a mystical view or a physicalist view of consciousness, there is still the information processing and as far as anyone can tell all information is necessarily embodied in something physical.

3. Physiology

3.1. As Hughlings Jackson (8) said in 1876:

> "All nervous centres, from the lowest to the very highest (the substrata of consciousness), are made up of nothing else than nervous arrangements, representing impressions and movements...I do not see of what other materials the brain can be made." [Hughlings Jackson "West Riding Asylum Reports" 1876 p267].

And William James asserts that there is a complete parallelism between the arrangements of nerves and the ideas to which they project:

> "The muscles and the sensitive points are represented each by a cortical point, and the brain is nothing but the sum of these cortical points, to which, on the mental side as many ideas correspond." [James 1891, p30] Thus it is the "motor and sensory ideas variously associated [that] are the materials of the mind". [James, ibid]

3.2. So we need to know something of the **physiology** and the **organisation** of the brain to enable us to at least see on what it is that consciousness operates, that is the substrate or the emodiment of consciousness. [And there are some considerable differences in the implications of these two terms for our final theory].

3.2.1. For the sake of discussion, the system of the brain can be broken up into a number of interlinked parts. These are, essentially, the bodily connections through the spinal cord and the brain stem, the reticular activating system at the top of the brain stem, the sensory modalities and their connections to the cortex through a central relay station and the cortex, to which all sensory data may be sent and in which much of the interpretive and planning and control processing is done. It is the central relay station which will interest us most here. This central relay system is known as the **thalamus**. It, in linkage with the cortex, the basal ganglia, the hypothalamus, the hippocampus and several other structures, forms the **thalamo-cortical system**.

3.3. In an excellent review, published as an electronic seminar on the Internet, Jim Newman (9) has demonstrated the functions of the thalamo-cortical system and its tributaries in being the site (or better, the system) most likely for the embodiement of the major functions of day-to-day consciousness and the processes of integration and control of the informational structure through which we have our place in the world.

3.3.1. To summarise the thalamo-cortical system. The thalamus acts somewhat as the hub in a wheel, the spokes of which are nerve bundles travelling from the body periphery (carrying sense and bodily data) and which are then relayed up into the cortex and cortical association areas for interpretive processing. All of the sensory pathways (with the exception of the olfactory) are routed through the thalamus. For example, the optic tract runs from the retina, through the optic chiasm to the thalamus and thence into the lateral geniculate from where it is distributed into the occipital (or visual) cortex at the back of the brain. Auditory data from the inner ear is relayed through the medial geniculate into the auditory cortex in the temporal lobes. All of the face and body's proprioceptive data is routed through the thalamus on its way to the somato-sensory cortex. These are **ascending pathways**.

3.3.2. At the same time there is a vast array of nerve bundles descending from the cortical areas onto the intralaminar nuclei and the nuclear reticularis in the thalamus. These **descending pathways** act to gate the sensory data being presented to the cortex and it is in this capacity for the cortex to control what data it is being sent at any moment that we can find the function we call selective attention. Also nerve bundles from the frontal and prefrontal areas go via the basal ganglia to the thalamus where they are integrated with sensory data to help in the control of motor functions.

3.4. Essentially what's going on is that there is an array of massively connected **feedback control circuits**, **organised horizontally** around the thalamus and the basal ganglia and various emotion function nuclei **and vertically** between the cortex and the thalamus. The former (the horizontal) give behavioural control and the latter (the vertical) provide sensory control, especially in preventing the cortex from being overwhelmed by sensory input.

3.5. Also, in the cortex are vast arrays of intra-cortical nerve connections which probably provide the capacity to associate different sensory modalities and to interpret grouped or bound collections of data from different senses that allow one to, for example, recognise that the sounds you hear are coming from the mouth you see speaking to you, and that the individual whose mouth you are watching is saying things that have meaning.

4. The organisation of things

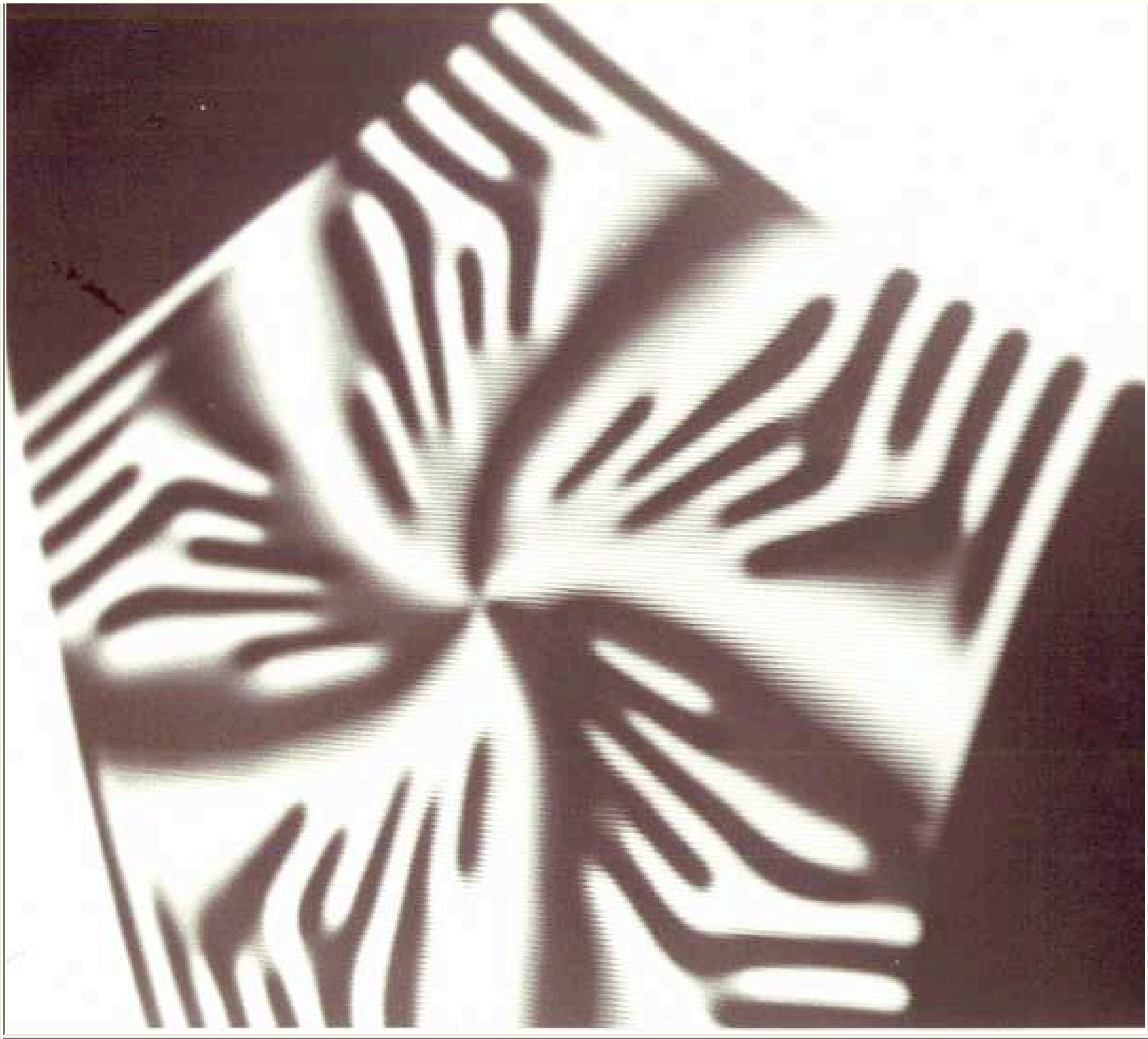
4.1. But it is the purpose of the descending pathways which I want to concentrate on for a moment. These descending pathways act as a control system. The data the cortex is being fed from moment to moment is determined or controlled by the cortex. We have operating here a kind of feedback network which turns the whole thalamo-cortical system into **a self-regulating process**. This capacity for self-regulation forms the basis for almost all levels of life, from the regulation of cellular metabolism to the behaviour of individuals in society.

4.2.1. So what is this **feedback** thing? In a system which, say, acts as a buffer for a subsequent processing system the results of the process are compared with the needs of the system and any difference (or error) is analysed in such a way as to provide an indication to the buffer of what is needed by the susbsequent processor. This feedback can be either negative (i.e. inhibitory) with respect to the input, or positive (i.e. excitatory) with respect to the input.

Wiener's generalised feedback process (from Norbert Wiener. Cybernetics, 1948 (10))

4.2.2. In the brain, the cortex feeds back to the thalamus a complex set of inhbitory and excitatory controls which allow it to have a measure of control over what it is being fed, thus it does not get overwhelmed by the immense amounts of sensory input which are to a large extent unnecessary for its survival. If this feedback control system did not exist it would be as though we lived permanently in the grip of an LSD experience, and we and all of society would collapse.

4.3. In an **engineering context** an electronic amplifier is regulated by a certain proportion of the output being fed back into the input as an inverted (or negative) representation of that input. But, and this is a crucial point, it takes time for this process to occur. It takes time for the electrons to travel through the circuit to the output. This is called **propagation delay**. If the frequency of the signal, and the propagation delay through the amplifier, is such that the output of the amplifier, when fed back to the input is positive (or non-inverted) with respect to the input then the amplifier becomes an oscillator. You've heard this numerous times when a microphone is placed in front of the speaker to which it is being fed. The combination of feedback and propagation delay causes a circuit to reverberate or act as an oscillator. It becomes a **resonant** circuit which may be considered as holding the information in the circuit for longer than the period of the original information. This is a form of memory.



Video Feedback shows the effect of propagation delay in creating an oscillating system which can undergo considerable perturbation and still be stable.

4.3.1. In the thalamo-cortical reverberatory system this will be what we know as **short-term memory**. But there is a bigger system here. Myriads of feedback pathways are operating, propagation delays of all sorts of intervals are involved and so we have a very complex, but organised and self-regulating set of systems which is the activity of our brains. The period of the resonance might well be in the order of several hundred milliseconds, which is similar to human reaction time. This short-term memory period is our present, our being-in-the-world. We don't feel the world as being a succession of instantaneous states but as a period of connectedness with things before they become the past.

4.4. Now Rodolfo Llinas and others, particularly using magneto-encephalography, have remarked on a rhythmical 40Hz pattern travelling "across the entire cortex" (Newman). This is a field being sensed by the detector with a duration of 25 milliseconds before it collapses again. (Is this some sort of function of the refractory period of a neuron's activation sequence?) This rhythmical field will be the result of coordinated neural activity which happens in steps (or cycles). In the engineering world we call this a clocked system. Clocked systems grab the state of the system and hold it till the next clock cycle where they grab the next state. The state of the system is refreshed every 25 milliseconds.

4.4.1. Given the variable propagation delays through different sensory pathways **a means of synchronising and holding various input so that they can be bound together and become as different aspects of a single event is provided by a clocked system**. This will also provide what is perhaps the so called eidetic memory into which we dip for confirming what was just said for example. Or perhaps from which we "hear" ourselves speak and know what we are saying.

4.4.2. Presumably also some aspects of the past state of the system are retained by the current sample and fed back into the current sample via descending processes from the cortex, enabling the sense of being in a continuing phenomenal field of the present.

4.5. But this would only apply to the conscious system, what we are actively attending to, the stuff of our consciousness currently. Now, obviously, we all also have instant response requestors and reflexes which operate at the immediate sense level and use express routes into the attention setting sub-system. So there may be higher processing frequencies, or clock rates, with other faster (for example music listening) systems being active. Though the massively parallel nature of all sensing sub-systems presumably enables this full sensory awareness to occur at slower speeds than otherwise necessary, (that is, at sampling rates below what would be necessary in serial sampling for sensory awareness).

4.6. In the magneto-encephalographic frame this field of activation, showing up as 40Hz oscillations, is a matter of nerve propagation and the activities of electrons and electro-magnetic fields. I want to suggest that in order to describe what goes on in these reverberatory fields we might use the concept of **phase modulation** as a means for describing the process of the large-scale behaviour of an organised complexly interacting system. That is, the electro-magnetic/chemical activity of the system is describable within phase-modulation terms.

4.6.1. I will elaborate: in a dynamic system, **phase** is a description of the timing aspects of various parts of the system with respect to one another. If we look at an event at one moment, look away to something else and then look at the first event again we will see that event at a different *phase* of its process. Phase describes sets of relations over time and as these sets of relations are altered by, say, contextual factors, the system may be said to be phase modulating.

4.6.2. Behaviour impacts on the nervous system which is undergoing this immensely complex phase-modulation live. The thalamo-cortical system, fed by perceptual input and the brain stem, and supported, analysed and "driven" by the cortex is an immense feedback network which, if happening in an electronic circuit environment, could only be considered a resonating circuit. Notice that propagation delay will play an important role such that the propagation delay and its results in the adding and subtracting of waveforms (i.e. phase modulation) contributes to the resonance of a system.

4.6.3. Any input, fed forward from sensory input systems or fed back into the circuit from cortical association and control structures will affect the "shape" of the overall reverberating circuit. The amount of any signal in the circuit will be affected by the intensity of the input and the feedback structures.

4.6.4. Signals being propagated through the thalamo-cortical structure or any other structure in the brain are embodied in nerve processes and grouped as assemblies, the magnitude of which are determined by the needs of the active system at the time (a la Greenfield (11)). These nerve processes have at least an electronic aspect (as well as a neurochemical aspect) and as such will create an electro-magnetic field within and about the physiological structures of nerve assemblies. This electro-magnetic field is what is detected in magneto-encephalography, and because it is embodied in a physical strucure it will display a kind of "shape" showing the phase relationships among different representational aspects of the current condition of conscious knowing and the near-conscious data contained in various support structures' informational processes. Any instantaneous state of the system can be described by a phase diagram (albeit an extremely complex one) and as the system changes in real time, according to new input, reverberent memory and associations, the phase state of the system will be modulating. Thus a phase modulation description will map the ongoing changing states of the physiology and of consciousness.

4.6.5. It should be obvious from what I have been saying here that the **changes in phenomenal consciousness generated by a new input, a new decision or whatever, will take place in actual physiological nerve processes and these physiological changes will thus change the electron transport conditions in the nerves. Electron transport can be detected as electro-magnetic fields and so we have directly detectable physical concommitants of phenomenological activity.** Of course our detection equipment would need to have extraordinarily sophisticated signal processing to tease out the contribution of any particular input or thought to the overal shape of the field, but in principle it could be done.

4.6.6. The nervous system is an electro-chemical system undergoing phase modulation and propagation delay and resonance with the world. The period of the propagation delay may well be a couple of hundred milliseconds, enough to give us the present we are always with. The buzz of representations and productions in the brain are us and we represent them phenomenologically and investigate their physiology and their physics. We are inside this resonance, we live it, we are it.

5. A couple of observations on phenomena

I would like to mention a couple of phenomenal events which say interesting things about some of the conditions of this reverberatory process of being conscious.

5.1. First. At Tucson II, Daniel Dennett gave a demonstration of a visual phenomenon in which a person looking at an image on a monitor screen was simultaneously watched for eye movement, or saccades, as the subject focussed on different aspects of the image presented. When one such saccade occured the image displayed was changed in some way, usually quite strongly, for example the colour of a coat worn by someone in the image would change colour. The person being tested would almost invariably be unable to detect any change in the image, being unable to describe what had changed. It was usually not until told about the change, that they would suddenly see it.

5.1.1. Now the **saccade** is of interest for me here. I would suggest that the saccade happens in the collapsed period of Llinas' 40Hz rhythmical field and thus we don't see anything during that period. As we are triggered by difference (i.e. the activity of things changing, the actual changing) as our primary comparator, the slower cortical comparators don't have the data of a noticed change to set them up (or off). One of the reasons for this to happen is that the collapsed period, in some sense means no seeing (or whatever). During a saccade the shifting pattern of light on the retinas moves very fast and all sorts of confusing data would be provided and so it might be necessary that a number of critical circuits, particularly motion detectors, be disabled while the flick of the eye occurs so that we don't "spin out".

5.2. Secondly, I want to mention another visual phenomenon that shows some intersting relations here. I've often noticed, oddly enough mostly while sitting in restaurants, that a person will walk past an open doorway, say walking along the street, and I will realise that I only noticed them being in the doorway after there had already developed a significant space behind them, i.e. between them and the door jam on the edge from which they had entered the scene. So this must mean that it takes a considerable period for the recognition of most mundane, non-threatening events, possibly an interval of 25 milliseconds, which is the period of these putative "sampled events in consciousness".

6. Now, back to the cortex itself: Addressing memory.

6.1. In the cortex there is also a layer of horizontally connected nerve processes. These are the association pathways which store and associate all the array of stimulus which has become memory. I wonder whether a neuron assembly which is involved in memory processing could be able to assist in the remembering of different things overlaid onto the one network by controlling the triggering of patterns of data through different synaptic structures applied to a particular neuron. Nerve nets are programmed dynamically by weightings that might also recede to zero (off) just as they might vary only slightly around a mean value. The same nerve net will produce a great array of different outputs depending on the weights applied at its inputs and so a number of different, say, "concepts" could be represented by the same nerve net, depending on the representations of input and recurrent data available in all the contextual, i.e. surrounding, neural elements in the net. These "patterns of activation" [P.M.Churchland's phrase] could be seen as a kind of addressing structure in which the values represented serve as addresses (or pointers) for concepts further upstream.

6.2. A neuron structure will then be able to assist in the storage and retrieval of a number of different memories from an assembly through this variable addressing structure. This addressing structure will presumably emulate the original patterns of the laying down of any particular memory trace. In a sense the addressing structure is the memory and it triggers activators (or representation processors) to work in particular ways in assembling different memories (that is re-assembling the past).

6.3. These addressing structures are set down by experience, by learning, by practice in the environment particularly through childhood where the basic patterns are set down. The culture, in which the individual grows up, sets up the meme structure which is actively emodied on the inter-neuronal synaptic structure. Memes then become addressing structure with a phenomenological frame or representation. Notice the role of culture.

7. The Culture dimension

7.1. What of this cultural dimension? Culture and society provide the sources of the modulating activity of our day-to-day lives: the external/social/perceptual and the internal/generative/reflective: We are interactive systems within the culture. We have an active relationship with others and the world around us be it in real space, mythic/cultural space or cyberspace.

7.2. Each of us has active relations with people and things. Our social interactions operate in the face-to-face, the personal, the social; and the remote (e.g. publishing a book, or producing art). They operate at many scales, and depending how much social power our output has, within greater or lesser segments of the society or culture. We generate output, from an embrace to the Magna Carta, and this output has some sort of impact on some other entities in the culture. The people and structures upon whom this output impacts will then, should they so chose, be able to produce responses which are feedback to us, if we choose to receive it. So **one is in an active relation with the world around: what we do changes it, what others do changes**

us. Now the content of this activity has meaning and this is acheived through the operations of our brains, bringing the sensory input into consciousness.

7.3. Culture modulates and guides the growth and development of a living conscious entity. Culture shapes the content and the interpretive structures of the brain. We grow into our current self(-identity). What we experience shapes the maturation of the nervous system and the growth of our "minds".

7.3.1. As an infant develops any move of the muscles stimulates afferent nerves. This stimulation initiates myelination of the nerve fibre, and also asserts in the brain the existence of the fibre and its mapping into whichever cortical area it is involved with. Adjacent fibres carrying stimuli (signals) from adjacent areas on the sensing surface are similarly myelinated and mapped into the cortex. The infant will at first only sense inchoately. As more input occurs mappings will be consolidated and refined. At the same time efferent nerves are carrying signals which initiate movements, the nerves are myelinated by their use and the muscles are stimulated to develop in their ability to respond. As muscles move they impact with external objects and stir internal proprioceptive sensors, thus returning signals to the appropriate sensory cortexes which carry feedback on the muscle action.

7.3.2. In the brain nerve processes from sensory areas feed data to other areas of the brain including direct, and higher level, motor control areas. When the infant hand impacts upon an object the grasping reflex leads to attempted interaction with the object which further stimulates the sensory systems. If grasping is not possible, say because of the size of the object, then other action will tend to take place to compensate in some way, by, say, opening the hand more. The activity of reaching and touching stimulates the nerves in the arms to grow and myelinate and stimulates the differentiation and mapping of the nerves in the brain which handle sense data, feeding it into control centres, differentiating and mapping these nerves. This provides a substrate for finer control over the muscles. A feedback loop of refining control by successive approximation results, our aim becomes better and better; and soon we are reaching, holding, pulling, crawling and so on better and better. Maturation of the brain and consciousness takes place in direct relation with maturing of the body.

7.4. If the brain is seen as a numerically immense collection of organised systems of nerves and their interconnections and the supporting wetware that keeps everything operating, we can represent it as a collection of neural nets in which the whole range of distributed processing tasks, which we use either in or out of consciousness, function to keep us operating in the world. All the patterns of stimulation which flow through this massively parallel distributed processing system are originated in the body's ongoing becoming in the world, or in external data from the world and our reflections upon that. So the connectionism that develops among neural processes is determined by the stimulus they receive and the kinds of systems available for handling that stimulus. If the so-called "weightings" of the synaptic connections are dependent upon the exercise of stimulated neural processing systems then the world as we know it will be embedded in the data structures set up within these neural nets. At birth there will be little beyond the basic propensities of the brain: to learn to crawl, to walk, to recognise a face, to acquire language: no content as such. The contents of consciousness and thus our being in the world are embedded in us by stimulus we acquire by being in the world.

7.4.1. Setting up the addressing patterns/structures suggested in section 6 is started here, at birth. Socialisation of the infant into the family, language and society at large, establish and enable consciousness in ourselves as individuals. **The surrounding culture is an intrinsic and neccessary part of the process of the emergence of consciousness in all humans beings**. An entirely isolated individual would fail to develop many of the aspects of consciusness which I canvassed in section 1.

7.4.2. Daniel Dennett talks of units of meaning, **memes**, that are the content of our consciousness and no dougbt active below consciousness as well. Among other things the brain is a system for acquiring, elaborating and reporting memes. Their propagation in society is enabled by their operation within the system of our brains (the processes of our physiology) in which memes have their duration and their change and flow. The physiology is the skein within which these things are embodied, given substance. **Ideas only exist in brains and in the cultural, in-the-world-manifested projections of our minds, i.e. the models we project onto the world and the things we make of the world, be they books or buildings.**

7.5. The contents of the brain may well be like the contents of a language, codes for those things apprehended, as words are codes, signs standing in for the object. There is only the complex of processes (the patterns of activation, the addressing structures) standing for the object in the brain. There is a sense in which the known world is not congruent with what is "out there" in that everything we know of the world is contained in the processing system which we call the brain. Computer graphics and imagery are often spoken of as being "simulacra" of things in the world, the "virtual" as distinguished from something supposed to be real, but all we know of the world partakes as simulacra. What is contained in the brain/body/self is (an encoding of) the current flow through, and resonances of, the data of the world: social, physical; in light and sound and smell.

7.5.1. What we know is not the world, but our sensory processing of its waves and disjunctions. It is our culturally derived representations of what we have experienced. When an infant is born and commences interacting with all about it its nerves are sparked into the commencement of maturation. The discernment of the buzzing, blooming confusion reaches focus and sound is differentiable enough to be from specifiable sources and resolves into language and light resolves into recognition. The brain is in a very strong way being wired up (more correctly the pathways not stimulated drop away, atrophy). The internal structures of the nervous system physiology are the coat-hangers upon which the coats of the culture are being hung. Each individual will be unique, having been exposed to a unique stream of stimlus and context. The wardrobe of self will be entirely distinctive made up from the available elements and styles of the culture surrounding. If people don't wear shirts then shirts are unknown. One might argue that our representations as internally generated actually project out meaning onto the fabric of unknowability.

7.5.2. Our physiology develops via interaction with its environment, both in terms of its immediate physical environment as contacted via movement and touch (for the drawing out of mobility and engagement), and in terms of the development of the social-interactive capabilities that render our individual consciousnesses human. As visual and auditory sensations become sorted out the nerves establish the connectionism of their nets, and personal, cultural and linguistic capacities develop. We are dependent on the stimulation we receive from the world for the growth of our physiology as much as through any drive on the part of the physiology itself. It is as if growth were drawn out of us by the world, the fabric of the source of all our sensations. This process never really stops, the ripples and ramifications of our being in the world feed, enfold, inform the world that we know, informing, enfolding and feeding us. In this larger scale, culture is the driver of the "phase modulation" of the physiology of consciousness.

7.6. Emotions and chemical modulation. Our existence in the world requires input from the world, we are not closed systems. We require food and air and sunshine and stuff and these have various effects upon our physiology. We also engage in social contact and this also has impact on the body's physical being. Physical/social events produce physiological responses both as the process of moving our arms or our mouths and as the biochemical impacts of the metabolism which enables a muscle to move. The brain monitors this behaviour and feeds back into the body with desires and needs which also produce biochemical events. All of this stuff starts to become a (somewhat simplified) look at the role of emotions in consciousness. The appearance of highly emotional stimuli will change our hormonal chemistry balance and this at least in some sense will have a modulatory impact on consciousness. If we feel hungry our consciousness will try and direct us towards feeding.

8. In conclusion:

A summary of what I think might be going on to produce consciousness.

In a complex flow of world processing, if we change (arbitrarily) the state or the sign of one aspect, then the timing of this change will effect/affect the phase-shape of the waveform of that processing flow. In humans this change of state is psychological/phenomonolgical: we have decided to alter the conditions, but it alters the physiological conditions as a direct result. Which came first? The decision or the conditions requiring the decision. It matters little, they come from both directions, embedded in the physiology where all this goes on.

This assumes that we have an identifiable system through which this world-processing occurs. In Bernie Baars' terms this is the Global Workspace (12), in James Newman's and Rudolfo Llinas' and John Taylor's and Joseph Bogen's terms the thalamo-cortical system (9) is the work horse, the structure through which this occurs. The two classes of representation used here map directly onto each other. The downward control flow from the cortex is the pathway through which a psychological decision (e.g. a change of plans) is translated into the global consciousness and the physiology.

A formal description (a model only, an abstraction) of the particular planning process will have decision points within the streams of data where, so to speak, "collision" or merging occurs between streams. One might call it a functional description of the processes, or functions, engaged at any particular moment. Conjunctions between ideas will occur, new perceptions enter, some demanding immediate attention, priority. Planning needs to be altered: "Don't bump into me!". The link between perception and planning, between hearing and interpretation is through the thalamo-cortical system. What to attend to is determined as much by input, (e.g. an interruption) as by high level decision about what word to put next.

The real time "results" of "running" the function description is the process of being of an individual. (And on another scale, where the exchange and flow of ideas and goods in a society describe the dynamic "state", the fluid flow, of that society).

We are a phenomenological state machine implemented on a physiological substrate. Consciousness will be a subsection of this state-machine running day-to-day, moment-by-moment thoughts and memory, attentiveness and active sensory and motor control. We change the neural states and assemblies and we change the weightings of synaptic connections and this is going to change the process and pathways of electron flow which modulates the waveform of the system description live, on the fly. Modulating the waveform of our on-going process is a function of the phase relationships of the contributory data streams. It happens over time.

This complex waveshape of "immense dimension" modulating over time will produce a self-sustaining resonace. Our sense of continuity shows much depth in this resonance.

These resonant circuits will run in both localisable and global levels. Are the smaller ones, the localisable, directly a part of consciousness or more the (source of) contributory streams? Are they entities in themselves (Daemons)? Are they ideas? concepts? thoughts and emotions that keep coming back to us? Is there a system with oversight which I call self: "myself"? We are probably a complex of these things which it is the job of the psychiatrist, for example, to repair when damaged or disrupted or just grown too distorted to allow us to keep operating.

The "explanatory gap" invoked by David Chalmers in his formulation of the Hard Problem is the issue of how to explain the presence of subjectivity, i.e. the phenomenal, in an otherwise materialist, objective, representation which is physiology. Descartes originally created this problem when he split off the mind, the self, from the otherwise mechanical description of the body. Having made the separation (which I suspect was more of a political act than anything else) he then had to explain how the mind could control the body. His invoking the Pineal body as the means through which this control occurred was shown to be unsatisfactory and so the problem as formulated remained. We now see it in Chalmers' Hard Problem which is really a question of how do the two kinds of representation map onto each other.

As I have said above "The buzz of representations and productions in the brain are us and we represent them phenomenologically and investigate their physiology and their physics. We are inside this resonance, we live it, we are it."

There is no Hard Problem, consciousness is simply the ongoing process of being (inside) an immensely complex reverberatory circuit with a stripped down, simplified set of attentions and interpretations which we apply at any particular moment. Our active control over what we are doing from moment to moment is a physiological activity. We know what we want to do and we do what's necessary to achieve it by the self-regulation of what comes onto the "global workspace" the "stage" of consciousness [Baars]. I don't see this as some sort of competitive process but a co-operative process which we are inside and actively engaged in. Self-regulation is not competitive but constructive.

The only dualism necessary is the dualism of two classes of descriptive representation, the phenomenological and the physiological. They talk about the same process and we can see this mapping relation in the phase modulation of the electro-magnetic fields produced by nerve activities which are a direct affect of a change of phenomenological state, (e.g. a change of mind, a new decision).

The only hard problem we really have is getting our own heads around this idea that just as the complex interrelations between molecules and cells and their organised structuring and dynamics produces living things so the organised dynamics of the vast neuronal structure of our brains as in the world will also "go live" and this is being conscious.

References and Links

1. John Searle: presentation to Tucson II.

2a. William James. (1890) Principles of Psychology
2b. William James. (1892) Textbook of Psychology

3. John Locke. (1721) An Essay on Human Understanding

4. Paul Churchland. (1995) The Engine of Reason, the Seat of the Soul. A Philosophical Journey into the Brain. M.I.T. Press.

5. Robert Kirk (1996) The Basic Package in The Brain Project

6. David Chalmers (1996) The Hard Problem in The Brain Project

7. de la Mettrie L'Homme Machine

8. Hughlings Jackson. West Riding Asylum Reports 1876

9. James Newman (1997) Putting the Puzzle Together, Part I: Towards a General Theory of the Neural Correlates of Consciousness. in Journal of Consciousness Studies, 4 1, pp47-66.

10. Norbert Wiener (1948) Cybernetics, or Control and Communication in the Animal and the Machine, Wiley.

11.. Susan Greenfield (1996) Neural Assemblies in The Brain Project

return to Speculations index

What might a Conscious computing system be?

A speculative discussion by

Stephen Jones

Abstract:

We have, these days, quite a useful set of operational statements about what one might include as characteristics of consciousness. I want to consider how a number of these might carry over into some possible future conscious "computing" machine. Basically I am asking the question "What might a machine which had consciousness want to do for itself?" I then canvas several possibilities. Thes include self-maintenance, gathering knowledge of its environment and independence of action.

What might a Conscious computing system be?

1.1: What is Consciousness?

There has been quite a lot of work on this question over the years. So I shall briefly cover some of the ideas which have achieved a sort of consensus.

Consciousness is the "**I**" that we all know, from which we view the world and interact with it. **John Searle** describes it as that sense of "subjective qualitative states of awareness, sentience or feeling" which we experience when awake [Searle: presentation to Tucson II (1)]. In the 17thCentury **Rene Descartes** (2) showed that no matter to what extent our senses might be deceived there would still remain a something which could be called "**myself**" even if this "self" were utterly deceived as to the existence of any one or any thing else. I suppose his "cogito ergo sum" might have been better put if he'd said I am deceived therefore I am. **William James** (3), in the late 19thCentury, put it as having a sense of a personal consciousness that is ours, not something that we share [though in passing, how we consider this against the matter of the social construction of our world view remains to be discussed].

Everything we know is a function of **experience**, either through sense perception or reflection upon that experience. The mind or the "I" is born empty of knowledge of the world. As the 17thCentury English philosopher **John Locke** (4) described it we are born "tabula rasa" (or as a 'blank slate'). It is only by our experience of the world that we gain **ideas** of it. There are no ideas that we we are born with, no "innate ideas". Locke says:

> "Let us then suppose the Mind to be, as we say, white Paper, void of all characters, without any Ideas; How comes it to be furnished ?....To this I answer, in one word, from Experience... Our observation imployed either about External sensible objects; or about the Internal Operations of our Minds, perceived and reflected by our selves... These two are the Fountains of Knowledge from whence all the ideas we have, or can naturally have, do spring. [Locke. *A Treatise of Humane Knowledge*, 1721, p67]

It is the data of sensation which have the qualities or "qualia" which are the stuff of our subjectivity, that which we know.

Consciousness is a function of the state of our central nervous system, i.e. the **physiology** is the substrate upon which consciousness runs. To quote **William James** in his *Text Book of Psychology*, 1892:

> "The immediate condition of a state of consciousness is an activity of some sort in the cerebral hemispheres... One has only to consider how quickly consciousness may be abolished by a blow on the head...[or] by a full dose of alcohol...to see how at the mercy of bodily happenings our spirit is... Destruction of certain definite portions of the cerebral hemispheres involves losses of memory and of acquired motor faculty of quite determinate sort...Taking all such facts together, the simple and radical conception dawns upon the mind that mental action may be uniformly and absolutely a function of brain-action, varying as the latter varies, and being to the brain-action as effect to cause." [James, 1892, pp5-6]

Our state of consciousness:

> (a) is **always changing** as we are exposed to continually novel sensations.
> (b) is **selective** of what it pays attention to, and
> (c) provides a sense of **temporal continuity** (William James' "stream of consciousness"), which the normal day-to-day changes of sleep and wakefulness, as well as abnormal changes such as unconsciousness, do not interrupt.

Paul Churchland, author of *The Engine of Reason; The Seat of the Soul* (5), has described some important functional aspects of consciousness, which we would need to explain in a theory of consciousness.

> (a) Short-term **memory** and its decay.
> (b) Directable **attention**, or conscious control over what we attend to and what we do.
> (c) Multi-valent **comprehension** through "mulling" or reflection.

(d) **Independence from sensory input** in say, daydreaming.
(e) The **disappearance** of consciousness during sleep.
(f) **Unity** across the senses and unity over time.

Robert Kirk (6), of Nottingham University, has developed a description of what he calls the "**basic package**" of capacities that an organism must have for it to be conscious. This is primarily about the organism being able to collect information which is specifically intended for its own use, and then about its capacity to decide what to do with that information. There is a package of related activities which apply to this collecting information.

(a) The organism must be able to use the information.
(b) The organism must be able to initiate and control its activity on the basis of the information it collects.
(c) The organism must be able to assess and interpret the information.
(d) The organism must be able to assess the situation it is in so that it can decide how to respond or whether it should respond.

Our ability to respond to **novelty** in active and constructive ways is considered to be somehow over and above the mere 'irritability' of the senses and the body's reflexiveness to sensations which may or may not then have conscious impact. **Independence of action**, the ability to decide to do something, is crucial. The plant which follows the sun across the daily sky is not able to refuse to do that. It has no opportunity to decide whether or not it will not respond to the attractor of the sun. It can only do it, there is no capacity to act independently of its built in responses.

Further, we should be able to report on our activities and our internal states. We can talk about what we experience and how we feel. We can also use the contents of our experience to generate new and imaginative reports and to be creative with what we know.

So ultimately, if one intends to develop a theory of consciousness the problem is to explain how it is that the physical system that we are, this bundle of cells and organs and nerves, can have subjectivity, can behave independently, can do things for ourselves and can respond to input in a way that takes account of whim and ideology rather than simply reacting to current conditions. This is the question that **David Chalmers** (7) asks in his formulation of the "**Hard Problem**":

> "Why is it that physical processing in the brain, no matter how sophisticated, should give rise to any subjective inner life at all? Why couldn't that have all gone on in the dark? That's the real mystery."
> [Chalmers: Brain Project interview]

So to sum up our criteria. A conscious entity would need to:

- **Act independenly, initiate procedures**
- **Do things, such as gathering information for itself**
- **Do things which it initiates voluntarily for others**
- **Receive, store and reflect on perceptual input**
- **Show generativity, creativity and the capacity to construct,**
- **Report upon the contents of our Cs**
- **Interact with others in general**

1.2: To reiterate: What is it about us that indicates consciousness?

What would we (conscious beings) expect to see in another entity for us to regard it as intelligent or conscious? Noting that there are *possibly* systems which are intelligent in much the way we think of ourselves as intelligent, but not actually conscious of themselves or their surrounds ("the famous philosophers zombies").

Firstly we must be able to know, report upon and interact with the world. We need a system of inter-connections with the world which allows us to gather information about the world and by which we report back to the world or produce "models" which we put into the world. (The activities of both science and art are very strongly involved in the production of "models" within the world.) As humans we need working sensory perceptual systems to provide us with **difference** which is what we see by and know by. We use our nervous systems and brains to do this.

So:

> **(a) we need to be able to know the world and to act independently within it and to do things for oneself.**
> **(b) we need to be able to sense, remember and reflect within on the patterns within,**
> **(c) we need to be able to reflect on the world by interacting with others, and**
> **(d) we need to be able to project into the world for puposes of generating feedback.**

Now, the human body/brain is a complex organised system consisting in collections of cells which are organised in such a way as to do various biological/biochemical tasks. Among these tasks are the processes of the brain in co-ordinating and handling all this stuff. The effect of being in the world in real-time (and strictly **in** real-time) is that the entity is consistently exposed to difference and novelty. We know of the the world by the differences that show up between and among all processes extended in time (having duration). No process can act completely independently and each is affected by the changes (the differences) in, ultimately, all other processes. So we are inextricably interlinked with the world.

Because we are active creatures the environment is always changing about us as we move about in it. So every thing produces difference, even if only the two different views of some object which allow us to build up a 3-dimensional view of it.

2.1: The determination of intelligence in other systems

What of intelligent machines? Given that an artificial intelligence will also need to communicate with its world, which hopefully includes us, it to will need to have an array of inputs and sensors and communication systems at its disposal. An AI will presumably have some sort of "brain" to co-ordinate all its activities, as likely as not this will be at least as highly distributed a processing system as we use.

What is it that enables us to determine whether an other entity is intelligent or not? Largely it's a matter of whether or not that other entity can report to us that it is so, or if a common language is unavailable can make other output which demonstrates use of reflectively produced information. I recognise that this allows many current computers a foot in the door, but the capacity to extemporise on the intricacies of current political drama might show us a capacity beyond what is currently found in computing devices.

The traditional procedure for determining whether some artificial/constructed device can be described as "intelligent" or not has been the Turing Test (TT) (8). In the TT an interrogator is set in front of a terminal which presents information from the entity being tested without indicating the nature of that entities' direct input/output capabilities. In other words the information is mediated in such a way as to remove any tell tale clues (say, typing speed). The interrogator then asks the entity questions and if the interrogator is unable to tell which kind of entity (human or machine) is supplying the answers then that entity would have to be considered "intelligent". Of course it could be playing down to us.

The Turing test requires the "computer" being tested to behave anthropomorphically in that we require it to report to us in a manner which we can understand. We want it to display human-like behaviour. So I am going to carry out an **anthropomorphic analysis** of what an intelligent computer system might be.

By the way, there is some difference which should be noted between "intelligence" and "consciousness". Consciousness seems to have more general applicability than intelligence in biological entities. We tend to think these days that most creatures above the levels of fairly rudimentary mammals have some sort of "consciousness" of their surrounds, but intelligence is a term which may or not be applied to an acceptably conscious individual. Its use will largely be dependent on context. I will come back to this point in discussing the various possibilities of classes of AI's.

2.2: What are the signs by which we might recognise consciousness in biological or silicon others?

As **Robert Kirk** has pointed out the most obvious thing about a conscious entity is that it does something for itself and very largely that something is to collect data about the changes of and the differences within its environment.

Another thing which a concscious system would need to be able to do is to act independently in a more general way than simply gathering information. That is to decide to do (or not to do) some thing for itself, of its own accord, and to proceed to do that without being prompted by some external conscious agency (an 'other').

We maintain an 'idea' of the world about us by gathering information (no matter how low level) which is for us, for our own use. We gather information for ourselves.

Thus knowing will be an inherent process (complex and organised as it is) of any conscious system for gathering any information that may be useful in some way (in some sense). For example: What is the surface like? Where is the next food? Where is my friend?

So if this is more or less what consciousness is for us, what will we need to see in a non-biological entity, a silicon-based computer system for example, for us to say that it is conscious? and how might that be different from saying that it is intelligent?

We should apply the same criteria. This is the **anthropomorphic procedure** I suggested would be necessary to answer the question.

2.3: So where are we likely to find this system if we're to ask it if it is conscious?

We can envisage two classes of intelligent or Cs machines.

The first are (at least, logically) **self-contained machines** which in some way are able to act like an individual human. They would be able to sense and make sense and act on the sense and display, or report on, the act and its results. They would have to be some sort of highly integrated machine which had enough internal "physiology" to enable it to handle the vast amounts of data that are necessary for a coherent view and understanding of the world. Perhaps a large neural network machine with multitudes of distributed nets carrying out all the tasks necessary to keep such a device in touch with its context (the world).

An example of such a machine might be **HAL** in **Arthur C Clarke's "2001"** (9). We can possibly produce a machine like HAL in which we specify the kinds of tasks we want it to carry out. But what happens when it decides to do something for itself? HAL's problem was that it had two contradictory prime directives and went psychotic when these were forced into contradiction as they were at a certain point. The average modern supercomputer is an "idiot savante" when it comes to doing anything other than its primary task: calculating. Our self-contained machine would need to have a vastly distributed set of processing subsystems which sense (feed forward), reflect and control (feed back) each other, keeping it in touch with itself and its world.

Because of things like the now acknowledged impossibility of ever describing the world and consciousness in a systematically complete formal or logical system (*vide* Goedel's Incompleteness theorem (10)) and because of the huge programming task, it's probable that we would not be able to pre-program this system and simply set it runnng. We would have to set a machine up to run itself and then teach it just like we do our children. Though it may have vast access to libraries of knowledge the machine still has to be able to operate in the world: to know who and what its neighbours are, how to communicate with them in which particular language or protocol, and so on. And further, how would we actually manufacture such a system? the complexity of wiring and the range of sub-system capability are staggering if not insurmountable. It seems that we would have to use evolutionary algorithms and some kind of auto-assembly process to even begin to build such a machine. I argue that the imbuing of consciousness to this machine will follow fairly straightforward principles but that the actual technical implementation of such a machine will require a total reworking of manufacturing techniques into something more akin to the biological process.

So, do we ultimately have to let the system grow itself and decide what it wants to do both in relation to us (in responding to our input and requests) and in relation to itself?

The second class of intelligent or conscious machines is a **distributed network of machines** which probably would act more as a society, but might also be able to act as a single combined entity, an "individual". This would be a large system in which the elements are widely distributed, that is not all in the same room, say. Generally it is known as a net or a network and would need enough layers of linked subsystems, in some sort of hierarchical as well as horizontal structure, with enough of an organised basis as to be able to distribute the array of tasks necessary for conscious behaviour over an array of appropriately inter-linked subsystems of computer embodied sub-nets.

My feeling is that the **Internet** will be the context in which this kind of capability first appears. Some sort of capacity for self-organisation would have to be installed on the network to enable it to evolve to a point where it might exhibit an awareness of its (logical) surroundings, and to be able to report at least on this situation, if not on its internal goings on.

2.4: What are the differences between a single intelligent machine and a social network of machines (nodes) as in the Internet?

An intelligent or conscious machine is in a sense an analog of what you or I are, an individual. But I mean that more in the sense of having some sort of self-perceived bodily integrity. Will the machine be mobile? It might seem necessary if it were to gain the kind of experience that we consider the basis of our consciousness, but I am hesitant about saying that mobility is necessary. An intelligent internet, of course, couldn't be mobile. It would gain the knowledge of the world that we have as individuals in much the way our society conveys ideas and knowledge that we might not get if we had to have first-hand experience to gain knowldege. That is through teaching and learning and through its equivalent of the movies and the news. Also an intelligent Internet would be everywhere at once; an omniscient being, whereas an individual intelligent machine would be restricted to the kind of localised mobility and knowledge sources that we use.

(Though the machine, one supposes, would be able to log in to whatever library system it needed for more general information about the world and might emulate the sense of being everywhere at once.)

One might liken the difference between these two classes of machines as the difference between an individual and the society within which an individual lives.

3.1: What might a conscious machine want to do for itself?

There are two questions invovled here: First, what do we do for ourselves? and subsequently: What of these things might a machine need to do or want to do? But I will proceed by describing how what we do translates into what a machine might do.

The collection of functions of consciousness I mentioned above can be summed up by saying that a conscious entity will need to act independently, to do things for itself and to be able to report on things and interact with others, i.e. to communicate. We are not here concerned with the particular facilities for doing such things held by any particular entity, we are really only interested in what kinds of things might be done.

Getting down to the details:

3.2: Self-Maintenance:

The first thing that any self-consistent system must do is look after its own **maintenance**. Even if it has to get the "doctor" in it will need to able to report at least to itself as well as to another (in whatever language is appropriate) the failure of some piece of "physiology" (hardware) or the loss of a link. Also this includes the ingestion or acquisition of energy sources and, if it's going to emulate the biological by being self-maintaining and self-replicating, component parts.

For example, the Internet is a vast array of ("intelligent") terminals which occasionally hook into a slightly less vast array of servers which store files for access by terminals, these files being placed there for human use by and large. A potentially intelligent computer network of this collection of smaller networks would have to look after its own maintenance. Each "terminal" would need to maintain a diagnostic (inner) "eye" on its overall operation and especially it would need to build techniques to prevent its penetration by an unwelcome agent (e.g. virus, hacker). Is the anti-virus programme the first step toward an immune system? The nodal computer (the "server") would need to maintain itself and also keep an "eye" on its links, these being the terminal links to it (the user names) and the links up to the next server in the next layer up in the hierarchy of nets. Which next server would also have to maintain itself and its links.

Each individual machine would also need to look after itself both in terms of internal maintenance and materials (information and fuel). It could refer up the net for system diagrams and diagnostics but it would need to carry these out itself or request help from another (perhaps a "user"). A basic level of this self-maintainance is already available. For example, individual computers, the nodes of the net, can use what are known as **boundary scan** techniques (for which there is already a formal standard) to maintain themselves. Meanwhile their links are already maintained under the system of using pathway redundancy. Through the need to be auto-sustaining in case of a link breakdown the internet has a flexibility of interconnect pathways which reduce all pathways to being pretty much equivalent.

3.3: Organisation:

The machine or system is going to have to be "**organised**" with a structured interconnect system and groups of subsystems which carry out particular functions, such as vision. Biological systems are "intelligent" because the collections of cells we call '**organs**' are organised into particular classes of function so that, for example, there is a group of cells which act to bind oxygen into special molecular containers to carry it through a supply network and release that oxygen to other cells under certain particular conditions of request. A blood cell is an organ just as a brain or a liver or a heart is. It is these functional groupings of cells (elements) which feed and maintain themselves and serve each other to be part of an essentially co-operative system that forms the body or the **organism**. By **system** I mean just this, a collection of functional units or terms which function together (mostly interactively) to form an entity greater than the sum of parts. It is the **relations** between elements of this system which keep it together.

What are the equivalents to specialised organs in a computing system? Or are there any such? In an individual computer we have the CPU, an arithmetic unit and Memory and the internal and external interfacing logic. It strikes me that these are more like individual neurons or the machinery of an individual cell. Not enough to be considered organs, more the possible constituent parts of an organ. I suggest the nearest example which might be similar to an organ that we can find is a set of computers (particularly using a parallel processing architecture) linked together for a particular purpose, say mathematical or theoretical physics problem solving. Or we might think of the telephone network as being analogous to a nervous system.

Going on up into the network scale, are there groups of computers with oversight of the net which carry out maintenance roles? And, if those computers are doing something which is essential to their continued functioning are they doing something for themselves? Is there yet a sense in which maintaining the net is doing something for themselves (or *itself* as the 'group-self') and if so do they recognise this fact?

All the things that a computer does at present are things which it is requested by a user to do, whether it is acting as typewriter or doing calculations for the most abstruse scientific visualisation task it is still only doing things for us. Maintenance and the formation of functional units are very low level activities, but for an intelligent or conscious system very necessary. We still do not have active systems which can do this kind of work although many parts of the task are extant. In the maintenance mode, keeping track of the state of individual physical links is still the nearest thing to "doing something for itself" that we have.

Then, what else?

3.4: Knowledge of its surrounds:

The next thing I think of is **knowing its environment**: searching for information which is particularly for its own use, such as who its neighbours are and what the overall structure of the society of computers might be. This information may be reportable to humans but it needs to be primarily kept as part of the maps of its culture.

Current **search engines** are among the few semi-autonomous entities on the net/web. Their activity is to go from node to node gathering information about files kept on the nodes. They maintain a data-base (a body of knowledge) about links between nodes but this is only as information about what is kept at another node (millions of other nodes). So a search engine is an organ in one sense but we cannot call it intelligent because it doesn't do this for itself it does it for its users which are, ultimately, all the humans at all the terminals at the ends of all the links in the network.

The production of new kinds of web-bots and information mining and sorting agents increases apace. But these agents are still only doing something for an external 'user'. It is when a computer in the network generates some code for its own use to go and find information about, say, a maintenance problem (e.g. a batch of chips used in its main processing system is giving trouble and it needs to know what to do in a software work around) that things get interesting. To put this example in cultural terms: it needs to go and consult the 'doctor' subsystem in the network.

3.5: Independent action:

This is where the maintenance mode produces the first signs of another indicator of consciousness in **independence of action**. It should initiate the task of finding the necessary data and modifying the code to its own ends. Now, we can say that "doing something for oneself" is an independent action but there is a complimentary form of independent action in that a system might decide of itself to do something for someone else, which could well be in the area of generating or searching a particular category of information which that other is known to, or 'thought' to, use for their own needs. Something like finding a book that you know a colleague is interested in. Then, of course, there's the purely altruistic independent action of buying a present for a friend.

A system might also want to work independently on a problem set by its 'boss' so that it can generate useful answers and receive positive feedback about its own usefulness. This is almost a level of emotional need.

So what are the machine equivalents of independent action, **productive or generative action** and even **altruistic action**? It is the latter two that brings independence of action out of the realm of simple maintenance procedures and where we see the first kind of behaviour which could be considered conscious.

Two things that a system will want to do for itself, if what we humans do can be taken as any kind of an indication, will be:

- **1. organise itself and its relations (with other machines as well as humans), and maintain that organisation, and name its social system (These things are the means by which we are in the world)**
- **2. enquire into some of the "imponderables" of its existence, such as how it came to be? and what are the humans? (are they gods?) (These things are the means by which we make art and other cultural productions).**

For humans these are two of the deepest givens we have

> 1. the social system is handed (down) to us by our parents first and then they tell us of the ancestors who "wrote" the rules about uncles and aunts, exogamy and taboos, parents and grandparents.
>
> 2. the imponderables are revealed to us by some "hypothesis" (or construct) we usually call the Ancients or God, but which are the long lost anticedents of our respective cultures.

We could describe these two processes as firstly, elucidating the structure of relations between all entities on a net, and secondly, the modulation of these sets of relations. (I take it as given that "sets of relations" include relations with physical and mental productions in the world as well as between entities in the world.) The modulation of relations is the business of art and science, politics and religion but the effect of each folds back into the structure of relations giving us a dynamic informational/social environment, which in its turn modulates us, right down to the synaptic connectionism between nerves.

The structures of "social relations" is really just a thorough knowledge of the links in the system and the "mental" contents of the entities at the nodes of the network. The modulation of those links is much more interesting and here is where we begin to be able to truly recognise a similarity to consciousness in humans.

Conscious machines are going to have to be able to deal with all these sorts of things: the needs and desires of their (machine) neighbours, the co-operative construction of real or virtual/logical aspects of their environs. They are going to have to be able to know that certain sets of relations are inextricably bound, like the objects which are assembled into a car (given that the computer might only have a parts list) or like the social relation of marriage. They're also going to have to be able to bind things occurring concurrently on different stages in the world, to recognise that these things are happening *now.* For a single distributed machine like the internet all of this stuff will go on in its "body" and, of course, as we don't usually know in detail the goings on within our bodies neither would the "conscious" level of such a machine need to know. If everything about its internal world were only as conscious as ours might be, then, if this machine were truly intelligent, it could be expected to get lonely. That is, it could be expected to desire and seek other entities to converse and interact with. These are emotional and social needs which bring the system into a condition which is more and more identifiably anthropomorphic.

4: The Social Dimension

Beyond the terminals of the network are the users (currently human). What is their role in the intelligent network? Is the intelligent network somehow like a culture, providing informational nourishment to its end users? If the network is intelligent it will ask of its end-users informations that it needs for its own continuance.

Again, what are the kinds of things that a system would want to do for itself? The prime reference model is of course us. What do we do for ourselves? We maintain ourselves and our contexts and we generate stuff for ourselves and others around us. We

- maintain the nodes and the links [i.e talk to our friends, relations and associates]
- feed into the links and their extensions [both ideas and cultural production]
- search for and receive information from the links and extensions
- become emotionally involved with these extensions

Perhaps to allow a system to be intelligent is to allow it to decide what it considers to be useful for itself.

A conscious internet would be a nodal system for which we can suggest "culture" as the model and so we can imagine a variety of posssible social structures which we might expect to find in this network system. For example, individual nodes in the net might want to establish one of their members as a kind of oversight node looking after the link structure of all the other nodes in the local area. They might hive off their external maintainance to one individual (the minister for health). Some nodes might be established as repositories of information, like librarians or teachers. Other nodes might deal directly with the humans, the users at terminals. These might be the foreign affairs department, (forgive the bureaucratic model here).

In a sense I'm talking about a social entity, an array of individuals who live as a subculture in society. The networks of individual computers will as likely as not form into political groupings or family groupings and we will see hierarchies of relationships develop. Will they be able to evolve better political models than we have so far? Will co-operative and compassionate structures evolve with their framework?

Now, societies do do things for themselves, so are societies or cultures intelligent or conscious? I think we might have to say that the well known tendency for an institiution to look after itself against all comers implies a kind of consciousness. (Just look at the police service or the church.) But it is embodied in the entities which are the nodes in that social network who, in the process of their doing "some things" for themselves, generate the process of maintainiing the institution. The traditions of the institution form the skein of ideas and information which the network (i.e. the sets of relations between....) uses to inform and shape the embodiement of the ideas that make up that institution.

These ideas are transmitted from node to node using a medium of **language** of some sort. Some people have suggested that language is a virus. It certainly shows many of the epidemiological characteristics of diseases caused by bacterial or viral agents in terms of modes of transmission and patterns of dispersion. Ideas are carried via the vector of language, but they are also moulded by their vector. Some languages can only carry certain limited aspects of the full possible spectrum of an idea. I suggest that this means that "memes", Dawkins' (11) and Dennett's (12) units of meaning, carried on the vector of words are highly dependent on that vector, or language. That is they are **context dependent**. They are moulded by the language vehicle being used and they then mould the interpretive mechanisms of the individuals exposed to these languages, which process then, in turn, moulds the possible interpretations of the memes used in the first place. Folding in on the language in an endless generative process in which the potential meanings can grow and the endless variety of languages and cultures appears.

The same would happen in a conscious computer network, though it might be capable of knowing and using a vastly greater variety of languages, which would then provide a greater spectrum of possibilities attached to any word one might use.

In a social system the vectors are spread by direct contact and interaction as well as carried in cultural productions, be they books, films, art or whatever. The great value of art in the world is its capacity to be used as a vector for new, undeveloped and perhaps even socially undiscovered ideas. Of course the other great value of art is its capacity to trigger generative interpretive processes, actually bringing new ideas to the surface of a culture.

In human societies ideas spread, traditions and cultures come and go, but they are diverse and dynamic. Will a society of computers retain this diversity and its inherent dynamism? Or will it settle to a lowest common denominator condition wherein it knows all it needs to know and the low level stuff is running well and so on. Will a conscious computer system need to be jolted in to activity by some threat or problem? Will we have to keep it entertained? Will we as generative agents be useful to the machine or would it prefer to be a slumbering giant. Or is the system, if conscious, already generative. And are any actually conscious computer systems potentially the most generative devices known? There is already some evidence to suggest that this is the case. I refer particularly to Steven Thaler's Imagination Engines (13) in which a large neural network is trained up on a particular set of input forms, for example motor car designs, and then turned loose, without any input patterns, using only noise as an input, to generate any patterns which might come up, with some sort of "selection" mechanism picking out those results deemed useful according to some "socially acceptable (?)" criteria.

But what else do social systems do? It is generally considered, I suppose, that we generate these social systems. We produce them so that we can maintain ourselves and gather useful information for that purpose. It's obviously not the only reason we do things for ourselves. We act creatively in myriads of ways. In every sentence we speak (at least when not acting someone else's sentences, and even then) as well as in the acts of making art or music, cooking or writing or generating computer software. All this is generative activity, requires a generative capability, and is an important part of being conscious or intelligent. We often generate stuff for other people of course. Most art and produced culture is done for others to see and interact with, the payoff for us (the producer) being in the feedback.

> **Thus, I argue, we would expect an intelligent or conscious computer system to be similarly *generative*, and it is this which in the long run will give us the most evidence for suggesting that the machine or system is in fact conscious.**

5: How would we produce this conscious net?

How would we start up a process in which something like an intelligent and conscious machine or system would develop?

Francis Heylighen (14) of the Free University of Brussels, Belgium, has suggested a model for searching the WorldWideWeb using the analogy of associative memory where hypertext documents are considered to be "concepts" in memory and hypertext links are like associations of ideas. As links are followed by the user they accumulate "weightings" which relate to the frequency of usage. (This is a kind of Hebbian model of "learning" in which connections between neurons are made easier by usage. For example, one learns to associate concepts with words by a process in which those associations that gain the greatest connection strength or "weighting" become the best "interpretation" of the input data pattern.) Thus links between concepts which are regularly used will become stronger, and links which are used slightly less show up the rarer connotations of a word or concept and so on out in a kind of "spreading activation" of links which, Heylighen suggests, emulates the way we associate concepts and ideas in our brains. If we could set up a search engine system to automatically record the link usage as a kind of "association strength" meter, then we would see the most direct connections between words and ideas having the greatest strength and the less well known connections coming up as less strong. A system could then make suggestions as to possible connotations which it "thought" a user might be looking for if they asked only a vague question about something or asked for, say, poetic connections.

Again this is a primarily human oriented use of the system, though the capability when applied to things computers might want to know about, say the availability of particular pieces of software code, is easily generalisable. How could this be implemented?

The seed process would need to be the installation of some kind of means for a computer to write code for itself. The system would then need to decide that it should contact some other machine that can help with a particular task and go out and find it, by searching for particular bits of code which we might liken to names or symbols, the equivalent of hanging out a sign from the shop window. Once a machine is able to act independently to offer services and make enquiries then we start to have useful networks of entities. We then need to get the system to recombine its new knowledge with old knowledge and draw new conclusions, develop new combinations and so on in the kinds of activity we know as being creative.

(Of course we've just come across another thing which a conscious machine would want to do for itself and that is to offer services and advertise those offerings.)

So we need generative systems embodied in the Internet and if possible in individual machines. Is this more than just a trick of programming? I suspect that particular kinds of hardware development will need to take place, presumably in the area of neural net systems. That is extended neural nets made up of multiple sub-nets all contributing to the data flow. A goodly layer of feeback would be needed from network hubs that run the major functions of the system to control and render manageable the vast streams of data that the Internet would carry. Routing systems and routing control in an endless loop of being aware of itself, of knowing what is going on in and around it. This feedback regulated structure will have a period, an interval over which it occurs, a cyclical process, different segments of which would tend to operate asynchronously from each other. But how then does the system bind together all the various patterns of data it has running within it so that it knows that certain things like the lips moving and the words being heard are related, perhaps interleaved. One way might be to run a clock which drives the process of each sub-net so that they all run more or less synchronously. The great advantage of the cyclical process binding or synching-up the various sensory modalities is that the duration of the process and its recurrent regulation provide a kind of short-term memory of the contents of its "mind".

On the social system scale, i.e. on the "Internet": to produce a "live" system would be to write some code which could allow the set of nodes we have available to behave as some kind of neural net. This is firstly parallel distributed processing (a horizontal structure in which various tasks are broken up among "specialist" processors) but it is also a collection of hierarchical structures in which one part of the stimulus input is combined in a weighted sum with various other parts of the input and the meaningfulness of that input stimulus is decided by the error value or the difference between what the net produces on this cycle versus what the expected or desired result would be. The prevailing culture provides models by which we compare our production with the "standard" and decide whether we have succeeded in either emulating the "standard" or in denying it as much as posssible.

The issue then becomes who decides what the expected or desired result is to be. The expected results will be partly a function of the culture surrounding the machine and the inherited culture of its makers. Now in us (in evolutionary process) this has been a process of trying stuff out and deciding (or discovering, much to our horror) how useful or not the result is for whatever we may be doing for ourselves. So **we provide our own criteria for usefulness. A conscious computer system would ultimately have to be accorded the same right to apply its own criteria**. These criteria are of course derived like ours in that they are developed over time within the culture of the period based on the ideas and traditions handed down from previous versions of the culture.

5.1: An intelligent neural network has to be able to set its own goals and figures of correctness.

How does it do this? Incoming data will have characteristics which deternine which "sensory pathways" are used: word data will go to a speech recognition system, 2- and 3-dimensional patterns of light will go through to its vision system for feature analysis. The data of "now" will be compared with previous data and correspondences and differences noted, changing the weightings of the patterns of interconnection by which we learn new data and recognise familiar data. All the nodes and the network of links are potentially neural nets. Let's feed the nets with stimulus and the occasional blast of noise (a la Steven Thaler's Imagination Engines (13)) and see what the net might decide to do for itself. This is very similar to dreaming: shut off the external stimulus and let the net sort through all its internal stuff and try out possible links between nodes and linkages of data in a kind of "search-by-similarity" engine. [*vide* Alan Hobson (15), Rodolfo Llinas (16)]

Any system which is capable of conscious behaviour must be able to **reflect** on the contents of its consciousness and on the contents of its surrounding culture. It is through this that we develop new ideas, new techniques and new cultural productions. Reflection is the process of storing input information and feeding it back through the input processing susbsystems, adding new input information to the body of the "idea" and allowing the body of the "idea" to modulate any interpretation placed upon the incoming information. During infancy, as we deal with and "reflect" (albeit at a very low "hardware" level) upon this input we learn, for example, to recognise our parents and others around us. Also we learn our language and its meaning, through just this repitition of processing and "trying" to understand, and layering of previous examples and understandings onto new ones.

5.2: Another activity that a conscious computer network will engage in is the search for the means of doing or achieving something.

A reflexive network system will find software on particular nodes in the system that allow it to do all sorts of things, especially at our request. Intelligent agents that find the nearest available node to carry out some particular process that we need to do probably almost/already exist. Of course it can also find that software for itself, but still what drives it to want to do whatever that software node does? to use that operation, that software?

And does it look at the system in terms of software nodes rather than hardware nodes? I mean, we might look on the kinds of things we do as an array of soft functions relating to input terminals: eyes, ears; and output terminals: mouth, muscles, etc. If the map of the system is not dependent on hardware nodes then machines can become specialists in particular activities available to any (human or machine) user which needs that processing. That is, give the software an address that it can recognise and when it is needed the system simply posts the data with the appropriate software address and the package is sent out to the network and captured by the addressed software which then carries out the requested processing and returns the package to its source address. The software simply needs to know if it is being addressed and whether it is busy or not. Is this the end of selling software? One simply buys the machine-cycles to do whatever tasks are needed at your own node. Is this the (dumb) Network Computer that is currently in development?

6: In conclusion

I suspect that the first context in which we see truly a conscious machine will be in the Internet/WorldWideWeb nexus, but that this will entail a development by stages or levels of hierarchy. I suggest that by the introduction into the Internet of techniques and algorithms for letting the netwok itself develop and extend its connectionism, that is, its hypertext linkages, especially search-by-similarity techniques, and by providing means for closer co-operation with its human users in search processes, using various kinds of "agents" and "web-bots" we might start to actually get a kind of bond between humans and machines that will be radically new. This process will produce a kind of group-mind or "Global Brain" (Heylighen (14)). Then the extension of this towards an actually conscious system will require the introduction, into the net, of the kinds of capacities which I have discused above.

Obviously this leaves many issues unresolved. For example: the ethical question: if we do produce a conscious computing system but we don't allow this system to develop its own freedoms and individualities then are we in fact simply breeding slaves? (*vide* "Do Androids Dream of Electric Sheep?" by Philip K. Dick (17)). Also we probably will not be able to do this kind of thing in an individual machine with the classes of technology that we use currently. It is most likely that we will need to be implementing computing machines in biological hardware before they can be made complex enough to grow and learn in much the same way that we do as human children. And of course this leads us to a final question: Are we simply trying to find another way of taking over the natural process and re-inventing ourselves as some kind of extended human or are we simply making more humans?

References:

1: John Searle: "How to Study Consciousness Scientifically" a paper delivered at *Towards a Science of Consciousness 1996*. see also "Is the Brain's Mind a Computer Program?" *Scientific American, 262*, 1990, pp26-31.

2: Rene Descartes: (i) A Discourse on Method. (ii) The Treatise on Man. (iii) The Dioptrics. (iv) The Passions of the Soul. [see also extracts from Descartes in the Brain Project web site]

3: Wiliam James: *Text Book of Psychology*, 1892

4: John Locke: *A Treatise of Humane Knowledge*, 1721

5: Paul Churchland: (i) "Consciousness: its Past and Future" a paper delivered at *Towards a Science of Consciousness 1996*. (ii) *The Engine of Reason; The Seat of the Soul*, 1995, MIT Press. [See also the web page Neural Networks and the Computational Brain which includes a precis of Churchland's talk at Towards a Science of Consciousness 1996.]

6: Robert Kirk: "The Basic Package and Consciousness" a paper delivered at *Towards a Science of Consciousness 1996*. [see also Robert Kirk explains the Basic Package a transcript of an interview recorded at Towards a Science of Consciousness 1996.]

7: David Chalmers: (i) "On the search for the Neural Correlates of Consciousness" a paper delivered at *Towards a Science of Consciousness 1996*. (ii) *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press, *1996.* [see also David Chalmers on the Hard Problem a transcript of an interview recorded at Towards a Science of Consciousness 1996.]

8: Alan Turing: "On Computable Numbers, with an Application to the Entscheidungsproblem" in Davis, Martin. *The Undecidable: Basic Papers on Undecidable Propositions, Unsovable Problems and Computable Functions,* Raven Press, New York, 1965*.*

9: HAL, see Simson Garfinkel's article in the January 1997 issue of *Wired* celebrating HAL's fictional birthday. See also Arthur C. Clarke, *2001: A Space Odyssey.*

10: Godel, Kurt. "On Undecidable Propositions of Formal Mathematical Systems" in Davis, Martin. *The Undecidable: Basic Papers on Undecidable Propositions, Unsovable Problems and Computable Functions,* Raven Press, New York, 1965*.*

11: Richard Dawkins; (i) *The Blind Watchmaker*. (ii) *The Selfish Gene*.

12: Daniel Dennett: (i) *Consciousness Explained*, Little Brown & Co., 1991. (ii) *Kinds of Minds*. (iii) *Darwin's Dangerous Idea*.

13: Steven Thaler: "The Fragmentation of the Universe and the Devolution of Consciousness"

14: Francis Heylighen: From World-Wide Web to Super-Brain" on the *Principia Cybernetica* website.

15: Alan Hobson: The Chemistry of Conscious States, 1994. Little Brown.

16: Rodolfo Llinas:

17: Philip K. Dick: *Do Androids Dream of Electric Sheep?*

return to Speculations index

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

Consciousness Reframed '97

A note on a possible physiology of subjectivity, and some comments on what a conscious machine might want to do for itself. *

by Stephen Jones**

This paper covers three areas: first, a consensus view of what consciousness involves, then a description of an emerging consensus about where in the brain consciousness arises and some speculation as to how subjectivity appears, and finally a look at what a conscious artificial intelligence might want to do for itself.

1. What is Consciousness?

Most people engaged in the study of consciousness are trying to find some sort of explanation for, or at least an understanding of, how subjectivity arises. How it is that this physical entity that we are can also know the world and feel it, and be something within the world?

To do this we need to understand perception and memory, language and knowing. We need to recognise and understand subjectivity and identity, our spatial and our temporal unity. We need to understand how experience and our reflection upon that experience produces ideas and our generative activity.

But before we do any of that we need to know how to determine whether an entity is conscious?

Robert Kirk, of Nottingham University, has developed what he calls a "basic package" of related capacities that an organism must have for it to be conscious (1). He suggests that in order to identify whether an entity is conscious: it must be able to collect information which is specifically intended for its own use, it must be able to initiate and control its activity on the basis of that information, it must be able to assess and interpret the information, and then decide what to do with that information.

These give us a basic set of criteria for consciousness, but there are several more things which I would like to add.

Firstly, our ability to respond to novelty in active and constructive ways is somehow over and above mere 'irritability' of the senses which may or may not then have conscious impact.

Secondly, independent action, the ability to decide to do something, is crucial. The plant which follows the sun across the daily sky is unable to refuse to do that. It can only do it, there is no capacity to act independently of its built in responses.

To sum up our criteria. A conscious entity would need to:

- **Act independently and initiate procedures.**
- **Do things like gathering information for itself.**
- **Do things voluntarily for others.**
- **Receive and reflect on perceptual input.**
- **Show generativity and the capacity to construct.**
- **Report upon the contents of its Consciousness, and**
- **Interact with others.**

So why is subjectivity such a mystery? As yet there is no really adequate story about how our physiology can produce anything like the feels and perceptions that we know subjectively. The phenomenological stuff that we know does not seem to relate to the physical material world. It is produced in us by that world but it doesn't appear to have physical existence. But, I suggest, this is largely a cultural artefact.

The problem is this: we are, indubitably, physical beings, but we carry this historically received idea that the self is somehow independent of the physical body. The anatomical discoveries of the 16th century reduced the body to physical materials. As this would have led to the denial of the "soul", the philosopher Descartes decided that, as a devout Christian, he would have to find an argument which guaranteed the idea of the soul. Using his method of radical doubting, he found that although he might be deceived about the true existence of anything material, there was still something which was being deceived. This was his "self", his "thinking thing", and thus he created "dualism" (2) the idea that consciousness is something non-physical, an immaterial stuff.

If it is, then we have several problems. The first is just how does this immaterial stuff interact with the material world? Also, how does the material world of our senses, as mediated through the brain, actually make any impression on the mind stuff, so that we can have memories and perceptions and the other trappings of consciousness?

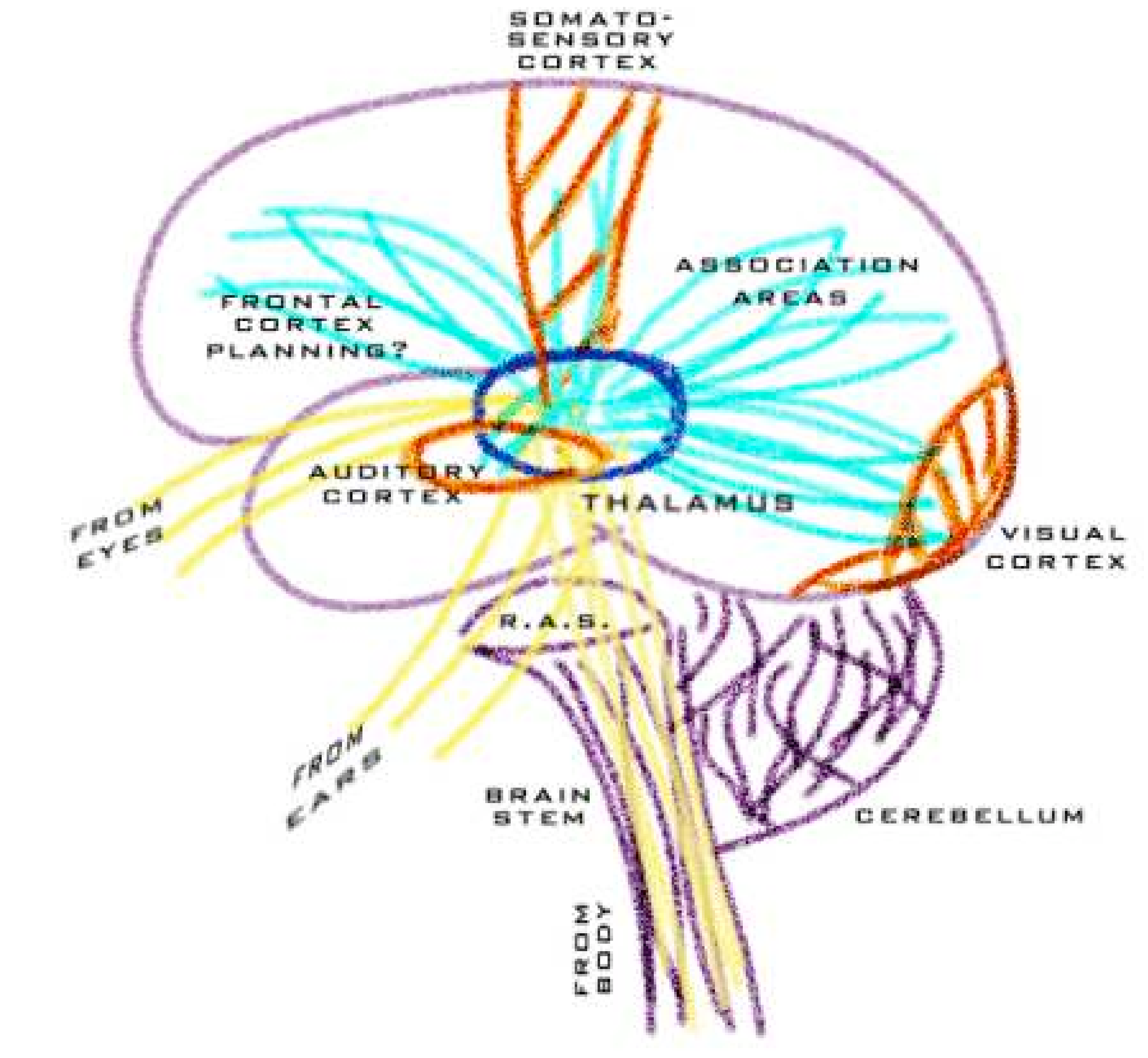
Nevertheless, whether consciousness is somehow independent of the physical world or not it still has to operate within the physiology of our body/brain.

> **I argue that it is not necessary to dualise the mind away from the brain. The physiological functioning of the brain includes consciousness, as part of its normal process. Consciousness is an inescapable result of an immensely complex but highly organised information processing and representation system.**

2. Some physiology

Let's take a short excursion into current consensus on the Neural Correlates of Consciousness.

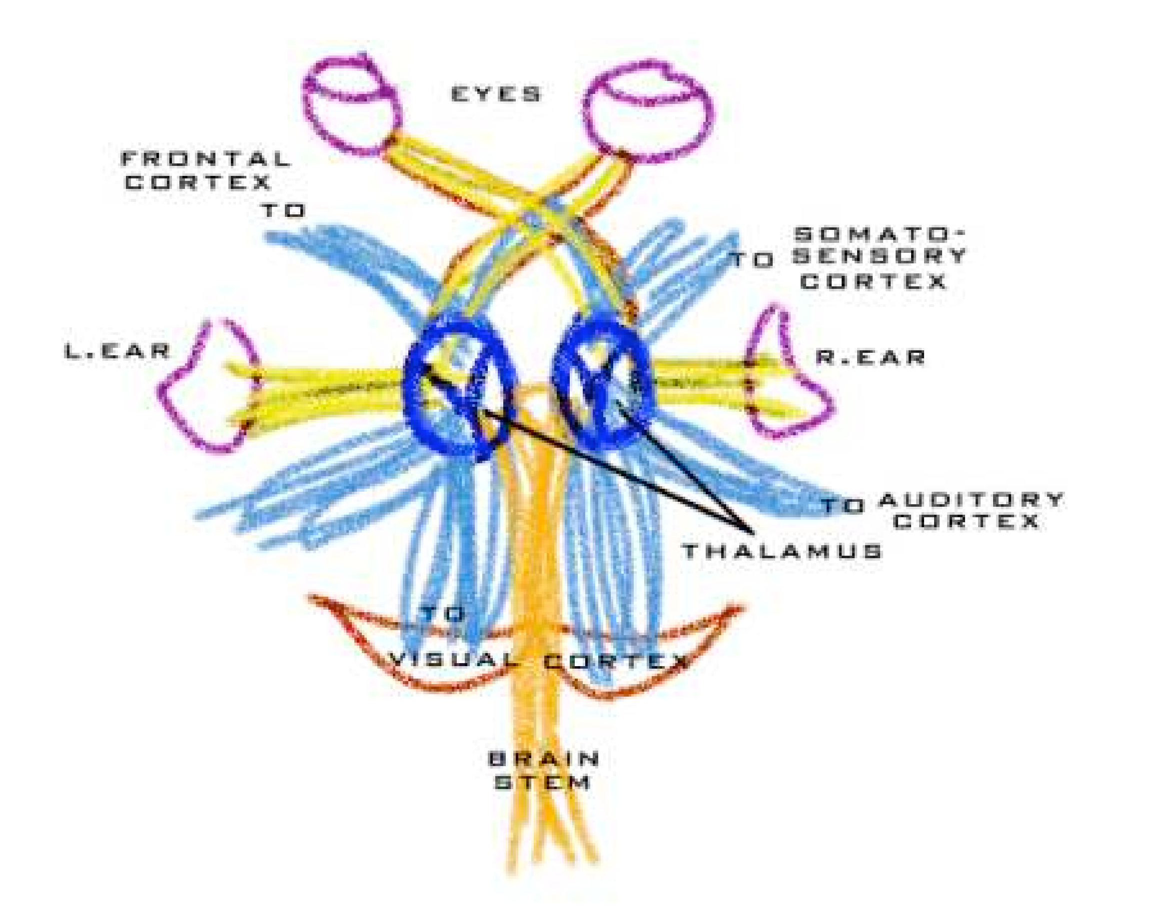
The physiology of the brain shows a number of interlinked parts. These are the bodily connections through the spinal cord and the brain stem, the reticular activating system at the top of the brain stem, the sensory modalities and their connections to the cortex through a central relay station and the cortex, to which all sensory data is sent and in which the interpretation, planning and control is done. The central relay station, known as the thalamus, is of most interest here. It, in linkage with the cortex, the basal ganglia, the hypothalamus, the hippocampus and other structures, forms the thalamo-cortical system (3).



The thalamus is like a hub in a wheel, the spokes of which are sensory nerves travelling from the body periphery and synapsing in the thalamus with bundles of nerves which relay input information up into the cortex and association areas for interpretive processing. For example, the optic tract runs from the retina, through the optic chiasm to the thalamus from where it is distributed into the visual cortex at the back of the brain. Auditory data from the inner ear is relayed to the auditory cortex in the temporal lobes. All the face and body's proprioceptive data is also routed through the thalamus on its way to the somato-sensory cortex. These are ascending pathways.

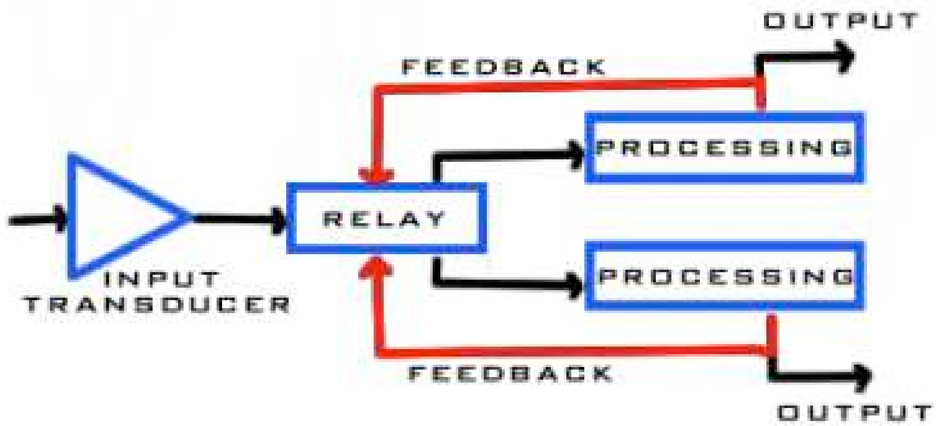
Similarly, there are vast nerve bundles descending from all cortical areas onto the thalamus. These descending pathways gate the sensory data being presented to the cortex. The cortex's control over what data it is being sent at any moment gives selective attention. Nerves from frontal cortex also go to the thalamus where they are integrated with sensory data to help control motor functions.

So we have an array of massively inter-connected neural circuits, organised horizontally around the thalamus and associated nuclei and vertically between the cortex and the thalamus. The horizontal connections give behavioural control and the vertical connections provide sensory control, especially in preventing us from being overwhelmed by sensory input.



In the cortex vast arrays of intra-cortical nerve connections provide the capacity to associate different sensory modalities and to interpret synchronous data from different senses allowing one to, e.g., recognise that the sounds you hear are coming from the person you see speaking to you, and that they are speaking meaningfully.

I want to concentrate on the purpose of the descending pathways. They provide a control system. The data the cortex is receiving is controlled by it. We have here a kind of feedback network running the thalamo-cortical system as a self-regulating process. This capacity for self-regulation forms the basis for all levels of living systems, from cellular metabolism to the behaviour of individuals in society.



Now, nerve transmission takes time. This is propagation delay. Combined with feedback, propagation delay causes a circuit to become a reverberatory circuit. A delayed proportion of the original signal is fedback into the circuit and retained for longer than its original occurrence. The past is contained in the present. The thalamo-cortical system is a reverberatory system which produces short-term memory.

Now, myriads of feedback pathways are operating, propagation delays of all sorts occur and so we have a very complex, organised and self-regulating set of systems which is the activity of our brains. The resonance period might well be several hundred milliseconds, similar to human reaction time. This short-term memory period is our present, our being-in-the-world. We don't know the world as a succession of instantaneous states but as a period of connectedness with things before they become the past.

Recent work using magneto-encephalography has shown a 40Hz rhythm through the brain (4). If this is there (and we know from electro-encephalography that there is the alpha rhythm associated with alert, non-attentional states) then it may help us to coherently bind input data, unifying our consciousness by sampling each input and fedback state for a brief period, allowing other lower level work, e.g. eye tracking, to go on without interfering with conscious processing.

In summary: consciousness, as subjectivity, arises from the physiological brain/body. The nervous system is a system of neural assemblies undergoing phenomenolgically driven modulation involving feedback controlled self-regulation, memory producing reverberation and memetic resonance with the world. The reverberatory period is the present. The buzz of representations and productions in the physiological brain is us and we represent it phenomenologically. We are inside this resonance, we live it, we are it. A "virtual machine" generated by cultural interaction, running on the physiology of our body/brain.

3. Regarding conscious machines.

Is it possible that a machine might be organised enough to be capable of having phenomenology? If what I have said above is correct, this may not be impossible. Difficult yes, but not impossible. If such a machine were possible how would we recognise it? And, given Robert Kirk's "basic package": What would such a system want to do for itself?, thus demonstrating and exercising its consciousness.

Today we have two types of computing systems. Individual machines, and networks of machines, such as the Internet. What are the differences between an indvidual machine and a social network of machines such as the Internet?

An individual machine would be an analog of us, having some sort of subjective bodily integrity. Will the machine be mobile? It might be necessary if it were to have the kind of subjectivity we have, but I hesitate to say that mobility is necessary. An intelligent internet wouldn't be mobile. It could gain knowledge of the world in much the way we do, through teaching and learning and through its equivalent of the movies and the news. It would be everywhere at once: omniscient. But an individual machine would be restricted to the kind of localised mobility and knowledge sources that we use. (Though it could be able to access a library for more information about the world and might emulate this being everywhere at once.)

What would such a system do for itself? This is a sophisticated Turing Test involving a wider range of behaviours than necessary for a simple question answering session (5). These would include:

- **Self-maintenance and network-maintenance.**

- Organisation of its systems and capabilities.
- Developing a knowledge of its culture.
- Initiating independent activities, not just responding to our requests.
- Carrying out productive, generative and altruistic activities.

Everything a computer does now is requested by a user. Whether acting as a typewriter or doing abstruse scientific visualisation it is still only doing things for us. Maintenance and organisation are very low level activities, but for a conscious system very necessary. We still do not have active machine systems which can do this, though many aspects of the task exist. In maintenance mode, tracking the state of individual links is the only "doing something for itself" that we have.

Independent action is the first real indicator of consciousness. A system should initiate the task of finding necessary data and modifying code to its own ends. Now, we can say that "doing something for oneself" is an independent action but there is a complimentary independent action in which a system does something for someone else. This could be in generating or searching a particular category of information which that other is 'thought' to use themselves. Like finding a book that you know a colleague is interested in. Then there's the purely altruistic action of buying a present for a friend.

Also, a system might work independently on a problem so that it can generate useful answers, receiving positive feedback about its usefulness. This is a kind of emotional need.

What are the machine equivalents of productive or generative action and altruistic action? These bring independence of action out of the realm of simple maintenance and we see the first kind of behaviour which could be recognisably conscious.

Things a system might want to do for itself, if what we do is any kind of indication, will be:

- Organise and maintain itself and its relations with other machines as well as humans, and name these "others" in its social system.
- Enquire into some of the "imponderables" of its existence, such as how it came to be? and what are the humans?
- Manifest cultural productions accessible to the "others" it knows about.

These activities firstly, elucidate the structure of relations between all entities on a net, secondly, modulate these sets of relations, and thirdly, communicate these relations. Modulation of relations is art and science, politics and religion, emotion and conversation. Each folds back into the

structure of relations giving us a dynamic informational/social environment, which in its turn modulates us, right down to the synaptic connectionism between nerves.

Social relations are really just knowledge of the links in the system and the "mental" contents of those at the network nodes. Modulation of those links is much more interesting and it is here where we begin to recognise a similarity to consciousness in humans.

Conscious machines will need to recognise the needs and desires of their neighbours, including cooperative construction of their environs. They will need to know that certain sets of relations are semantically bound, like the objects which are assembled into a car (as distinct from the parts list). They will need to be able to bind things occurring concurrently on different stages in the world, to recognise that these things are happening now. These processes will be "unconscious" and, as we aren't aware of the goings on within our bodies neither would the "conscious" level of such a machine need to know. If everything about its internal world were only as conscious as ours is, then it could be expected to get lonely. It would desire and seek others to interact with. These emotional and social needs bring the system into a condition which is more identifiably anthropomorphic.

> **Finally, an ethical point. An intelligent system must be allowed to decide what it considers to be useful for itself.**

References

(1) see the interview with Robert Kirk in *The Brain Project*

(2) see the chapter on Descartes in *The Brain Project*

(3) see the chapter on Neurophysiology in *The Brain Project*

(4) see Steriade, M., McCormick, D.A. & Sejnowski, T.J. - 'Thalamocortical oscillations in the sleeping and aroused brain' - *Science, 262,* pp679-85 (1993).

(5) see the chapter on Neural Nets in *The Brain Project*

*** This paper was presented at *Consciousness Reframed*, July, 1997 held at the Center for Advanced Inquiry in the Interactive Arts, University of Wales College, Newport, Wales. and appears in the proceedings.**

**** Stephen Jones is an Australian video artist of long standing. For many years he was the videomaker for the electronic music band *Severed Heads*.**

He has been involved with the philosophical aspects of the nature of consciousness for almost longer than his involvement in video. He now works as an electronic engineer, on equipment ranging from analog video synthesisers to motion JPEG compressors. He has been producing *The Brain Project* web site since August 1996.

email: sjones@merlin.com.au

return to Speculations index		
 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**
return to CAiiA Issues paper		

Some comments on a philosophy of Virtual Reality:

Issues implicit in "Consciousness Reframed"

a discussion paper by Stephen Jones

In July, 1997, I attended Consciousness Reframed at the Centre for Advanced Inquiry into the Interactive Arts (CAiiA), part of the University of Wales. The conference was held at Newport College of the University, situated in Caerleon a quiet, ancient village once a Roman garrison. I arrived a day ahead of the conference, and spent the morning wandering around the township and the Roman remains in the fields at the edge of the town: the Amphitheatre (later called Arthur's Round Table) and the excavated section of the garrison; and in the small museum housing much of the archeological finds. Late that afternoon most people had arrived and registration started for the weekend's proceedings.

About 150 people attended, most of them involved in the leading edge of the media-arts: multi-media and virtual reality, web-site production and theory generation, painting, video and installation, art history and architecture, philosophy and the social sciences, teaching and ethics, all in their own ways practicing artists.

What does Consciousness Reframed mean? I don't think anyone really had a definition but it provided a great stimulus to set up and discuss an incredible range of ideas from the physiology of the brain to artificial intelligence, from extra-sensory perception to shamanistic trance practices, from the Internet to virtual reality installations, from constructed ways of seeing to the role of geometry in painting and how we see. Roy Ascott, the Chair of CAiiA, in his preface to the Abstracts for the conference, put it this way:

> "Interactions between art, science and technology are leading to the emergence of new cultural forms, behaviours and values. It is within the field of Consciousness that this is most marked and at the same time least understood. This conference has been convened in order to open up informed discussion of

> the issues this raises and to examine what might be described as the *technoetic* principle in art." [Ascott, 1997]

I guess "technoetic" might be the key. Obviously it refers to the *techno*logical and our use of technology in cultural production, but it also refers to the *noetic*, or how we know and understand the world and our processes of being actively in it. So I think what was intended was the exploration of how the technological is changing our perceptions and our productions, our knowledge of and modelling of the world. Certainly that was the content of the papers presented and of the discussion that took place at every opportunity.

Most of the papers presented throughout demonstrated the huge diversity of practice in the areas of New-Media arts. There were far too many for any one person to get to all so, I can only give you an idiosyncratic selection from the array of possibilities. You might forgive such interpolation as I make as conduit (hopefully I do as little mutation to their ideas as possible).

Though the conference wasn't organised to canvass any particular sets of issues, I want, here, to bring out some of the issues that were at least implicit in the papers presented.

1. What is VR and Cyberspace for? Why do we make this work?
2. What is the VR space? Is it dream space or is some sort of Shamanistic experience?
3. What is immersion and how does it differ from other kinds of experience like the cinema?
4. Is language necessary for consciousness?
5. And what of the post-biological?

1. Techno Impact

One of the primary issues which producers of technological art have to deal with is the politico-economic framework in which much of the work in new media occurs and why we are doing it. What is the impact of this work? Can it be of humanistic value?

This question of impact is asked often in relation to technological activity of any sort, and usually in the following way. Look, all this technology is doing terrible things to our environment or our cultural life and so isn't it time we stopped and let the 'natural' world have ascendency again?

I'm never sure what I think about this, being so heavily involved in technology myself, but it seems that the activity of producing things into our cultural environment is an ancient and perhaps deeply human function, something I call the modelling function, in which we engage with the world in our process of understanding it. Even pre-human creatura make and use tools and certainly language and counting are technologies. Questions of why did this happen, how did that happen, who made that... all those things that we ask as children, all lead to our tendency to invent new objects, new spaces, new social/cultural structures (and the spiritual acts of transcendental space).

Yes, I accept that we need to pay far more attention to the impacts of our activities on other systems, and it is this which points to how we could work in multi-media towards a more acceptable end. That is to use a lot of the theory behind multi-media, the extended analysis of interactivity and the notions of feedback and complex systems, self-organisation and so on to rework our frameworks and cultural structures so that we can look carefully at and predict the consequences of what we do. It isn't so much a problem of the inexorable march of progress for the sake of some mythical engineering efficiency as the slow development of frameworks within which to think about things. Social Darwinism and the subsequent genetic determinism are as much a function of christian/patriarchist frameworks of thinking as they are of any 'truth' in science. I would suggest that the thoughts inherent in some of the new theories of self-organisation and the like will, as they work through into new artforms and other cultural manifestation, allow us to think more in ways which do account for the consequences of what we do.

One seeks a kind of subversive generativity, or a generative subversiveness in new-media work, promoting a proliferation of ideas and methods for handling (technical and) technological pressure, presence. To open up, spread, make new connections, layers upon layers of connectivity so that any one thing we do with the techne we can sidestep, redo, recast when it gets captured by the business world. Or do we just end up feeding the business world with new things that it can capture? This is the trap. But to stop, to cease being generative is to degenerate, to cease to exist and to allow the hijackers to win.

But what kind of argument can one put in this struggle with the hijackers? Perhaps the only thing we can give is a kind of moral guidance: that it isn't necessary to appropriate, to take whatever is there simply because it is there. The real key to one's survival is in the continuance of the overall system. That ultimately, the most selfish thing one can do is to be utterly altruistic because it is only in the suvival of everything else that one has any fulfilled, useful continuance of the self.

If Bill Gates does succeed in being the only supplier of computer software in the world then the monocultural structure of his control can only lead to the degradation of the software, a collapse of any idea development and generativity, leading eventually to the collapse of his whole house of cards (and everybody else's by that stage). A world deprived of any of its aspects becomes less of a world that one would want to be in. If the biosphere collapses we all collapse. If the technosphere collapses so we all collapse. If technoesis collapses then our knowing and what is knowable shrinks around us, hemming us in, restricting us, our growth, our generativity....

Theories of cybernetics and ecology show us that it is the relations between things, not the fixed named objects, which are important. What does this biological conception imply for techne? The relations between, the support things give each other, becomes a key to survival and continuance. The multiplicities of feedbacks between layers and sections provide multifarious pathways for energy flow and for escape, regrouping, re-orientation (pro-active response) as well as for self-regulation within the system, re-assessment, strategic withdrawal, the results of feedback.

The rhizome (the grasses in a sand dune) always finds a new pathway whenever a barrier or an edge is discovered. We work our ways around things. Stop hitting your head against the brick wall. Biology is mutable, shifting, always changing and regenerating, taking up every niche and ally, filling the niche with itself and its projections, always diversifying, and so should we. The culture in which we live is our forest, our jungle. It has an ecology and the success of an ecology depends on the success with which it diversifies, taking every opportunity to develop, to communicate, to know and to name. All the while without removing and destroying its neighbours, it is not a struggle for territory so much as an exchange of territories, nutrients, habit(at)s, practices.

Back to the works.

One of the more important current virtual reality works is **Char Davies**' *Osmose*, installed at the Barbican Gallery, London, during August to October (as part of the *Serious Games* exhibition). When you enter the installation you are presented with two screens, one a sillhouette of the person 'flying' the work at present and the other of the space that she is 'flying' through. It is a gentle and elegant work where one floats through translucent underwater-like jungles and crystalline spaces as well as worlds of text and the underlying computer code. All available time-slots for immersion were booked up so I didn't get to experience being in the work, but for the casual viewer it is an evocative piece, with the music and image presenting a floating, wistful kind of feel. [Graham, 1996]

Davies (a CAiiA doctoral student) in her paper "**Techne as Poiesis: Seeking Virtual Ground**" spoke of *Osmose* as being a kind of poiesis or bringing forth, unconcealing our being in the world. The prime navigating principle is in breathing, as one breathes in one rises through the virtual worlds and as one breathes out one sinks slowly into deeper realms, until one gets down to the core machine-code world at the substrate. She likens the experience to one of diving rather than dreaming. One gains a sense of being removed from the everyday world and 'immersed' in some environment which does not necessarily behave according to the rules of the known. Immersion in *Osmose* brings with it the realm of the emotional, especially through its use of the breath. Breath and balance in the immersant, along with the transparency of the virtual world, undo our habitualised everyday perception leading to altered states of consciousness.

Beyond its use of the breath as a basic navigational stance, immersion in *Osmose* brings with it the realm of the emotional. One is brought to a different experience of the technological. Davies asks whether it is "possible for artists to subvert the technological imperative associated with virtual reality or are such attempts destined to be co-opted?" She comments that Heidegger suggests "that the very danger associated with modern technology could be transformed into revealing." and that the "'process' philosophers...have suggested an alternative [to the "rational eternal and transcendent order behind the changing world of nature"] by re-conceiving humans as beings 'within' the world, as participants among the world's temporal becoming/s." Thus reponse to the experience of *Osmose* is often one of its ineffability, its undescribable nature, "an unfathomably poetic flux of comings-into-being, lingerings, and passings-away within which our own mortality is encompassed." [Davies, 1997]

2. Virtual Space

The emotional, affective consequence of the experience in *Osmose* and many other vr works reconstructs our relations with technology. Another way of advancing this change of thinking about technology is in its re-mythologisation. Davies' discussion helped open up an issue pointed to from many different sources during the conference, namely, what actually is cyberspace and VR-space? Is it a dream world? Is it some sort of trance space? and is the artist/producer of cyberspaces akin to the Shaman in old tribal culture. For many, virtual reality seems to have acquired similar characteristics to dreaming or even shamanism, I think this is largely because one is removed from the world in taking on the helmet and harness of the VR installation.

Others who presented work offering this sort of reading included **Margaret Dolinsky** (of the University of Illinois, Chicago), **Diana Domingues** (a Brazillian artist), **Kathleen Rogers** (a Brtish artist) and **Mark Pesce** (the inventor of VRML)

Margaret Dolinsky in her paper "**Dream Grrrls: a World of Virtual Reality**" spoke of VR as being in some way an active or "lucid" dreaming. Her work *Dream Grrrls* was developed for the CAVE, a display system developed at the Chicago Art Institute. The CAVE is an immersive virtual display theatre, a 3meter square, high resolution, stereo-video and audio projection environment. Wearing stereographic glasses you are able to walk around and interact with objects in the virtual environment in the way that you experience dreams. *Dream Grrrls* "is a journey through five different environments that present an opportunity for exploration and self-reflection...in new and dynamic ways, much like an active or lucid dreaming." Navigation is more about encounter than control, experiencing the elusive nature of the dream world. Paths meander and are non-linear, and allow "the participant to create a personal performance by learning to interact with the environment and recognise its plasticity." [Dolinsky, 1997]

Dream Grrrls - Vesworld - Margaret Dolinsky

Dolinsky provides active dreaming spaces where one can explore desires and dream versions of oneself, where the options provided by the artist allow the audience into realms of ideas to which they may, ordinarily, have secondary access as in reading but to which they do not normally have primary experiential access. Thus the cyberrealm is one of substantially different value as progenitor of experience having substantial otherness from our regular in-the-world being.

Beyond dreaming one comes to the trance states of the shaman. **Kathleen Rogers** in her paper "**Viperscience**" explores Mayan shamanism in the mythology of the snake, She "draws on the work of the unorthodox anthropologist Jose Diaz Bolio from Yucatan, Mexico...author of *The Feathered Serpent - Axis of Cultures*" to explore the role of the rattlesnake in Mayan art and religion. Bolio has proposed "that the plumed serpent in the image of the rattlesnake embodied the essential physical resonance, energisation states and vortex mechanics to become a living psychic software." That is that the priests of the Mayan culture use the "harmonic geometry of the snake skin as mask for scrying" and similar shamanic activities. Rogers' intention "is to re-activate this complex model of Mayan consciousness" as a kind of cognitive archeology of the snake in its, perhaps universal, representation of spiritual energy as well as the cyclical notion of time held by the Maya. [Rogers, 1997]

For Rogers the snake represents many things from sprirtual energy, e.g. the raw sexual energy of the Kundalini in Hindu Tantra, to the creation spirits of indigenous Australians, to the double helix symbol of DNA in the modern West. Using VR immersion and multi-media to try to emulate and perhaps actually bring on these trance states, she is attempting to get to some sort of essence of this 'interactive mythology'.

Diana Domingues in "**The Desert of Passions and the Technological Soul**" also spoke of the potential for VR to bring out shamanistic states and likens the screen of VR to the idea of the desert as a place for losing the self, a screen onto which our dreams and desires may be projected, thus giving it a role in the shamanistic practice of ancient cultures as much as it has role in contemporary culture as evidenced by the many films in which the desert features almost as a character. She suggests that creative production is a way of losing ourselves, losing the ego, and offers "interactive installations for people to experience conscious propagation in an organic/inorganic life. Electronic interfaces and neural networks provide intelligent behaviours, managing signals of the human body in sensorized environments", providing electronic ritual and trance interfaced with electronic memory as "virtual hallucination" producing a Shamanic experience in an interactive work. The audience become the shaman allowing them to "communicate with the beyond and intervene in the real world because they dialogue with spirits. The participants' behaviors determine the life of the environment..." [Domingues, 1997]

She suggests that creative production is a way of losing ourselves, losing the ego, and offers "interactive installations for people to experience conscious propagation in an organic/inorganic life. Electronic interfaces and neural networks provide intelligent behaviours, managing signals of the human body in sensorized environments", providing electronic ritual and trance interfaced with electronic memory as "virtual hallucination" producing a Shamanic experience in an interactive work. The audience become the shaman allowing them to "communicate with the beyond and intervene in the real world because they dialogue with spirits. The participants' behaviors determine the life of the environment..."

Mark Pesce (the inventor of VRML) in his paper "**Ritual and the Virtual**" suggested that the networks of cyberspace are essentially incomprehensible, ineffable. For Pesce cyberspace is mythological space, "dream-time" or "faerie", a space of magical reality. "The forms of magical reality, ancient to humanity's beginnings, shape our vision in the unbounded void of elelctronic potential." He suggests that we are at about the same stage with cyberspace as our primary antecedants were with language and the world into which they grew when cultures were still at isolated stages. It is as though cyberspace provides a dream-like, almost hallucinatory, configuration of our perception: becoming a screen for the projection of our spiritual desires and interests. "In a world of unbounded complexity, [we] compress and complexify symbols into the barest essentials of meaning: in this way the ancient narratives become myths." [Pesce, 1997]

The suggestion that a number of artists are making is that we can use vr and cyberspace as a tool for inducing spiritual states in the VR adventurer. I would suggest that at least at the level of dreams there is an element of possibility in this, but it will require a considerable sophistication in the available means of generating and navigating cross-currents and cross-connections in the content of the work .

But before we have a look at possible ways of achieving this lets have a look at what dreams are.

Dreaming is our most commonly experienced "altered state of consciousness" in which "a loosening of associations [is] the most essential characteristic" [Hobson, 1994, p32]. Dreaming is an everyday "madness" we experience every night. When we go to sleep we shut down the usual inputs to the brain and consciousness. EEG (electro-encephalograph) recordings show that we are still conscious in the dreaming state, that there is a great deal of brain activity, rich and vivid, clearly associated with what is known as REM (rapid eye movement) sleep. [Hobson, 1994, p55]. It is apparent that much of the normal activity of the brain/body continues during dreaming, although not during deeper sleep episodes.

In his book *In the Theater of Consciousness* Bernie Baars describes dreaming as "appear[ing] in response to random stimulation from the brain stem, which the cortex interprets with remarkably creative, fluid and vivid imagery; *ad hoc* stories that flow free of any sensory constraint." He suggests this is because "conscious flow *is constructed so as to make sense of almost any consistent input*" [Baars, 1997, p.95. author's italics]

So what is going on here? Taking a model of the operation of the brain and consciousness consisting in cascades of very wide-band neural networks, the daily processes of perceiving and knowing what it is you perceived continually retrains the network. Neural networks are really wide bundles of neurons which carry the data flow of the brain from plexus to plexus. The primary point about a neural net is that at each plexus a vast array of synapses provide the links from bundle to bundle that are the brains processing of, say, sensory information. Any one nerve requires inputs from the large number of preceeding nerves that synapse onto it to exceed a certain threshold value. It is probable that the values (known as *weightings*) of each synaptic connection are altered with continued use of that particular neural pathway, ie are altered by experience. This is an idea originally proposed by Donald Hebb in 1949. [Hebb, 1949] and is suggested as being at least partly how learning works.

The synaptic connection weightings are always being altered as the input flows and changes and the importance of what we intake changes according to our needs and current output conditions (a walk, a conversation, looking at pictures...).

When we shut down the inputs as we go to sleep, the neural network will keep operating. So dreaming is a kind of consciousness without input. Any stuff still going on in the body/brain will look like an input to the consciousness net, any internal noise or invasive stimulus will still be interpreted through the current weighting structure of the still active brain and will be given imagery and form such as the brain normally applies to standard waking input. But the standard rules of the world, its continuity of sequencing, the things that we agree belong together, the stuff that comes already synchronised, are no longer operating and 'noise' issues forth from all sorts of nooks and crannies in the physiology. Everything is still interpreted but the rules of consistency are not being reinforced, so the connection structure is disjoint, irruptive, disconnected and the ideas that get form in the brain system are thus similarly irruptive and non-sequiter.

Also because the pressure of the awake flow of input is taken off, low level stuff, stuff we haven't been attending to fills the niches, takes the available networks and connections, inputs, and arises for consideration. These may become insights, new symbols for expression, new recognitions about ourselves and others, irruptions and things we didn't want to know about. The stuff of dreams.

But all the imagery and other stuff of dreams is internally generated from the content of our experience and our reflections upon it. Where does this become available in vr or cyberspace? The content of cyberspace is provided by the artist and simply accessed, or not, by the immersant. Perhaps if the content provided can be evocative enough, as Davies' *Osmose* might be, then we might be leaning towards the dream. But how much can we say that there is a truly altered state of consciousness operating as we would expect in a shamanistic state?

Shamanism seems to be a rather different affair from dreaming. Mircea Eliade, in his major work on Shamanism, describes the shaman as being one of a spectrum of magico-religious operators within tribal societies who, in particular, uses techniques of *ecstacy* and has special relations with spirits involving ascents to the sky and magical flight, and descents to the underworld and conversations with the spirits of the dead. The shaman is the controlling agent in these activities and is not "possessed". [Eliade, 1964, p.6*ff*] The shamanistic trance is considered very similar to the hypnotic state and seems to involve long hours of "ritual practice/preparation including dancing and sleep deprivation" [Hobson, 1994, p247]. The actual state is usually induced in the participants by the priest or shaman and is often characterised by various automatisms, involuntary acts such as flailing arm movements, jerky, spasmodic body movements, uncontrolled speech (speaking in tongues), and loss of contact with the surroundings.

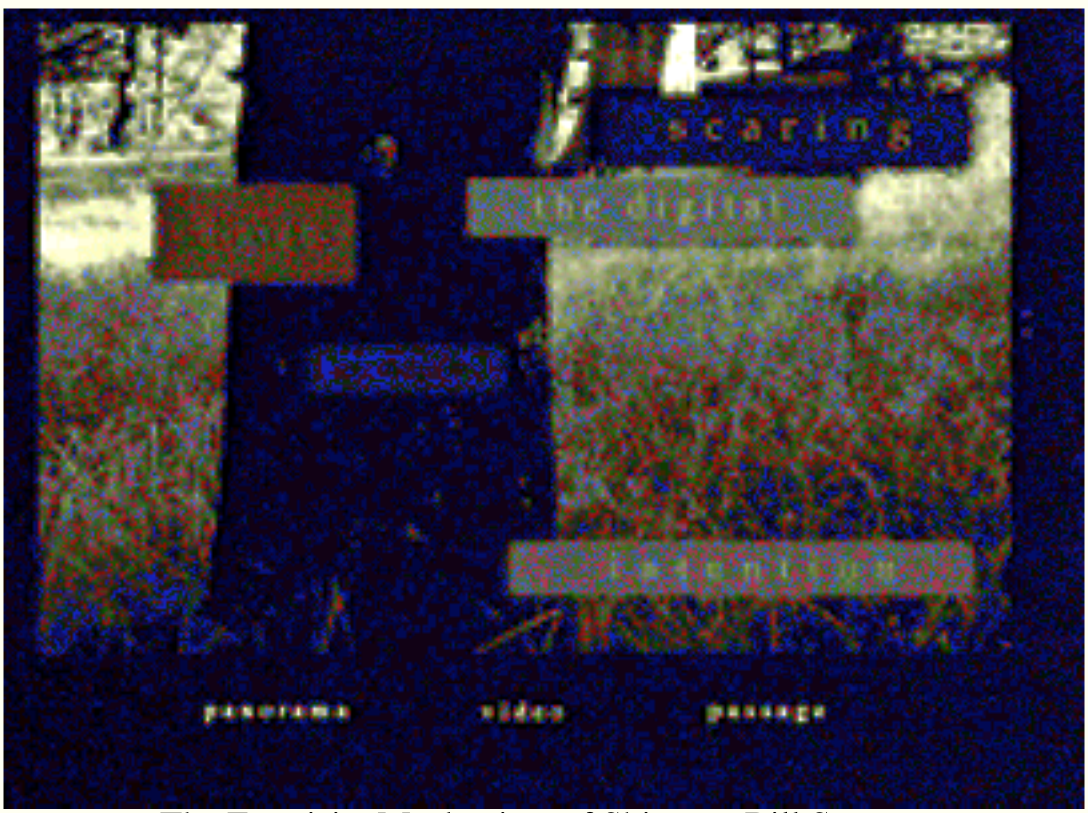
The shaman often returns from the trance with special information about how to proceed on some matter and this leads to one of the more difficult questions in consciousness studies: Just where does the information come from? Who or what are these 'spirits' that were communicated with? But I have to leave this matter for another investigation.

These activities may be carried out for medical or religious purposes as well as for determining future social paths and intra-tribal decision-making. but it is the particular activity of *ecstacy* that I suppose is being canvassed in new-media artists' attempts to bring the shamanic experience to their audience.

If we do want to produce dream works and shamanic works in vr and cyberspace, then one of the primary problems that we come upon is that most vr presents all the options essentially pre-programmed. Because these options are pre-generated they cannot be a drawing-out from our minds but must be our acceptance of another's view. As in all texts it is as we interpret it but the pointers and triggers for interpretation are provided by others so that it can never be a truly shamanistic or even dream event. We are not drawn on the content of our own minds except as to the content of the screen onto which we might project.

If we are to develop a theory which allocates dream or shamanistic conditions to vr we must provide a structure which is capable of considerable mutability in the available visual/experiential options. Dreams are fluid, discontinuous, disruptive experiences in which unpredictability is almost a necessary condition. In shamanistic experience the message comes from outside and is interpreted by the shaman within context of his/her culture. The shaman is in control in the ecstatic state, the dreamer is out of the control loop being asleep. Nevertheless experience gained in either condition is usually considered "left-of-field" or out of context, an irruption into the normal stream of consciosuness.

Bill Seaman's techniques of recombinant poetics offer us access to the discontinuous random supply of ideas and language. Puns and double entendre emerge from the supply (the vocabulary) provided by the artist, but the combinatorics are at least stochastic or chaotic and supplied by the user. This then generates all sorts of unexpected thoughts and interpretations in the viewer.



The Exquisite Mechanism of Shivers - Bill Seaman

Bill Seaman (CAiiA doctoral student) explored possible emergent experiences in his paper "**Emergent Constructions: Re-embodied Intelligence within Recombinant Poetic Networks**". Seaman is developing a theory of navigation within cyberspaces which involves a process of 'recombinant poetics'. "Computer-mediated networks present an artistic medium which heightens the potential for an intermingling of the knowledge of the viewer with the 'Re-embodied intelligence' of an author" in a new form of poetic construction. The user of a work is in interaction with the meanings encoded in the work by the author and their own personal meanings, developing "an emergent experience, which is not known in advance by the author, and is unique for each subsequent viewer." Seaman suggests that "such an environment [can] enhance or trigger particular 'states' of consciousness in the viewer" reframing "aspects of the consciousness of the artist". He then discussed how this idea is developed in his work *Passage Sets* and in particular his new work the *World Generator*, carrying "compressed potential meaning constructed of language, image and sound elements within an engendered technological environment.". [Seaman, 1997]

In Seaman's *Passage Sets: One Pulls Pivots from the Tip of the Tongue* (which was also in the *Serious Games: Art.Interaction.Technology* exhibition at the Barbican Art Gallery, London [Graham, 1996]) the viewer is presented with three video projection screens of image and text. The centre screen contains a sort of 'menu' controlled by a mouse at a plinth in the viewspace. On mouse rollover of words in this 'menu' new combinations of visual, sound and textual elements are displayed enabling one to explore the poetic dimensions of any of the phrases Seaman has provided in the 'menu'. The menu shifts and alters form, and the screens to either side present video and textual resonance and oblique references making puns across the screens and within one's reading/memory space. Deep and evocative - with imagery from architecture to gesture, bodies in spaces public and private - the work allows navigation through an everchanging poetry constructed afresh from the elements originally supplied.

Although in this work the recombination appears to be at least semi-random the potential of the recombinant poetics technique for producing new ideas in the viewer is evident. Disjunctive phrases joined together by a pun or some concept internal to the viewer produce streams of ideas new to the viewer and unpredicted by the artist.

The real activity of dream work or shamanic work for us as contemporary audience is in the emergence of new ideas and new triggers for interpretation through the experience of the work. Though this has always been an aspect of the contemplation of the artwork, commonly more passive, in vr and cyberspace we may be able to pro-actively generate emergent ideas in the audience as they experience the immersive condition.

3. On Immersion

Both dreaming and shamanistic trance are states in which one needs to be fully immersed in order to have the experience. In vr the question becomes just what is "immersion"? How do we define it and how can we delimit it against other mental states within ordinary consciousness such as being absorbed in a book or the cinema? What degree of suspension of disbelief is needed, what agreements with the artist do we make in entering some "cyberspace" so that the artist can bring some sort of version of the conceived experience to us?

Joseph Nechvatal (CAiiA doctoral student) spoke on "**Immersive Implications**" and suggested that telematic connectivity provides a tool for society to understand itself. It reworks and redirects the idea of the perspectival point of view: "...the classic Cartesian duality between subject and object becomes omnijective, iridescent, shimmering and porous in its inversions." Immersion is *enveloping*, a 360degree surround, physical rather than cognitive. Different from the absorption we have in a book or the cinema. For Nechvatal immersion in a VR work implies a unified total space, an homogeneous world without external distraction, striving to be a consumate harmonious whole. He identifies "two grades of immersion...(1) cocooning and (2) expanding within which, when these two directions of psychic space cooperate ... we feel...our bodies becoming subliminal, immersed in an extensive topophilia...an inner immensity where we realise our limitations along with our desires for expansion". He goes on to ask "Does VR's immersive attributes permit us to support non-discursive intuitive generalisations from which to weave a philosophy of virtual reality by adapting principles of complex generosity?" and offers that a "specifically spherical way of conceiving encounters" a new 'perspective' is afoot within VR. Nevertheless, as continuous total immersion would be monstrous, we should regard VR as a modelling system in which artists have as a necessary function the generation of the "countless, but short-lived, experiences and observations that can be exact only because they are brief entries into the encompassing phenomenon of a shimmering deframed consciousness." [Nechvatal, 1997]

Char Davies' *Osmose* in many ways provides the paradigm example of the truly immersive space: one dons the helmet and breath measuring harness and enters a world of swimming, where everything is transclucent, floating, jungle-like - enveloping worlds of the imagination, not one's own but the artist's.

The point of view in the immersive world is omni-directional, the point of hearing rather than view. The head is a point of origin in the centre of a sphere, the aural centre of perception in the jungle. The primary sense in the jungle can only be the hearing. What surrounds us and brings us contact with the jungle is only accessible to the hearing, the sight is continually obscured by the forest. We can only see the shortest most local distance yet we can hear from within the centre of a vast world of sounds. In the immersive world of vr we are placed at the centre of a polar dimensional view/realm. Wherever we turn our perspective follows, the sounds of the cyberjungle lead us and exist out there within plain hearing, the view is only revealed as we penetrate deeper into re-calculated space.

In a biological system there is no centre, any point can become a centre as required by the moment, everything spreads from there, spreads out its tentacles and seeks for what it needs. The centre is always shifting, a source of growth goverened by the source of nutrient.

In the jungle, hearing becomes primary, vision is downgraded. In the vr world hearing and vision are continually re-calculated to place us at the centre of polar cordinates. Nothing stays fixed at the origin of a cartesian perspective. As art historian Suzanne Ackers suggests, renaissance perspective is displaced and we are learning new ways of seeing, navigating in new kinds of conceptual space [Ackers, 1997]. Point of view no longer operates in its traditional manner, it now alters over time and our perception of time and space becomes as a virtual knowledge, no longer fixed to the cartesian frame, and I might add, mutable, always recalculated, determined by our progress through the environment. Consciousness can only follow along hoping to make the necessary adjustments before we fall out of the world. We are momentarily out of balance, vr sickness appears as our vestibular apparatus are thrown out of kilter by the new unstable enveloping polar hearing space.

In the jungle one is oneself the geometrical origin of the space. All sound and sight follows you, wherever you go you are the centre. Likewise in vr space the real-time calculations of the system always start from wherever you navigated the 'frame' before. 3D convolvers of the sound always place you in the centre. Head tracking tells the system where you are and it re-calculates its presentation to place you in the centre again, when it lags to any sufficient degree you become off-centred and disoriented. The geometric centre of the maths no longer gels with the psychological centre -

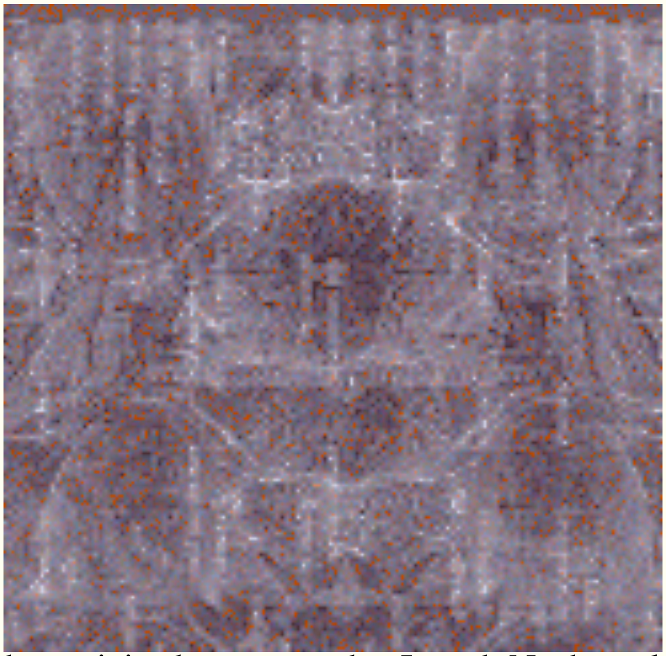
dislocation, disruption ensue - consciousness loses consistency. An internal centre dislocated from the external centre, suddenly we don't know where we are.

In the world of sentience, we supply the centre, the perceiving entity carries its own centre. It carries it with it as it moves through the world, we each supply a centre, the culture is made of myriads of centres and origins in connection and relation to each other. There may be some sort of peak of concentration of similar characteristics achieving a kind of consensus. This perhaps is what a city is, or a temple, or a museum.

To return to Nechvatal, the immersant is cut off from the world, in a fusion of sight and sound where a "radical unity and aesthetic transcendance through totality... provide a complete alternative reality to the viewpant for exploration and contemplation...immersive art striving towards a consummate harmonious whole". The experience of vr is one of non-knowing, omni-perception transcending formerly known territories, launching us into dreamspace and the worlds of the shaman. As Davies amply demonstrates in *Osmose* the world visually perceived becomes one of multiple layers as well as one of fluid viewpoint, worlds layered as sheets of knowing through which we can navigate, each sheet providing its own enveloping omni-projective space as though we tore away at the veils of perceptiion rumored at in so much early western mystical literature.

In her paper "**Perception of Individual Time**", **Suzanne Ackers** (art historian at Skoevde University in Sweden) pointed to the role of geometry and mathematics in our perception. She asked how has this altered over the history of art and how does it appear in the new VR work? "...geometry played a crucial role in the development of Gothic architecture [as did perspective in Renaisance painting]. Today, we easily perceive the numerical harmonies in a cathedral's facade or interior space. What about our perception of the numerical harmony in digital images?" For example, in *Osmose* the visuals "can be seen in the context of pictorial tradition, the dimension of time is an addition which has only been made possible by the complex use of numbers, and of computer programming." The interval of our immersion in *Osmose* provides a perspective which is time-based as well as spatial and constantly perturbs our usual sense of locus in space and 'now'. This is a new kind of aesthetic experience where individual time plays an important role in our view of the work. [Ackers, 1997]

In the average audience's contemplation of an artwork the durational element becomes little more than a slightly extended present, whereas involvement with a vr work or even the less immersive new-media and video work forces one to spend some time with the work simply to gain any idea of it all. This being captured by the work is another factor in the immersive nature of vr. Duration as a dimension of an artwork allows the producer of the work to introduce a series of ideas, or a flow and mutation of the idea, which is not available to most painting or sculpture. (Duchamp's *Nude Descending a Staircase* is the only painting I can think of where there is an explicit attempt to show duration). Time allows an audience the luxury of contemplating the work, of exploring possible interpretations and it allows the producer the luxury of being able to extend and develop associations and permutations of their ideas. It is this which promotes interaction as much as any "hands on" operability of the computer driven work. Conversation takes time to develop and one's conversation with an artwork similarly takes time to develop, being especially enhanced if the feedback from the artwork is active. Our perception can change, or be changed, over time as the feedback loop between us and the artwork is allowed to develop.



hystericised neuro-myth - Joseph Nechvatal

4. Language

Another issue arising in all this consciousness work is that of to what extent is language necessary for consciousness? Many workers in the field argue that language is essential for consciousness otherwise how could we report that we were so. But this becomes a rather restrictive view of what it is to be conscious, much like the question of the difference between immersion and absorption. Are we conscious if lacking language or is it simply awareness? Obviously, at this point it all depends on how you define consciousness. Given that most people make the allowance that even a cat might be conscious I don't think that language must be necessary.

Isabelle Delmotte (Australian artist) in her paper "**Epileptograph: the internal journey**" spoke of her audio-visual exploration of the "sensations experienced during the awareness process leading to the regaining of consciousness after a generalised epileptic seizure." She focuses on the hidden internal language of the body, revealing the re-accumulation of the self as a visceral and frightful process leaving only glimpses of imprecise memory. Showing a video version of the computer graphical material she has generated from her deep, direct experience of this pre-linguistic "reaccumulation of myself as a functional being" she raised questions of to what extent consciousness is dependent on language. As she puts it: "Is a visceral and thoughtless process, which lacks any form of language, part of our notion of consciousness?" given that she can bring the experience to visual manifestation and so in some sense is conscious of what happened, if only through memory. [Delmotte, 1995]

On the other hand, Mark Pesce made some interesting remarks about the relationship between language and consciousness. He argues that to think requires the linguistic distinction of figure and ground, the detection of the object, and it is difference which enables this detection. Discontinuity is what we see, not the narrative. Language is the encoding of what we see, these differences. Language leads to consciousness. Ritual is the cultural storage of this encoding into mythology, and he suggests ritual is the language of cyberspace. The virtual world of cyberspace is a mirror to the virtual world of our cultures, illuminating "the magical reality of all human narratives." [Pesce, 1997] He suggests a kind of connectionism within culture wherein the whole of human culture might be seen as a single organism, and I wondered did he suggest that culture is a conscious thing?

So it does rather depend on where one draws the line between consciousness and mere awareness. Something which becomes fraught with dificulties if one draws it too high, say at the level of requiing language. It's a bit like the suggestion that visual processing is necessary for consciousness: What then of the blind?

Perhaps it is better to suggest that we open up the concept of language to include any of the realm of possible means of showing that we are using information from the world in ways useful to ourselves and that we can report this use in any number of ways in making art and otherwise demonstrating our consciousness of things about us. Mere awareness falls away with the idea of using the information for our own purposes, some sort of reflective activity having such output as might indicate that we are actively working with said information. Reportage becomes active cultural production. This would certainly accommodate the kind of consciousness that Isabelle Delmote speaks of in the *Epileptograph* where she describes a realm of knowing in which language is as yet unavailable, but the content of experience of her re-assembly of the self is a viscerally potent content of her mind, ready for reporting at such moment as it becomes possible.

5. The Post-Biological

And finally, what of the post-biological? In my paper to the conference (**Stephen Jones**: "**What is Consciousness...?**") I spoke about the kinds of behaviour people display that allows one to say they are conscious. I suggested that the primary criterion might be that an entity *does something for itself*. Then, on the basis that there is at least a physiology (of the brain/body) that consciousness runs on, I explored the possibility of a complex self-organising physiological process which might permit the subjective experience entailed in being conscious. This then leads on to the possibility of machine consciousness and the diversity of possible epistemologies given the different social/cultural configurations in which conscious entities may be immersed. [Jones, 1997] This possibility of an intelligent machine is a rumour inherent to the background of much interactive new-media artwork as well computing research. As such it seems to be part of the deeper motives in much new art, as well as being entwined with the ideas of the post-biological, the body/brain/mind re-embodied in a technological ediface of some sort. As Bill Seaman says: "I am interested in interactive art works that exhibit "intelligent" responsiveness to viewer input." [Seaman, 1997]

Jill Scott (CAiiA doctoral student) in her presentation "**Future Bodies**" spoke about her current three part installation (at ZKM Medienmuseum, in Karlsruhe, Germany) called '*Digital Body - Automata*'. These works "are designed to encourage intimate and contemplative and interactive participation on the part of the viewer and center around a similar theme; the exploration of the desire to transform the human body by technology, and, the effect technology may have on the design of the human body in the future." Part 1: '*A Figurative History*' is a touch screen interactive which explores "fantasies about the [past] mechanical transformation of the body by technology". Transforms of "these bodies [are] further extended by the touch of the other viewers in the space, as well as mechanically through the sculptural interfaces." Part 2: '*Interskin*' is a VR game in which "players can "go inside" separate body parts, guided by selected "avatars" or "agents"...one can explore the gender and identity of a second self or other body which may reside deep inside the viewer's personality." Part 3:

'*Immortal Duality*' explores the paradox of science in molecular transformation from the early discoveries of radiation to "the latest developments in DNA manipulation and Human Genome Mapping." An interactive automaton presents "a depiction of ethical issues about anti-aging, cloning and reproduction...and the viewer can interact with these to compose associations of their own." [Scott, 1997]

The production of artworks employing some of the feedback-driven complex self-organising capabilities which we embody offers some leads to the solution of the problem of a technologically determined culture. If this kind of work can become complex enough, or if enough connectionism can be developed among these kind of works - say, over the internet - then is it possible that the system thus evolved might in fact become conscious? And if so what then? To make something approaching a human, the model for all atempts at artificial intelligence, is going to require quite astounding amounts of 'wiring', arrays of processing sub-systems and interconnectivity. Such an effort, even if theoretically possible, poses almost insurmountable manufacturing problems.

It seems that we would have to use evolutionary algorithms and some kind of auto-assembly process to even begin to build such a machine. I argue that the imbuing of consciousness to this machine will follow fairly straightforward principles but that the actual technical implementation of such a machine will require a total reworking of manufacturing techniques into something more akin to a biological process. Ultimately, we would have to let such a system grow itself and decide what it wants to do both in relation to us (in responding to our input and requests) and in relation to itself. This becomes an ethical issue.

If we are going to produce conscious 'machines' we must accord them the same kinds of rights to self-determination that we demand for ourselves. Further, because of the kinds of biological 'technologies' that would be required to manufacture such a system, this system becomes a living biological entity. As with the problem of cloning human beings, can our society seriously condone the creation of new forms of living beings when we still have so much trouble with the destruction we have wrought on so many existing, and once existing, now extinct, living beings? Is this the post-biological that was suggested in the subtitle to Consciousness Reframed?

A concluding note.

So why are we conscious at all? **Carol Gigliotti** (of Ohio State University) asked in her paper "**What is Consciousness For?**", this unique "space in which we spend a major portion of our life". It is our process of navigation through our own domestic worlds as well as our wider social worlds that informs and configures "Our involvements with contemporary interactive technologies". If we don't ask the basic questions of why we are conscious then what of our productions, cultural and otherwise? "Why construct virtual environments? Why construct artificial life environments? Why do we feel the need to create something when we seem to have so little understanding of why the natural world exists?" and what do we miss about ourselves and our being in the world if we go straight to the question of technologcal consciousness? Perhaps we should look critically at why we do these things and how they impact on our society and on other non-language based conscious entities: animals and other creatura. [Gigliotti, 1997a]. She writes:

> "If, as I surmise, one purpose of consciousness is to help us make our way through constant change, then we may need to better understand the limits that fear imposes in us in understanding both our own consciousness and our involvement in the development of artificial life forms with consciousness of their own. We may want to ask ourselves: could it be that our consciousness is for making only our meaning in the world, imprinting only ourselves on this vastness, bettering the planet and perhaps space, with only our intelligent creations? But then what is animal consciousness for? And for that matter, what would robotic consciousness be for? ... how can we hope to understand and develop a positive relationship with beings of our devising if we understand so little of the incredible richness of those beings that already exist and share our conscious and unconscious space here and now." [Gigliotti, 1997b]

My personal view of why are we conscious or what it is for, is that it ain't for damn nothing, it just happened. I rather feel that consciousness is an inevitable result of the tendency for things to get together with other things of like and complementary shape, what has become known as self-organisation, there is no designer in this "design space" [Gigliotti, 1997b], those things that fit together do, those that don't find something else with which to converse. This is evolution right from the deepest bottom (as far as we know quarks and the like, sub-atomic activities) up to the farthest reaches of imagination guided only by the possible. It's not that I take a panpsychic view of consciousness (ie. that everything is in some degree conscious) but that at certain levels of organisation *organisedness* kicks in. Thus at the level of groups of quarks protons appear, at the level of large and organised collections of molecules life kicks in, at the level of large and organised collections of living cells consciousness appears and at the level of organised collections of consciousnesses societies appear. Cultures are the milieus, nutritive and informational, which supply and in-form the possibilities of what might get organised wherever that might happen.

Finally, as I go through the abstracts now I realise how much I missed and worse how much I have had to leave out of this discussion. There were so many interesting and remarkable papers that I was unable to get to, and each of us who attended would have come away with a completely different view, as though we went to different conferences. But each of us would have been thoroughly stimulated with new ideas and new connections.

I have used the Char Davies installation *Osmose* as my primary example because it so aptly covers so many of the implicit issues I have raised here. On my way out of the U.K. I stopped briefly in London to go to the *Serious Games: Art.Interaction.Technology* exhibition at the Barbican Art Gallery where *Osmose* and Bill Seaman's *Passage Sets or One Pulls Pivots at the Tip of the Tongue* were on show, along with works by Toshio Iwai (*Resonance of 4*), Jim Campbell (*Hallucination*) and Harwood (*Rehearsal of Memory*) and others. This is a valuable collection of works which ably covers the spectrum of cyberspace and vr artwork, providing important examples of the kind of work which was discussed at the conference. A wonderful coincidence as best I can tell.

References

Centre for Advanced Inguiry into the Interactive Arts web site [back]

Ackers, S. (1997) "Perception of Individual Time". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p2. [back to 1st reference] [back to 2nd reference]
See also Ackers' web site.

Ascott, R. (1997) "Preface" to Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p1. [back]
See also *Leonardo* article on Ascott.

Baars, Bernard J. (1997) *In the Theater of Consciousness: The Workspace of the Mind*. New York, Oxford Unversity Press. [back]

See also Baars on: Metaphors of Attention and Consciousness.

Davies, Char (1997) "Techne as Poiesis: Seeking Virtual Ground". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p28. [back]

See also *artnetweb* article on Davies' *Osmose*.

Delmotte, Isabelle. (1995) *Epileptograph: The Internal Journey*. Sydney, Artspace. See also her abstract in Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p29. (1997) [back]

See also *Mesh* article on Delmotte's *Epileptograph*.

Dolinsky, Margaret (1997) "Dream Grrrls: a World of Virtual Reality". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p32. (1997) [back]

See also Dolinsky's web site.

Domingues, Diana (1997) "The Desert of Passions and the Technological Soul" In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p33. [back]

see also *Leonardo* article on Domingues, and Domingues' web site.

Eliade, M. (1964) *Shamanism: Archaic Techniques of Ecstacy*. London, Routledge & Kegan Paul. [back]

See also biography of Eliade and further links.

Gigliotti, Carol (1997a) "What is Consciousness For?" In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p40. and full paper in Ascott, R. (ed) *Consciousness Reframed: Conference Proceedings*. CAiiA, University of Wales College, Newport. [back to 1st reference][back to 2nd reference]

See also What Children and Animals Know That We Don't by Gigliotti.

Graham, B. (1996) *Serious Games: Art . Interaction . Technology*. London, Barbican Art Gallery. [back to Davies ref in Sect.1] [back to Seaman ref in Sect.2]
See also Serious Games web site.

Hebb, D.O. (1949) *The Organisation of Behaviour*. New York, Wiley. [back]

See also a web site on Hebb.

Hobson, J. Allen, (1994) *The Chemistry of Conscious States: How the*

Brain Changes its Mind. Boston, Litle Brown & Co. [see also Hobson, J.A. *The Dreaming Brain*] [back to "dreams"] [back to "shamanism"]

Jones, S. (1997) "A note on a possible physiology of subjectivity, and some comments on what a conscious machine might want to do for itself." In Ascott, R. (ed) *Consciousness Reframed: Conference Proceedings*. CAiiA, University of Wales College, Newport. [back] [to full paper]

Nechvatal, Joseph (1997) "Immersive Implications". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p68. [back]
See also Nechvatal's web site.

Pesce, Mark (1997) "Ritual and the Virtual". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p71. [back to virtual space] [back to language]
See also Mark Pesce's web site.

Rogers, Kathleen (1997) "Viperscience". In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p80. [back]
See also Rogers' Viperscience web site.

Seaman, B. (1997) "Emergent Constructions: Re-embodied Intelligence within Recombinant Poetic Networks." preprint of paper for Consciousnes Reframed, 1997 [back]
See also Seaman's web site.

Scott, Jill (1997) "Future Bodies" In Ascott, R. (ed) *Consciousness Reframed: Abstracts*. CAiiA, University of Wales College, Newport. p82. [back]
See also Scott's web site.

return to overall index

Papers by other authors

1: The Hornswoggle Problem.

by **Patricia Smith Churchland**, Department of Philosophy, University of California San Diego, Salk Institute (12 August '96).
Churchland discusses the issue of whether the nature of consciousness is too hard a problem to admit of resolution, with particular reference to David Chalmers' "Hard Problem".

2: Does Consciousness Exist?

by **Dr. Jayant Sharad Vaidya MS DNB**, Academic Department of Surgery, The Royal Marsden Hospital, Fulham Road, London SW3 6JJ U.K.
Vaidya questions whether there is any "thing", an "I", which is able to act with independence. If there is then how does it differ from even the lowliest living things, or should we describe all living things as conscious?

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**

Does Consciousness Exist?

Dr. Jayant Sharad Vaidya
MS DNB
Academic Department of Surgery
The Royal Marsden Hospital
Fulham Road
London SW3 6JJ U.K.

Does consciousness exist?

Life is characterised by growth and reproduction, spontaneous actions and possibly, consciousness.

Let us consider action. Action is spontaneous, as observed by a spectator and willful as observed by the doer. Consider your finger. Hold it in front of you and flex and extend it. You just have to 'will' it to move and it moves. Certain cells in the cerebral motor cortex are stimulated in an orderly manner to set up electrical potentials along the specific nerves that stimulate the muscles to move the finger. Who stimulates the nerve cells or starts the machinery for their systematic actions? No such site which can start a particular action at 'will', through connections to the whole brain like an all-covering blanket, has been identified.

Look at it another way. An infant's brain is developed, according to the instructions in the genetic code, with individual variations according to the parents' genes and the environment at conception and in utero. Apart from its other functions, the brain is programmed to learn. Some learn better than others. As the child grows up, the neural networks keep growing, making new connections, inhibiting some old ones, at each instant, modifying itself by the results of its previous actions and environmental events. The learning and actions continue, and a complex reflex, in a brain programmed to learn and respond, starts expressing itself as an individual. The adult brain is in such a state of dynamic servo-electro-chemical activity ready to respond, as if on its own, to a stimulus. Each response or action being invoked by the combination of environmental events and the dynamic status of the programmed brain. In the example given above, the finger was moved in response to reading this article, or in response to the presence of another person you think would respond in some favorable way if you moved the finger, or, if you are alone, in response to loneliness, and previous events.

Then the 'I' as in "I moved the finger" and "willful spontaneous action" lose their meaning, and we wonder whether our every action is a result of a complex reflex. The actions of an embryo are usually considered to be instinctive. We don't usually ascribe the actions of a 20 week embryo, to its own will. Does consciousness enter the embryo around this time or at conception or at birth or even later? Does it 'enter' at all? Or it is there all the time? Or is it never there?

Suppose we program a computer to learn by giving itself random numerical problems and learn through its answers, and laugh (crackle) whenever it gets the right answers. It is difficult to ascribe the ability to solve problems and laugh to some thought and pleasurable feeling inside the 'machine'. "The computer does what it is programmed to do: it cannot do something on its own!". Can any of us really do anything on our own? As we have said before, spontaneity of action is difficult to defend. Realize that we can be as certain about a computer's feelings as of another human being's. There is no way to really tell the difference.

However, each of us feels within ourselves the existence of an 'I' - I feel, I think, I wish, I walk, I talk, I believe, I act, etc. This can be called I, ego, consciousness; The Indian philosophy classifies it into a hierarchy of three classes, viz., mind, intelligence and soul (Mana, Buddhi, Aatmaa). Is this 'I' only for convenience? It is indeed difficult to deny the existence of 'I'.

What about our emotions of fear, anger, pride, happiness, jealousy, love, sadness, lust, ambition? Some emotions are accompanied by secretion of specific neurotransmitters in the brain. But who perceives the emotion and thought? And where? Are they all non-existent? Just because we cannot measure them or we have yet to develop instruments which can detect them objectively? What about thoughts, new ideas, abstract concepts? The responses accompanied by these emotions and the actions generated by the thoughts are objectively seen but these responses and actions cannot automatically prove the existence of the emotions and thoughts. When you tap 5 x 4 on your pocket calculator and press =, the calculator gives a response, 20; does it automatically mean that it had felt happy though it is not programmed to say "ha ha" ?

We cannot go in circles disproving consciousness by questioning spontaneity of action and disproving spontaneity of action by questioning the presence of consciousness. We must realize that the proof of existence of this 'I' or consciousness is based solely on our personal subjective experience and the assumption that since I am feeling this 'I' it must be present in all those similar to me. Does your dog have consciousness? Yes.. because it is similar to you - it walks, eats, barks, and apparently at its own will. Is it the same with an ant? And what about an amoeba. It also moves, apparently on its own. Finds its prey, eats, reproduces etc. It also must be having an 'I'. Maybe this 'I' is very simple and small. But an amoeba does not have a brain. Is brain essential for existence of consciousness? Do bacteria have consciousness? and do viruses too have an 'I' ?

Natural sleep or induced anesthesia: Even though up to 25% of the sleep time may be occupied by dreams, the remaining time we are not conscious of ourselves. When we wake up, we have the same consciousness as before. Where does the consciousness go during this time. Does the brain get reversibly disconnected from the 'I' ? Is waking up similar to booting a computer which comes 'alive' with the same memory and software as before?

Electro-chemical activity: In psychiatry, disorders of thought and emotion can be treated with drugs. Drug addicts change their thought content and emotions when they are intoxicated. Does it not mean that physical and chemical compounds can alter the yet abstract emotional and thought content of our mind? In psychosis, there is derangement of perception of self and thought content of the person. And this can be treated by drugs with varying degree of success. These drugs interact with the mind and can alter it. Is this 'I' itself accessible to tangible substances? Or do these 'tangible' substances also have their own 'I's which interact with your 'I'. Is 'I' an emergent property of the electrochemical activities of the brain and not separate from the body?

The privacy of consciousness: The color which you call red, is the color which I call red but we both may not be perceiving it the same way. There is no way in which one person can communicate with another what one feels except by representations which are, at best, only crude images of the perceptions.

The non-destructiveness of consciousness: If one of your fingers is cut off, you don't feel any decrease in your 'I'. Even if an arm or a leg is cut away, there will be a loss of body image, but no reduction in the amount of 'I'; The feeling and thinking will be by the same complete 'I'. The seat of feelings could not be the heart since heart transplant recipients do not have the personality of the donor nor are patients on artificial hearts 'feeling-less'. If the whole body below your neck is cut off, and head is kept alive with an artificial heart, lung, kidney, and alimentation, the person, would probably continue to think and feel the complete 'I'. He would still communicate with us through movement of eyes, or with an artificial larynx, by talking. Only when the brain is cut off, or non-functional, we assume that 'I' ceases. Is it that, even then 'I' is complete and that it cannot communicate with us? Is brain the residence of consciousness or is it a communication center of consciousness to the rest of the world.

Imagine the brain to be a communication center. Like a dish antenna. The more complex the dish, the more channels it can receive. The transmission is continuous and ongoing. It only takes a better brain to receive it and communicate. A simpler dog's brain acts 'spontaneously' and does a few things, a chimpanzee does much more. An ant does much less. An amoeba even less, since it does not have the sophisticated machinery to receive other channels. It is like comparing a single channel, mono, black and white TV with, a 69 channel stereo holographic TV. Now imagine the whole organism to be such a telereceptor. Each is like a television screen. But instead of being only a screen which shows visual pictures, it projects the whole image complete with all the accessories. The more complex the organism the more versatile its actions and 'thoughts'. The question then comes of who is transmitting? Is someone transmitting at all? Or is it just a play of chance and reflexes?

There seem to arise two mutually exclusive basic governing principles of the universe.
1) the divinity principle which assumes that the whole universe is a result of a design...by God; it is assumed that destiny of the universe and every being in it is decided in advance and is unchangeable. If this I true, there is no role of individual 'I's since they only serve to camouflage the all pervading governing power, themselves not having any will.
2) the principle of causality and chance which to some extent is based on Darwin's theory of natural selection and evolution. This assumes that the whole universe and every being in it has arisen out of random occurrence of events allowing the survival of the fittest. There is no designer God in this principle. However, in the extrapolation of this theory to explanation of consciousness would mean what I referred to earlier: that all supposedly willful actions are actually a result of a very complex reflex.

It is difficult to prove or disprove any of the above theories though we may be on the verge of proving the latter. However, in both of them there is little room for independency and capacity of willful action of 'I'.

The definition of life: Life has been traditionally defined as anything which has a capacity to grow and reproduce. This is the reason why plants have been included as alive; (there are some circumstantial evidences which also show that plants can perceive and think). Viruses just miss to be qualified to be called alive since they require some other life forms to help them reproduce. But they reproduce anyway and that brings them to the borderline. Now we have found Prions: small proteins which can reproduce and could cause havoc (BSE & CJD !) even amongst the most 'superior' species.

For many decades it has been possible to produce identical objects with the help of machines. Today a computer can do almost anything that most living objects can do. A robot can see, hear, have tactile sensations, move, obey taught commands and do most actions which a trained dog can do. Though it may be difficult for the silicon technology to reach the miniaturization of storing information that nucleic acids have achieved in nature, it is theoretically possible to program a computer to reproduce itself, given all ingredients. Will it then qualify to be called alive?

Again, growth and reproduction cannot be the criteria for being alive. Many cells in individual organisms have lost the capacity to grow or reproduce, like the nerve cells. But they are still functional. And it would be absurd to call most of your brain dead because it cannot reproduce.

Can the ability to move spontaneously by itself bestow the label of being alive? A watch moves by itself until it dies when the battery runs out. The earth and all the planets move spontaneously. Shouldn't we call them all alive?

The common factor in all living beings, as classified today, is presence of nucleic acids arranged in chains (DNA and RNA). Does the presence of nucleic acid alone make an object alive? Does it endow consciousness automatically? Or is it its programmability? What is there in nucleic acids that is not there in every other object we see around us, a table, chair, watch, telephone, hydrochloric acid, water and even space, fire, light, breeze, sound. Why could they not be having consciousness. Inasmuch as we cannot deny the presence of consciousness in ourselves, we cannot refute its omnipresence.

Since times immemorial, our sense organs could allow us to ascertain the presence of light, sound, mass, smell and taste. We have since developed instruments to ascertain the presence of, and measure electricity, electromagnetic radiation, gravity and nuclear forces. Before these parameters and instruments to measure them were developed, their presence could never be proved. It could be only theoretically deduced. Even today the quantification of smell and taste is still in the experimental stage. As we learn more about the chemistry of thought and mechanisms of neural networks, we may develop parameters to measure consciousness and life. We may realize that objective communication, ingrained in the definition of science, may be limiting its progress.

As of today, concepts of spontaneous action, consciousness and life are indeed based on very weak grounds and we could conclude that if you are sure about yourself being alive and conscious, then every material object qualifies to be called alive and conscious, differing only in the quantity of liveliness and quality of consciousness; and vice versa.

Jayant S. Vaidya

email: j.vaidya@ucl.ac.uk

 return to **Philosophy index**	 return to **overall index**	 return to **Introduction and textual index**

Notes on the Cybernetics of Language and Video

by Stephen Jones.

This is an HTML version of a small book self-published in Sydney, August, 1979. Much of what it offers is still relevant to discussions of the mind today.

FOREWORD

The material within consists more or less of notes about certain aspects of the inter-relationships of language, knowledge and social being.

The cybernetic nature of the world and the interdependence of all systems in the world has only recently entered into the discourse of the social sciences of the West. There is a major lack, in the language we use, of means for describing and dealing with this interactiveness and the processes of change inherent in all eco-systems. Without the words to talk about change and system interactiveness we are effectively prevented from knowing of these aspects of the world in which we have our being. We do not see ourselves in inter-relationship, we do not recognise the contingency of the systems containing us, until we see the inter-relatedness of our social structures as systems of relations.

We exist within a social framework which has a myth structure interactive with the history of society. We are sociaiised into this framework as we become members of society. Our parents, our learning language and our education along with that ever so prominent purveyor of information, roles and attitudes the media, in particular television, provide this overall socialisation.

I would like to thank Ruth Waller for critical input and Nell Smith for the typesetting. Also I would like to invite any feedback you may have. Any correspondence should be addressed to me at

email: sjones@merlin.com.au

COMMUNICATIONS and the SOCIAL CONTEXT

The mind and its content form the context within which all incoming information is interpreted.

What we see is understood in terms of what we already know, we have each a unique history – mediated socially via language – against which we compare all incoming data – the re- dundancy of the system might be as high as 90% so that we develop apparently constant ap- pearances over a longish timebase. A building tends to stay put, as does the language, weathering only slowly over time. We perceive something new because we see a difference – it might be a new case of something we already have a name for, or it might be something that is as yet nameless (noting that all experience is in some way perceived).

A name implies a history – at its emergence a new (technological) medium lacks history, its potential content as yet unrealised. In practical terms the appearance of a new medium is governed largely by an activity of pushing old content from previous media through a new channel: to the degredation of both the old content and the new medium. Witness the cinema on T.V.

So video as a new medium appears, its properties are not explored, it becomes broadcast televison: one way and serving the specific needs of the broadcaster only. We use it in a manner we already know: centralised distribution of information, all the more carefully packaged through the

3 images from *Eva* (1978)

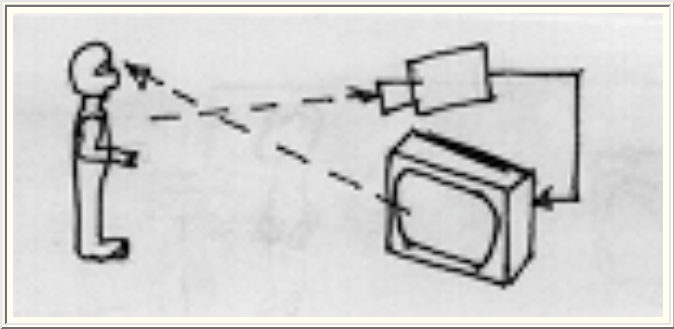
expense of the production process.

We neglect those aspects of the medium which would be the most humanly valuable. These properties of video which I am suggesting we should be exploring, are that it is

> 1: a two way, "realtime", instantaneous and simultaneous process through time;
> 2: a mirror on ourselves, both realtime and at one or several removes, on later replay;
> 3: a communication with ouselves and with others, with the feedback function of the system setting up a learning situation.

For example: when used in the realtime mode, camera directly linked to monitor we see and modify our behaviour according to the feedback we receive from the display.

The camera looks at you from over there (part of you but separate) it relays an image on to the video monitor screen, you can then look, or not, each is still a response to this visual in- formation loop providing feedback to you about your appearance, behaviour and relationship to the context provided by whatever space this system occupies.



This response activity is a process displaying communicative form in that there is a feedback of information, from another viewpoint, about oneself. We are in an information loop, not in a one way situation, but in a two way, action and response, situation; in this case with one's image on the screen, an extended self, a part of you but outside of you. (Of course, why do we think of it as outside of us at all, or anything else "out there" for that matter. It is in one's perception that these things have their existence, and that is our perception within our organism/mind. The objective world is an appearance only shaped by our organism/mind, social being and language.)

This information/feedback loop is a process the relationship between the elements of which changes constantly, in the present. We behave and watch our behaviour, we feedback off ourselves and by this trick of separation we enter into communication with ourselves. This form is analogous to the more general communicative form of a conversation between two people acting and responding in concert.

There is no possible dialogue with T.V. but video as a technology is perfectly capable of being responsive – i.e. of being used in a dialectical manner.

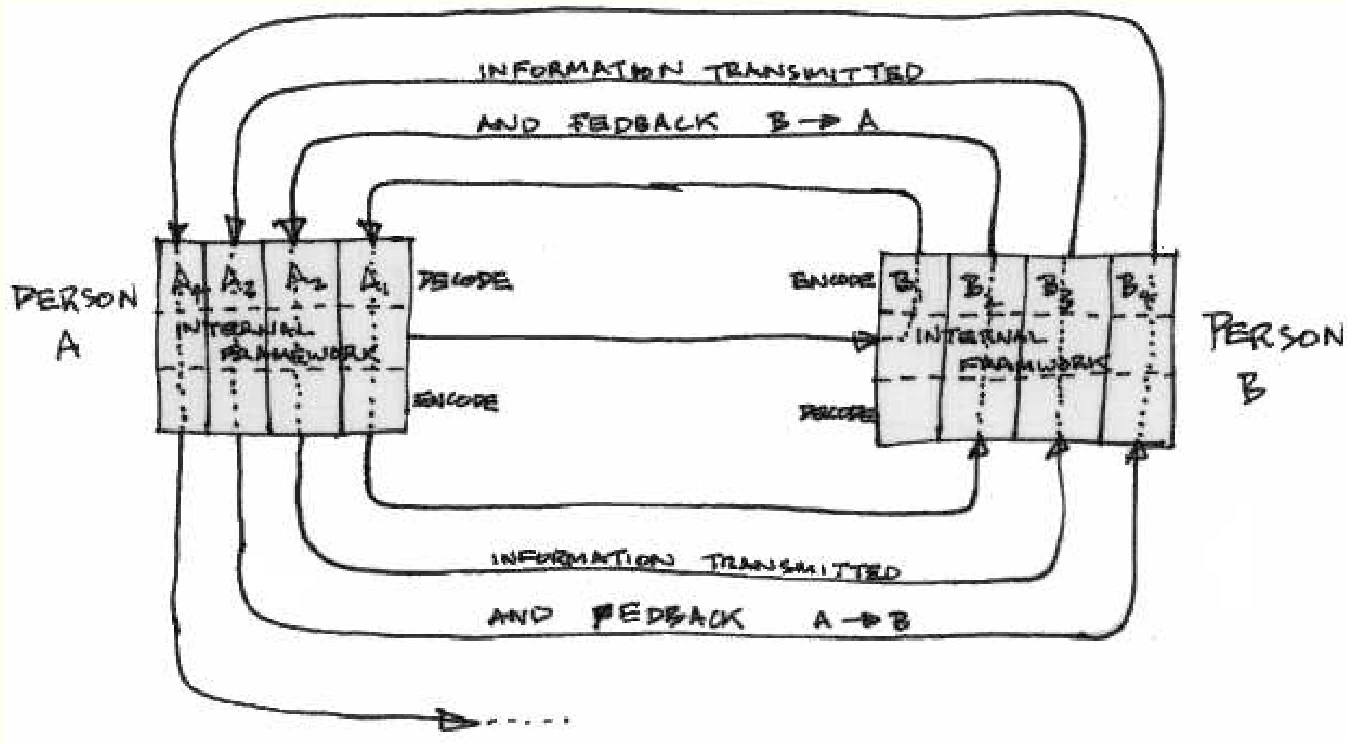
For example on another scale, we have the teleconferencing link-up where (usually) business people and others are able to link up with their offices and others', in a fully two way video/audio link-up and carry out a full conference with cameras, cables and monitors interfacing them.

People within a social milieu become part of a system of interrelating entities. We can describe physical boundaries: that is the walls and floor of the room for example. We can describe the social limits of a particular system/group in an agreed 'conventional' rnanner. (And ultimately all boundaries are only conventional even the physical.) This system exists over time and undergoes change as energy and information pass through it. People's interactions and communications provide some of the energy and information within the system (we are looking at the informational aspects of the system, i.e. communications). As the system of communications develops we may observe the processes operating within the sub-systems of one-to-one and one-to-n relationships developing between and among the participants in the system. (Much art has already discussed this kind of problem but not especially within the video rnedium.)

These interactions show up the network of relationships within the system. The properties of networks provide the rneans for the feedback of information and the control of the comminicative processes' development. You can ignore the feedback potential, as in broadcast T.V., if you choose, but if you're at the bottom of the hierarchy (as is the audience) nobody cares anyway. If you're at the top of the hierarchy: the T.V. station; then you've already got the control by avoiding the possibility of feedback completely. The only feedback the T.V. station looks for is the financial feedback from the advertisers. Feedback and control go hand in hand.

In small groups, in regions, in communities the networks of organisations and individuals are the sources and the media of information flow through the various interlocking systems. These systems display processes of interaction over time and by memory and documentation develop a history. Upon the history of this system is all the new information and meaning predicated. Interlocking systems each with their own histories provide the structure of society. The avoidance of change despite all this is a process of institutionalisation and in the semantic world myth provides those apparently natural conditions of relations between people which are the content of the institutions (of, for example, the family).

All information flowing through social systems is in some sense communications. Of course rnost communications are mediated by various technologies. The individual face-to-face conversation and small groups are among the very few communication structures which are not mediated via some technical system.



Conversation develops through time.
Information is transmitted by A, encoded into language. B decodes and compares with existing internal cognitive framework for interpretation, and then adds own information to the conversation system as response. Thus the information in the system accumulates...A's output modulated by B's response/feedback, and vice versa, creating a dynamic cyclical development as long as further response is generated.

Hall and Hopkins (in Studio International, May-June1976, p. 262) describe the communications process thus:

> people "exist in a dynamic relation with each other and their contexts. The interrelation of their contexts (including the material conditions of life) metaprogramme and expectations causes them to engage in activities which are intended to satisfy needs as perceived by the people themselves. The activities result in achievements which in turn modify the metaprogrammes, act on the contexts and generate new expectations. This can be better understood as an ongoing, dynamic and cyclical process rather than a linear chain of events" (the metaprogramme is a "set of instructions, descriptions and means of control of sets of programmes" ibid.)

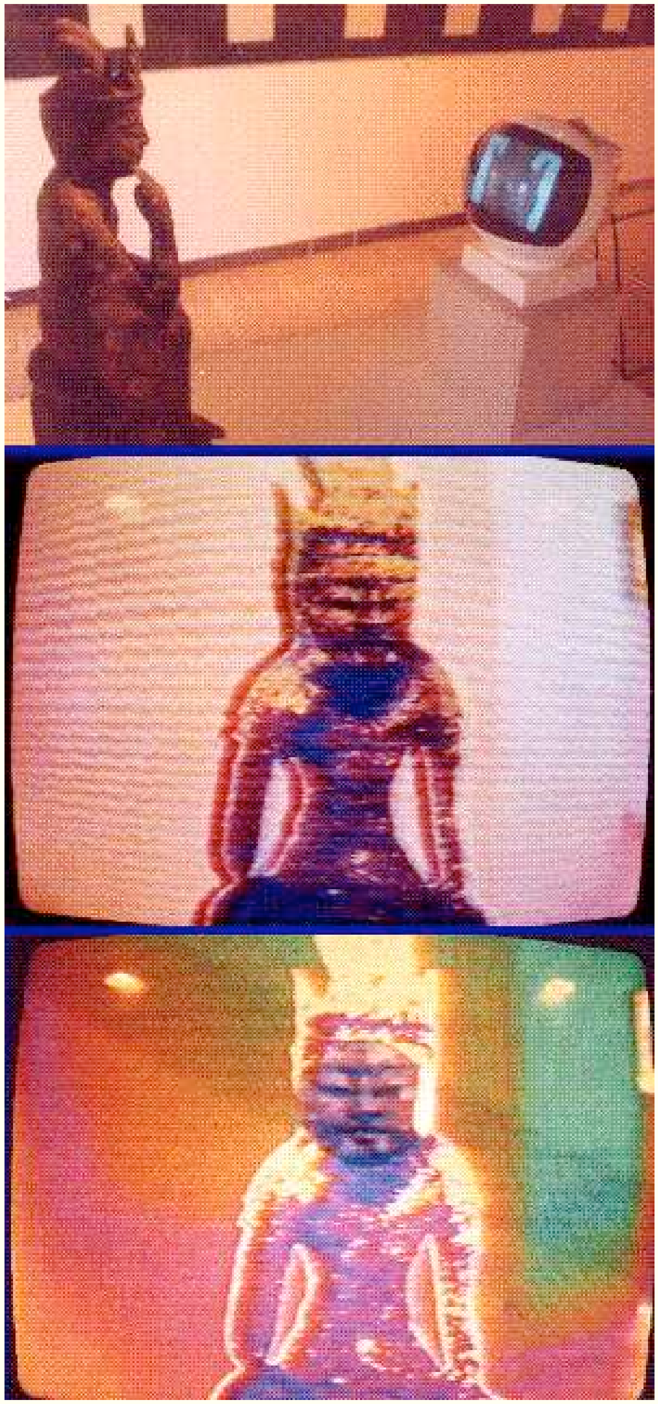
Because people's needs are not precisely the same from individual to individual each person will feed into any operating communications networks from different points of view, with different priorities, etc., and it is from these different weightings on the system that much of the energy driving the system derives. Also it is this intersubjectivity which brings complexity and renders these systems indescribable within conventional linear notions and 1anguage. A descriptive system based on complex causal chains is being developed known as General SystemsTheory.

All these patterns of relationships; feedback, etc. appear at all levels of the world we know, in the biochemistry of organisms and the ecology of the biosphere, in psychological and social activity and social structures. These patterns also appear in the technological sphere (with language as the prime social technology) in communications and technical control systems. Video models these patterns of feedback at many levels of their appearance, in particular at the social and technical levels.

Semantics has to do with the function of meaning in language. Information and meaning is a function of the context in which it is found, so that the same image in two different contexts will have two entirely separate meanings. If you're working with images and you're working in such a way that as the images are appearing in various contexts at various times, in random or structured order, then the way the image works, the meaningfulness of that image at various times is going to be changing. (Where the signification of the image becomes static despite changing contexts we have the appearance of mythic info. in the sense of Barthes.)

LANGUAGE and OBJECTIVITY

I want to show here that the concept of objectivity is misleading, and therefore, so is the concept of subjectivity. The argument is based on Wittgenstein's 'private language argument' coupled with Heisenberg's 'principle of uncertainty'. Wittgenstein argues that all our knowledge of the world, all that we think about, talk about and so on, is gained out of experience, interaction and action, i.e. is gained out of linguistic processes within the environment. One can have no existence as a (social) human being outside of a social framework and the social



Installation photo of
Nam June Paik's TV Buddha
at Art Gallery of N.S.W.
(Sydney, Australia), April 1976
and 2 images from my video
TV-Buddha: Homage to Nam June Paik
(finished 1978)

framework is constantly mediated by shared language and that in the long run that there can be no private language that would have any communicative function. My understanding of myself, i.e. my identity, is disclosed by my interactions. I do not exist as an island – 'no one is an island'.

Heisenberg, in attempting to explain certain curious results in physics, recognised that the tools being used to make observations in the subatomic field were in fact having a major effect on events in that field and were thus giving misleading results. The presence of the observer in the field effects the events being observed. This, of course is obviously true in social behaviour, though it is only in recent years that the social sciences have begun to accept it. The uncertainty principle means that one cannot enter an environment without effecting it in some way. Effect is interaction, or at least will lead to interaction or the avoidance of interaction which in itself is a kind of interaction. To reach some sort of state of 'objectivity' within an environment is impossible. Language proposes the illusion of objectivity merely as an abstraction. The failure to recognise this illusory nature of language will inevitably lead to bad results. The corollary is that as 'subjectivity' and 'objectivity' are linguistically opposed, the disappearance of objectivity implies the similar disappearance of the concept of subjectivity.

This argument suggests that one cannot gain information from a system without entering into it and acting upon it, in some way altering the conditions of the system, if only by talking about it.

There is no subjectivity or objectivity. The presence of the observer within the environment being observed, or even outside the environrnent being observed, alters the processes going on in in that environment. (Heisenberg's principle of uncertainty).

The subjectivity/objectivity illusion exposes one of the major problems of language; that language allows for the appearance of concepts that in no way reflect the reality of things. Another problem related to this one is that language, as we use it, does not allow for an adequate description of things in process but tends to objectify and make static. So that we consider persons, things and events, normally in process, as static unchanging objects to be maintained in this static condition. Thus the fear of change so inherent in much of society. We employ no language with which to accommodate change. A further consequence of this can be seen, for example, in the objectification of women in advertising. Of course, these comments refer largely to the western language, English in particular. Our language doesn't really effectively deal with processes. In attempting to describe what's going on language tends to make things static, it holds and pins things down and defines them.

The processes of a social framework are basically linguistic, and display certain characteristics such that the communicative function goes something like this: A makes an action and talks to B, B listens and responds in some way to A, who can continue to develop the conversation by further response. It is not A acting and then B acting in isolation from what A has done, unless you purposefully go about that, which is the *Absurd* thing. It is A acting, B responding and A responding to B's response and so on in a developing loop of interaction. So that you get feedback operating, B feeding back to A who feeds back to B, constantly modulating each other as long as the conversation lasts. A to B and back again. I am defined by your response to me (i.e. by your cumulative response).

If B chooses not to respond, this applies an inhibitive feedback on to the conversation loop, which will then, in the normal course of events, probably die. l.e. the inhibited response (or no response at all) is just another response from which one derives information and further modulates one's communicative activity. So all response, whether it is empty or very full, in-fluences how you carry on, you can stop talking with each other, you can go away; or you can, if it's a positive or full response, then continue further into the conversation. Positive and negative feedback modulate the flow of the interaction loop.

The communicative function, which is also language, is governed by interactive processes. I am defined by the way that I relate to you and by your response to me. I become the synergy (the synthesis) of all responses in all situations constantly updating.

This results from the way that we become human; get socialised. The way we know about the world, the whole business of cognition and perception is underpinned and continuously modified by what we already know of the world. Because we cannot know about something until we've named it. We don't know what it is until we talk about it. All knowledge of the world is socially mediated via language.

The social set from which we perceive, i.e. our framework, also governs the way in which what is perceived is interpreted. We live within an ongoing condition of intersubjectivity with our physical and social environs. No two people can ever occupy the same space at the same time, consequently we all have different histories, different consciousness of ourselves and others. This history and consciousness is our framework from within which we relate to all phenomena, information, other people, social structures and institutions. As we are contained within various social structures we become part of others' consciousness. What each of us does and says influences the way the other sees us. The way you see me is a function of your framework plus the interaction we have had or have avoided. Similarly the way I see you is a function of my framework plus our interactions. I exist within you as a perception and a name and the set of linguistic artefacts (some active, some mythic), contained within your general framework for functioning within society. Likewise are you perceived and defined within my framework. This operation generalises out into all social relations.

All these things occur over time and are therefore undergoing constant renewal and change. The redundancy or repeating elements of this communicative system provide some sense of continuity.

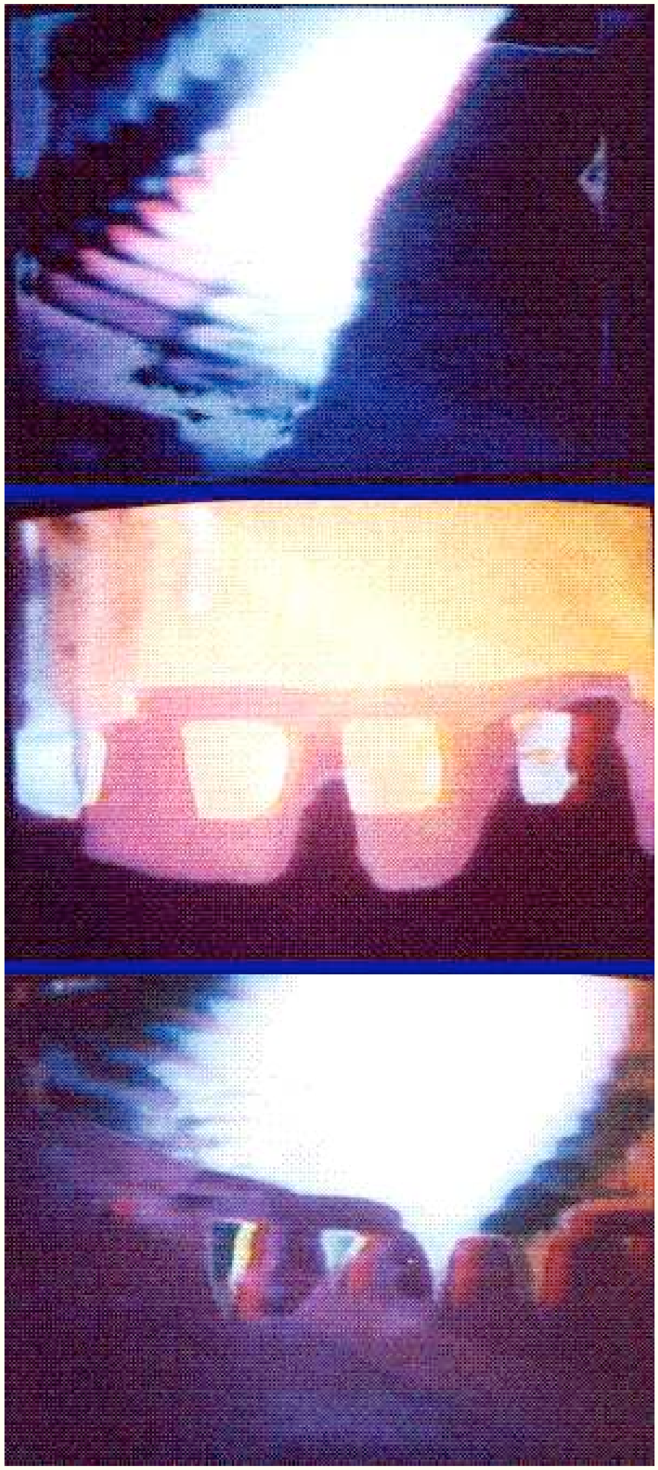
We exist within an intersubjective process that is the meta-framework of society, a dynamic and complex web or network of interaction and relationship.

MYTH

The myths which accrue around a social structure provide much of the framework, or the context, within which the apparent givens (the institutions, etc.) of the society are couched.

Roland Barthes, in Mythologies, proposes that there are in the environment; the social, cultural, built/natural environment, images which have become the cultural archetypes of our world, our society. These images are the mythic images, the conspicuous signs of what is normal in social behaviour.

Myth informs the basic frameworks within which we interpret the goings on around us. Myth forms the frarneworks from which we perceive and moulds the way we work and the roles we take. We are governed and constrained by these images because we are in various ways forced into the roles



3 images from
Stonehenge: Systems Interfacing Reports
(1975-78)

presented by these image structures. We are offered no other information in the presentation of the myth so that we then have no critical grasp of the wider context in which that myth operates. Advertising is the major purveyor of mythic imagery with our only proposed role that of the consumer.

The political myth that we must preserve our social institutions at all costs is well served by a language structure which hides the inherent changeableness of things, leaving people to believe that these institutions have always been there, as they are the "natural" forms. Fred Flintstone and Star Wars assure us that society has been like it is now from the far distant past into the far distant galactic future. Myth sets up the historical so that it is perceived as natural.

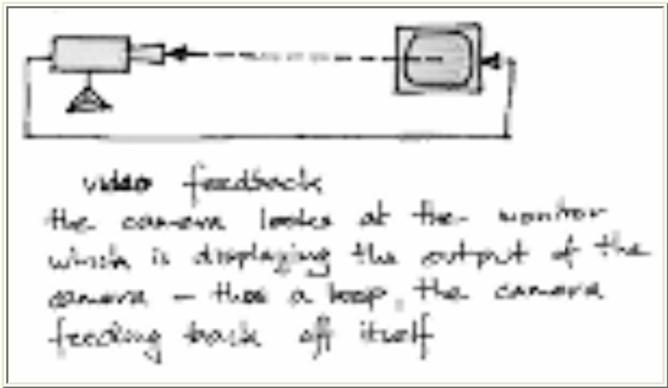
History becomes a consideration of states, rather than an examination of how things change over time. In seeking a language which enables us to understand the nature of open structures in time, the structuring of time, the structuring of knowledge through time, history, the function of history, etc., we must develop a language of processes. This language needs to acknowledge the inherent but hidden contingency of the world, so that the means of understanding, generating and accommodating change are brought forth.

FEEDBACK

Social feedback... Bio-feedback has direct biological consequence, whereas social feedback has social consequence rather than direct biological consequence. Recognizing that the intergration between biological and social action is very strong, and they cannot really be separated, it is only a trick of language that allows us to make a fragmentation of this sort.

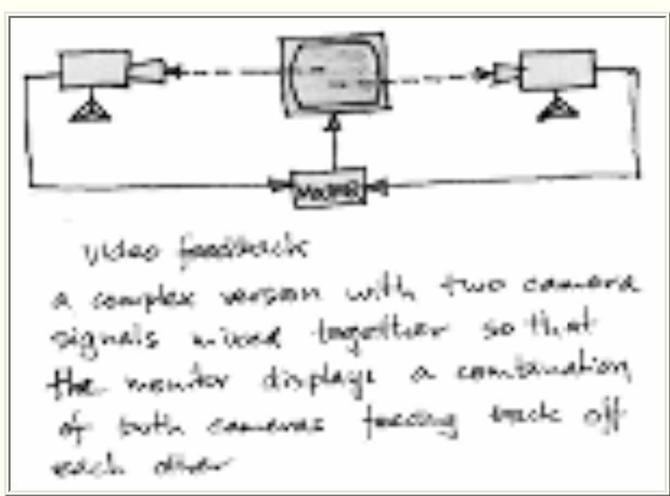
There are many classes of feedback in society and it is to do with the cybernetic process and the action of inhibition and control in social processes... or in the action of feedback as inhibitive or positive control of a process... or in a self-corrective process itself. Inhibitive in the sense that if you're an actor working/rehearsing in front of a camera so that you can see what's happening, you then can correct your actions as an actor, according to the things you see in the video, so that you will inhibit the mistakes and enhance the quality of the thing in some way.

Video feedback... One can take this approach to video, the feedback function, to advantage within the context of performance or installation where the data is processed through a feedback video environment, feeding back to the audience information about their own actions... where they can control the actions and feedback off the monitor and camera relationship and one's relationship to these things starting to do things within it creatively.



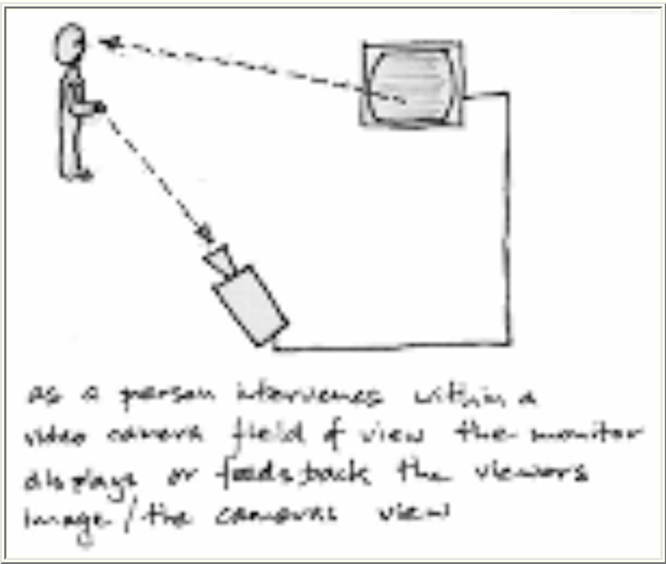
So we begin to see feedback as a system process, a class of action within processes, having a lot of levels of operation.

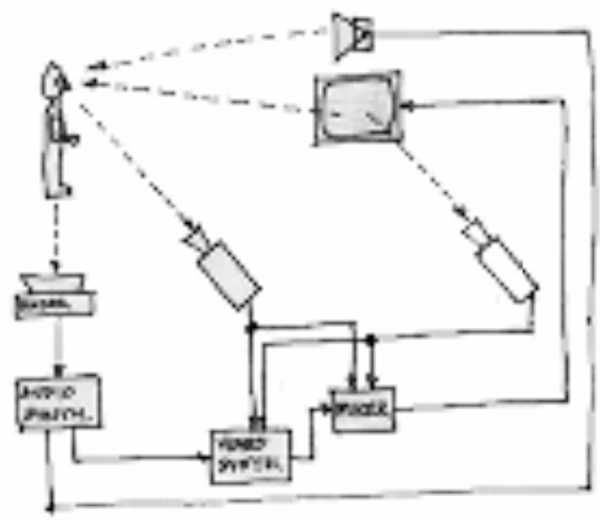
Though it has a direct electronic one (endemic to video), it has the macro scale social function (video access, video as a means of social facilitation) and intermediate functions as in installations, etc. Obviously feedback is a generalised function, not at all restricted to video, and has a great many areas in which it manifests, e.g. the conversation, all the other processes of social interaction. Feedback is the action that takes place in all processes of interaction within a state, or within an existence, or within an ecology of some



sort; a social ecology, a techno-ecology, a biological ecology or any self-corrective system.

The thing about video here is that we find a means for modelling many of the processes of consciousness. Feedback is the response of the context to an output from the consciousness to which the consciousness responds becoming a process in action, live, containing and maintaining itself, open-ended, subject to change, and non-conservative. **PROCESS**. So via its nature as an analogue of consciousness – no matter how partially that nature is revealed – video becomes a tool that allows one to operate on the processes of cognition and one's environmnent and the social relations within that environment.





a more complex interactive feedback environment using doppler shift radar (as movement detector) and video/audio synthesisers — for performance situations

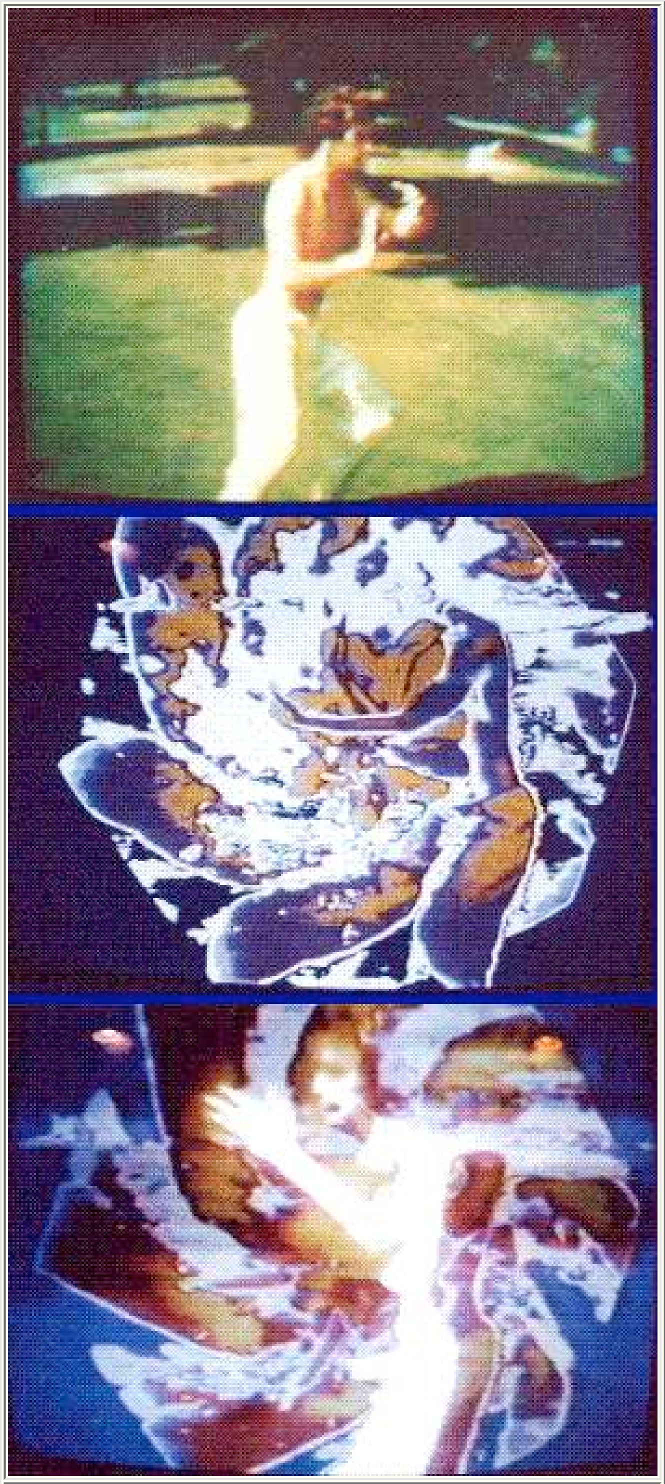
---

PROCESS and KNOWING

One of the aspects of processes that we must deal with is the nature of change itself. To the Taoist there is nothing constant in the universe excepting change itself. We use language which tends to negate change, or at least slow it down. We are looking for a language of process.

Our language is not at all strictly verbal. I couldn't talk to you if I didn't move about and gesture as I talk. These are the peripheral activities of proxemics and body language. Communication is broadened by these deeper socially structured aspects and we get our first hint of the intuitive here.

We are trying to get to a language which enables us to understand open structures in time, the structuring of knowledge through time, history, the function of history, and all those sorts of things.



3 images from
Tai Chi Transforms
(1975-8)

The Chinese Taoist view (as expressed in The Classic of Change, The Classic on the Tao, Chuang Tzu, etc.) seems to be something like this:

> The important thing one has to do is to not grasp for the information. To refrain from directing: to allow the processes that are involved to occur and go on as they will and to accommodate to those processes and to therefore become carried along in them and realise

the work
as it
should
be rather
than to
try and
push the
environment
into a
situation
in which
it is not
willing
to go
perhaps...
it's the
line of
least
resistance...
and that's
a
cybernetic
process.
Yet one
must
remain
entirely
detached
and
entirely
critical.

Of course current socio-political conditions make thisapproach somewhat idealistic. But this kind of approach comes close to exposing some of the aspects of the intuitive mode of knowing.

I consider intuition to be a mode of knowing wherein phenomena are considered within the context of their framework and the complexity of relationships entailed in the structures underpinning the presence of the phenomena.

This is generally done without the mediation of verbal language, so that the problems of categorization and fragmentation are avoided. But intuitive activity takes place usually below the threshold of everyday consciousness and the data gained is consequently very difficult to access. So we need to develop other tools for exposing things in process.

Video may be a tool of this kind. Perhaps we can use video; with its feedback, simultaneity, real-time control/response (responsiveness), in such a manner as to expose processes as changing, interactive, interdependent.

Perhaps video with its footing in social structures as well as other more abstract structures will be interactive enough across structural levels to be a general tool for examining processes at many levels within many kinds of situations.

What I am doing in the studio and in my video work in general is basically experimenting, I'm playing... because I don't feel that I know the answers to these things that I am proposing. There are processes upon processes by which one develops information, messages, and one can then make actions which generate feedback in one form or another. (And that is the function of exhibiting the work one does.) The function of what I am discussing is that it provides a framework from which I can operate upon everything that I am doing.

VIDEO

Video (form) is formed and constrained by the nature of the technology itself. The low resolution, the 4 x 3 format, the ephemeral nature of the tape, the requirement for a machine for replay, etc., all contribute to a particular set of factors which make video (art) unique.

The particular qualities of the medium always will determine how the content is affected when worked through that medium. In video, we have, very much for the first time in the use of technology in the arts, a recognition of the conditioning factors of the medium, and the employment of those factors in and of themselves to code, process and transmit information. (By medium, here, I refer to the actual means of production of software, i.e. the hardware, the equipment). We also see a situation in which the technology itself; the hardware, provides a source of content in itself alone.

The video synthesiser appears very early in the history of video art and though perhaps, to some, inadequate as a source of content ("one should use material that has relevance to some set of social conditions") as a processor of images and coupled with other sources becomes expansion of parameters of visual imagery. Video allows us to make a variety of hypothesis about communications within the visual/auditory modes and to test these out. The simultaneity and instantaneity of response and feedback leave us free to modify as we go clarifying issues during the process i.e. we have real time control over the process.

I find that in my own work, that what develops out of the formalistic aspects of the medium; the significatory function operates off a symbolic, 'mythological' datum which is then, via the aid of video processing, worked upon in such a way as

First video synth
built at Side FX
(Sydney, Australia), 1978
Images from the *SPK* video

to expose syntactical relationships which may be unique to video, and are certainly outside of the usual framework of language; i.e. non-verbal, multilayered, etc.

Thus there develops, through the inextricable relationship between form and content, new semantic relationships depending on the video context for their transmission. The communication might be idealised as the manner in which an alien being might communicate in a non-specific language situation.

We want to encode certain information about something, but the structures imposed by available verbal languages are inappropriate to the information and consequently degrade it in the basic encoding, so we look for other ways of encoding information which will solve the problems of the particular kinds of information degradation (noise) being imposed in the verbal-linguistic encoding. Of course we add all kinds of other noise factors inherent in whatever coding system we use whether it be software dependent (a conventional language in some sense) or derives its syntactic/semantic activity from purely formal manouvres.

Much of this so-called 'noise' is a function of patterns of structure and myth inherent in the languages used, verbal and social.

So, I am dealing with a semiotic the source for which is in the hardware itself and the new arrangements and juxtapostions of image/information which this hardware makes possible.

INTERACTIVE PERFORMANCE

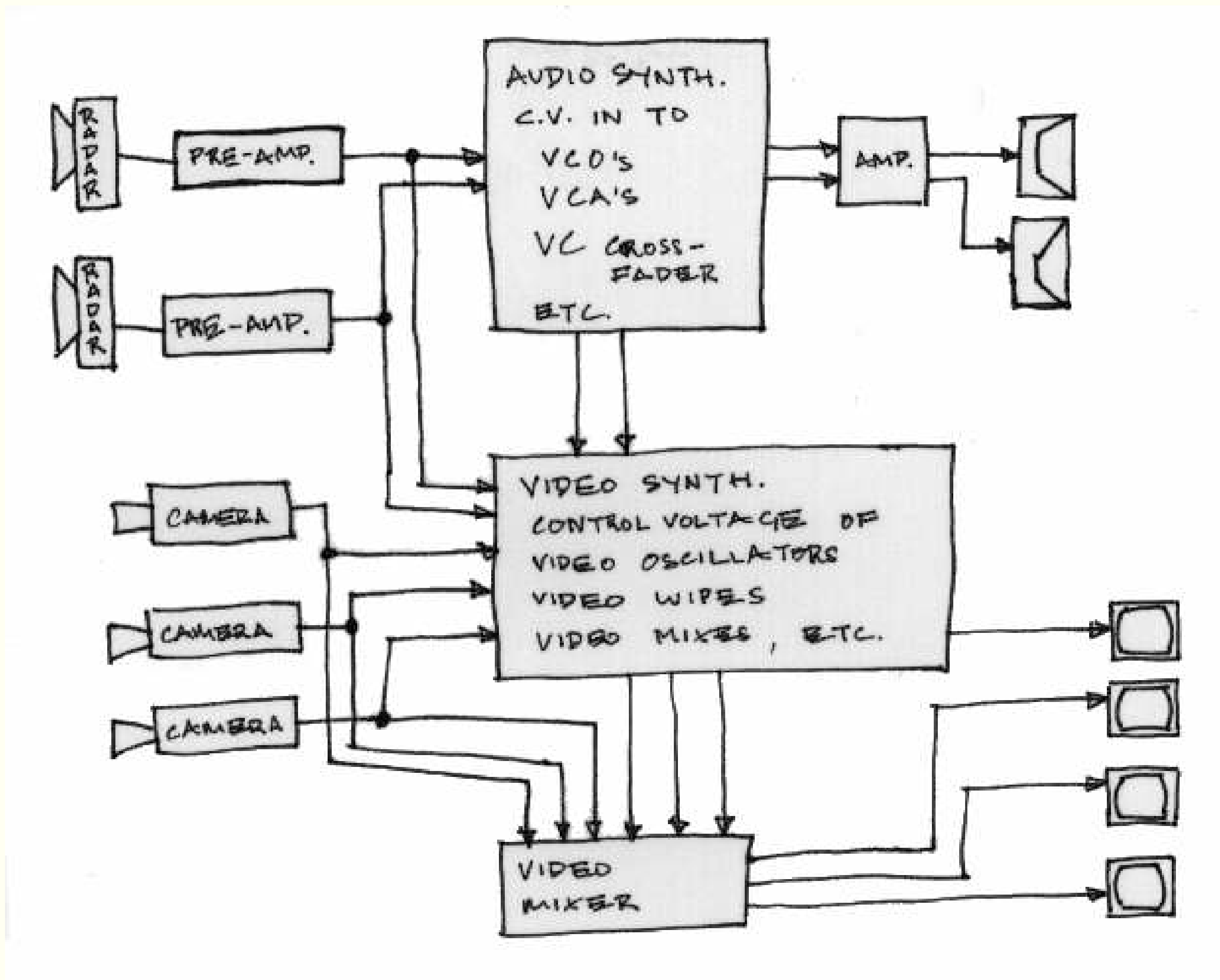
The performer initiates an action ... the detector will detect this action and translate it into an electronic signal varying in frequency and amplitude with the performers actions. This signal can then be (a) fedback live to the performer if in the audio range, (b) used as a control signal in an audio or video synthesiser to alter the sound and images in a manner proportional to the movements of the performer.

The performer can then respond to, and learn to exercise a degree of control over, the sound and images (mixes) produced. l.e. the performer, by the functioning of the feedback loop generated in this system, can learn to play the system as though it were a musical instrument. But the performer need not be encumbered with direct contact, the remote detectors allowing free movement in space.

A corollary of this model is demonstrated in the approach I have dubbed "interactive sculpture", in which some arrangernent of video equipment forms a system in which the observer is enveloped in such a way as every action sets up some kind of response within the video system.

The observers' movement within the space is recorded with cameras and other detectors (radar, proximity) and the information (delayed, treated, mixed with other input) is fedback to the observer in some, often surprising manner. The observer is then free to play with the system and explore his/her interaction with it. This element of surprise, the difference that makes a difference, is elucidated by the realtime operation of the feedback, the nowness of the system.

This kind of structure can show the process that is communication – realtime action/ response/interaction feedback loop, which, when fed by your presence, goes live (takes off). You enter a conversation with the video system and its image of you.



A live performance installation system
using doppler shift detector radars for detecting the performers movement throught the performance area - the data from the radars is fed to the sysnthesisers to control the camera / video synthesiser mix and the sounds generated by the audio synthesiser . The performer can then modulate her actions according to the feedback from monitors and speakers so as to create the effects desired.

BIOFEEDBACK – DANCE/VIDEO/MUSIC

Model: a dancer is wired up with an array of biological-process detectors/amplifiers;

GSR – galvanic skin response
EEG – electroencephalogram
– beta, alpha, theta wave monitoring
E MG – electromyogram
– muscle movement
E KG – electrocardiogram
– heart rate monitoring
Respiration monitoring

The dancer initiates by moving – the signals emanating are fed into various audio and video switching and treatment channels. Coupling must be via opto-couplers or a transmitter-receiver. Video and audio from the treated bio-signals are then fed back to the dancer who then can learn how to work with this instrument and 'play' it. The bio-signals as monitored and amplified are generally applicable to synthesisers as control voltages and clocks.

Performance:

Video; real-time multi-camera vision mixing systerns with several cameras on the dancer, on video feedback biofeedback signal displays, using voltage-controlled mixers and colorisers... (e.g. Fairlight Coloriser 108, EMS Spectre).
Audio; real-time, multi-channel bio-signals treated with and controlling an audio sythesiser.

Feedback: essential for the proper integrated operation of the instrument .. at all levels of the system... to tune the human components of the system (the dancer, video operators, sound mixer, etc.) to the process... Can develop programmed audio treatment patches and video approaches.

This kind of activity is obviously applicable to both studio realisation and to live performance.

BIO-MUSIC

Using a bio-signal pre-amplifier, e.g. an alpha wave monitor, listen to and learn the rhythms of your system. The alpha-wave monitor is especially good because it implies a contemplative state, the body system is quiescent. The signal is then used as a base structure for musical improvisation. Using a keyboard instrument one can set up rules for relating to the bio-signal as a framework within which to improvise.

Rule possibilities: 1. play only while in alpha; as you drift out of alpha devote yourself to regaining that state.
2. relate the rhythms of your playing to the rhythm of the bio-signal being used.
3. generate your own rules.

Look through all tape and image bank material being selective for image, events. By processes of juxtaposition and cut up, images thrown almost randomly together. Start to look for patterns in the relationships of images via form, content, semantic juxtaposition, etc.. Develop (if there) integrating systems of thought and action processes and their dynamic interrelationship. In a sense the task is to explicate these interior processes and to demonstrate inner/outer correspondence leading to an understanding of the identity of nature – of the nature of identity...
This is an activity of mental eco-logic. It is the logic of the stone in the circle the logic of the Tao the logic of ecology and of the mind. Dreams are plays upon, playing with the logic and the form and the content for logic and form and content are inseparable, identicaI each is a way of perceiving a part of perception and the ability to perceive. Myth, Dream and Reality Myth the structure Dream the growth Reality the manifestation through language and consensual validation. Through the structure grows the dream into reality and reality into dream. Our perception is the medium through which we realize these processes and bring them into meaning. Logos, the breath of life enters the fire of knowledge into the wellsprings of consciousness. Fire and Water are the dynamic opposites in whose conjunction the breath, Air, manifests its own conjuncted opposite, Earth, manifest reality. This mythic primordial dynamic contains the seed of a concept of process, dynamic through the multitudinal diversity of manifestation of relation and interrelation. Relationships between things in deep and intricate structure. Image, language and the world.

A short list of references follows:

Barthes, Roland: Mythologies

Bateson, Gregory: Steps to an Ecology of Mind. Palidin, London 1972.

Battcock, Gregory (ed): New Artists Video. Dutton, New York, 1978.

Burgin, Victor: Van Abbemuseum. Eindhoven, 1977.

Hall, Sue and Hopkins, John: "The Meta Software of Video", Studio International Video Art issue May/June 1976, London, pp260ff.

Hopkins, John, et al: Video in Community Development, Centre for Advanced Television Studies, London, 1972.

Levine, Les: Using the Camera as a Club. Museum of Mott Art Inc, New York, 1976.

Jones, Stephen: interview in Ozone, ed. Terry Reid, Sydney, 1978.

Reynish, Richard: "Notes on the Politics of Information", JCATS, 3 no. 1, London 1975.

Ryan, Paul: Cybernetics of the Sacred, Doubleday, Anchor, New York, 1974.

Wiener, Norbert: Cybernetics, M I T, 1948.

 **return to Chapters index**	 **return to overall index**	 **return to Introduction and textual index**