**A Scientific Basis for Personhood**

***This text was the introduction I provided for a discission I led in the Wairarapa U3A Philosophy Group on 4 August 2022***

I want to start this talk with a few general points:-

*First,* the term ‘personhood’ refers to that sense we have ‘of being a person’ (whatever that might mean). I am *not* talking about ‘personality’, the infinite diversity of personal characteristics and styles of thinking and behaving. That is a fascinating scientific issue, but not my topic today

*Second,* when I use the phrase ‘scientific basis’, this is not the only way to think about personhood. In any case, I am not just a scientist: I try to be something more complex, better integrated and more complete than that. Of course this is an impossible, never-ending journey for all of us. But I admit that from perspectives other than the scientific one, there may be views of personhood, which, at least in their own domain, have advantages over the one I will try to convey, wearing my scientific hat. As for alternative viewpoints, I suggest that you save your comments to the discussion after I’ve presented my account.

*Third,* as a scientist, I try to be rational, that means taking care to define my terms, so that they can be used rationally, that is, for reasoning. So for such reasoning, I should start by declaring my assumptions, some of which are philosophical, perhaps metaphysical; some of which are core assumptions built into scientific tradition. With regard to my philosophical assumptions you may disagree; but, again, I ask you to save your comments until I have presented my account, so that you can see what I build on those assumptions.

I hope to build up my conclusions, rationally, into a coherent picture, with no irreconcilable splits. To start with, I would not expect my basic assumptions to contradict each other. If they do, or if there do appear to be irreconcilable splits, I’d say there is more work to do to bring about a reconciliation; or you are starting with the wrong building blocks altogether.

*My philosophical assumptions*: The most important is about how we use those problem words, ‘mind and brain’, or more generally ‘mind and body’, or even ‘mind and matter (of any kind)’. It seems to me that there is an undeniable duality here, which must have been crystal clear to the earliest humans in prehistoric times – evident, as soon as we had words to describe things and our experiences, for instance the strange phenomenon of dreams, or the fact that we can have hidden desires or ambitions which are quite separate from actions we might take to fulfil them. So I call this insight ‘a fundamental ***duality***’. This universal experience then became the origin of thousands of years of philosophical rumination.

It was the classical Greeks, especially Aristotle, who, uniquely I think, stumbled on the idea that words could be used as vehicles for reasoning, in a disciplined way. However, I think he overdid it. Words are not very good tokens for reasoning, if only because words usually have multiple shades of meaning, derived from the long history of use of each word. Words for concepts are especially problematic. So, we always want to debate what words actually mean. This may explain why scientists preferred reasoning with numbers rather than with words. That is the first problem I identify with use of words for reasoning.

The second problem is that words are separate from each other; but the concepts to which they refer may be far from separate. Concepts designated by different words may be entirely interdependent. This I think is the core of the problem about the relation between mind and brain. Because the words are separate, the Greek philosophers imagined that what they called mind was completely separate from body or brain. This led to something different from that ancient duality – definite philosophical ***dualism*** (a separate term) as if mind and brain are each distinct substances; and in the hands of the Catholic philosophers this led to an entire philosophy, theology, cosmology and indeed state structure build around two-substance dualism. I believe this was (and for many still is) a profound mistake. Mind and brain are entirely interdependent, utterly different (one subjective – the view from inside, as it were; the other objective, the view from outside); yet they are inseparably blended in the organic unity of living beings such as ourselves.

So in this very old debate there are two strands. The first started with Pythagoras (who was a dualist), then Plato and Aristotle, then the Catholic philosophers, and then Rene Descartes, who while anti-clerical, was still a dualist. He and had some weird ideas about how these separate substances might interact. Isaac Newton was also a philosophical dualist.

The second strand may also have been very old, but was pioneered in Europe at the time of Newton by the heretical Jewish philosopher from Amsterdam, Baruch Spinoza. For Spinoza, mind and brain were utterly different aspects of a single entity. One line of his goes somewhat as follows: “*If two entities have nothing whatsoever in common (as is the case for mind and brain) there can be no causal interaction between them*” So the relation between mind and brain, always interdependent but known to us in utterly different ways, could not be ‘causal’. Successors to Spinoza were pioneer German scientists of the nineteenth century - Gustav Fechner, Heinrich von Helmholtz and Ernst Mach (all polymaths, equally at home in physics, physiology or psychology). In due course, these - especially Mach - influenced Einstein. The philosophy they deployed can now be called ‘psychophysical parallelism’: the two entities - mind and brain. - are never separable, always going about their ways in parallel in some mysterious way, which is not the same as permitting causal interaction.

Another philosophical assumption also goes back to the classical Greeks. Aristotle was concerned about disciplined ways to reach ‘truth’. He knew of two main methods of inference. *Deductive inference* started from general statements, deducing from them detailed particular statements; and *Inductive Inference* which derived general statements by putting together a collection of specific particular experiences. Inductive inference is inference by association, a very basic unconscious and instinctive process, which we are doing all the time. Aristotle was more interested in deductive than inductive inference, because - so he thought - it was a more reliable way to get at absolute certainty of conclusions. Today, few philosophers would argue that either of these methods guarantees absolute certainty. We come to the weaknesses of deductive inference soon.

The Catholic philosophers, and then Descartes, followed this idea in their quest for absolute truth; and then, in the seventeenth century – the Enlightenment, so-called Age of Reason especially in pre-revolutionary France) Reason was put on a pedestal - better than holy writ, possibly even better than sliced bread (if you could get any in pre-revolutionary France). Far away in the East Prussian city of Koenigsberg, then part of the Prussian empire, the philosopher Immanuel Kant produced his magnum opus entitled ‘*Critique of Pure Reason’*. The title makes clear his objective; but it is a long and difficult book. What I know of it is found in the first section, where he undermines the notion that in deductive inference conclusions follow with absolute certainty, and unconditionally and the assumptions. I use my own terms to paraphrase Kant, whose terminology is less direct. In deductive inference there are ***always*** deeply hidden assumptions, of which we might have little explicit awareness, but which are needed for that type of deductive inference. Kant argued that, somehow, those hidden assumptions were *innate knowledge*, perhaps another part of the search for and defence of the notion of absolute truth. More recently philosophers and psychologists in Europe (for instance Jean Piaget) have been influenced by Kant, and the requirement for those elusive hidden assumptions; but they reject the idea that they represent innate knowledge. Most would now argue, that such knowledge is acquired mainly *by experience*, even if at a very early age

I have my own term for those hidden assumptions. I use the term ‘***context***’. It can be used in an very wide range of circumstances, not only to enable specific instances of deductive reasoning. In short, no message or signal – none whatever - is unambiguous, its significance crystal clear, until itit is resolved by locating it in the appropriate context. This applies to single images, signals, words events, etc etc. The problematic word ‘information’ means nothing without context. *Context is the key to meaning.*

*My Own Journey:* Let me say a bit here about my own exploration of these ideas. I became interested in the brain even before I got to university, from a very good book I found in the school library ‘*Doubt and Certainty in Science’* by JZYoung, Professor of Anatomy at University College, London – his Reith lecture series from 1950. At university, as everywhere else, it was hard to get away from implied philosophical dualism – it is built very deeply into western culture (its language, and especially its legal systems); but by the age 20 I had grasped the notion that one cannot do really rigorous theoretical work on the brain with that as a basis; and although I knew nothing about Spinoza at the time, the ideas I was considering were really quite similar to his – but not with his immense subtlety and erudition, and not in Latin. The way of thinking which I developed, rather ashamedly at first, but with increasing confidence, comes straight from this. So I use two methods in coordination: *Introspection* – looking inwards about how my own mind seems to work. This is neither factual nor objective, and is certainly far from complete, but it allows me to develop *hypotheses* (a concept foreign to Aristotle, introduced by Islamic philosophers in the Golden Age of Islam); and then, in my study of the published research literature, *I seek observed facts about the brain,* which might support, or refute, or elaborate, or run in parallel with those preliminary hypotheses.

As for scientific assumptions, like Ernst Mach, I believe in the unity of the sciences, so that the same assumptions should underlie both physics and neurobiology. I don’t need to get into modern physics and relativity theory here. Newton’s ideas will suffice. His system was built upon three concepts: mass, length and time. In what I want to say here, mass is not relevant. *Time* is certainly relevant to the concept of personhood, because, whatever a person is, it endures over a long span of time – three score years and ten as the bible says somewhere. However, the brain deals with time in very different ways according to the span involved - very different for fractions of a second, compared to days, months, years and decades.owever, the brain deals with time in vdefry H *Length* is also relevant, as I will show, as an aspect of the the anatomical structure of nerve cells; and also when we come to the *function* of nerve cells. Given this, I want to build up my ideas as a physicist would, with proper theories, deriving higher level properties (such as those of the brain as a whole) from knowledge of lower level facts (such as those about nerve cells).

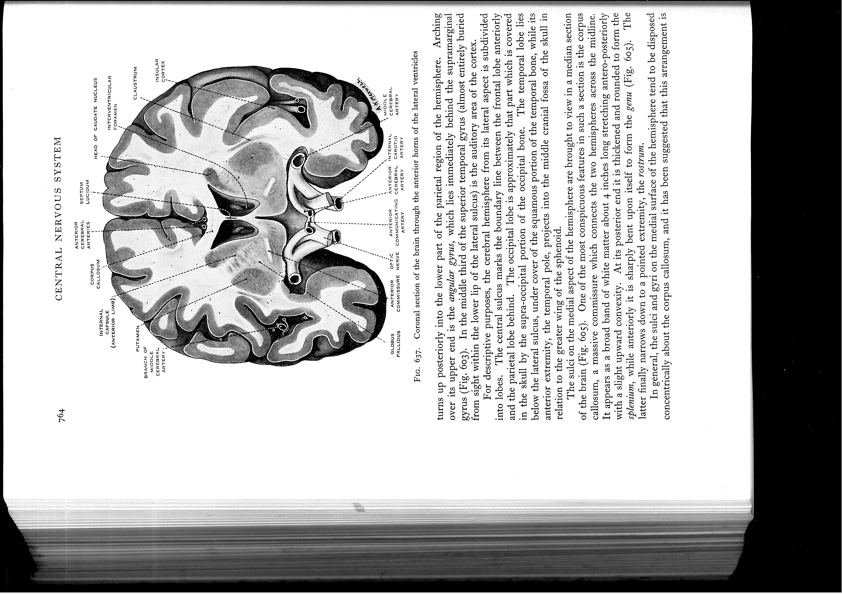
*Anatomy; Neurocytology*: So, lets talks about nerve cells. The ‘cell theory’ – separate cells as the building block for living things - had been growing since the seventeenth century. However, in the late nineteenth century it was not clear that it applied to the structure of the brain. There were obviously a lot of long fibre-like connections as well as cell-like structures, and some people thought that the cells were in continuity with each other – not really separate – a *syncytium,* as it was called.



In the 1890s, the Spanish histologist Santiago Ramon y Cajal solved this, and was one of the very first recipients of the then-new Nobel prizes. He used a staining method which stained only a small proportion of cells, but those it did stain, it did so in their entirely. In a very large number of drawings, he showed that nerve cells, despite all their long projections, are indeed separate from each other, as are all other cell types. I say ‘drawings’. Methods were available at the time to take photographs down the microscope; but Cajal preferred to *draw* what he saw, because it depicted what he saw with greater clarity. My first picture is one of his meticulous drawings, showing nerve cells in the cerebral cortex, that folded structure underneath our skull.

You can see the main body of each cell – the cell body – has a number of relatively short processes, the ‘dendrites’; and also a single fibre – the axon - which may be very much longer, and leads to somewhere else in the brain. You can see that the surface of the dendrites is covered with very many little stud-like things – dendritic spines - and sometimes you could see the end of an axon making contact with the spines. If you follow the axon from the cell labelled ‘F’, you can see it has a branch which turns back to the right, and makes contact with a dendrite of the cell labelled ‘C’. So – this is the junction – the synapse - where one nerve cell communicates with another.

Let us now step back, and look at the human brain in large scale: The cerebral cortex is a complex layers of cells about 2 mm thick – the grey matter – and beneath that is the white matter – in which are collected an astronomically large number of those connecting axons, going every which way from one region to another.



*Carl Wernicke:* Now let me introduce you to a special hero of mine – Carl Wernicke. He is best known in the English-speaking world as a pioneer of neurology; but that reputation comes from work he did as a young physician immediately after the Franco-Prussian war. There is a region in the left hemisphere of the brain, named after him, which supposedly, holds the representation of speech sounds and words. However, for the last 20 years of his life, he worked in a psychiatric institution in the city of Breslau in the eastern Slavic part of the German empire. Today it is called Wroclaw, in south-western Poland. Wernicke died tragically young in 1905, at age 57, of all things in a bicycling accident.

From my contacts in Germany I knew he had published his lectures to trainee psychiatrists in the 1890s; and that the lectures had never been translated into English. Well, a good friend of mine, John Dennison, was a very good linguist as well as an anatomist, and together, we produced a translation of Wernicke’s lectures; and it was published about 7 years ago by Springer Verlag. Far too expensive, but we are proud to have done that. Sadly, John died during our first lock-down period, not from COVID, but from a sudden heart problem. Here is Carl Wernicke.

I quickly realised that Wernicke was the best scientist I had ever read, dealing with psychiatry. He made big mistakes, but at the time his basic knowledge of the brain was quite limited. At his best, he is ahead of where most of the psychiatry profession is today. In particular, in the 1890s he was asking that fundamental question which has been bugging me, for most of my life. How does that remarkable organ inside our skull construct for each of us the sense of being a person? . . .and from the facts available in the 1890s, especially those astonishing drawings of Cajal, Wernicke concluded that the cerebral cortex, that folded structure under the skull, is ‘the organ of association’. In three words one cannot give a better description, although this is by no means the only thing the cerebral cortex can do.

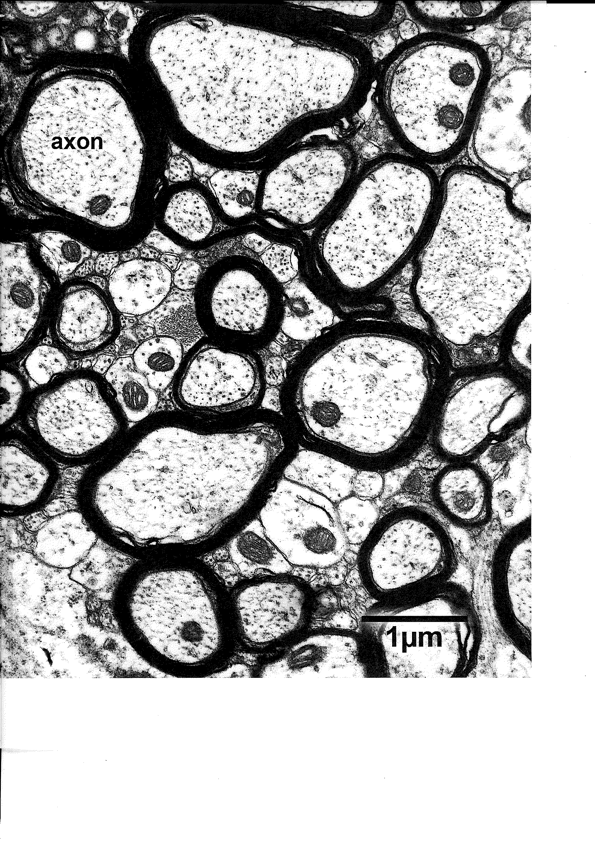
A person with a beard

Description automatically generated with medium confidence

***Association***: That is exactly what philosophers were talking about when they wrote about ‘inductive inference’. So, the most basic function of the cerebral cortex is association, connecting together things that are correlated, building patterns of the outside world and perhaps of other things, by induction, and then storing for future use, the fact that associations have been detected.

‘Association’ includes many things. When we recognise someone by their face, somehow, in an instant, we pull together most of the details which define that face into a single pattern. I say, ‘in an instant’ – and in principle all those details are presented to us in the same instant of time. But what about recognising a tune, or even more simply, a consonant speech sound? These are patterns extended across time. A consonant speech sound – ‘b’ ‘k’ ‘d’ – are a short sequences of acoustic events packed together with precise timing in a duration of perhaps 100 or 150 milliseconds. How do we recognise such patterns? Back to the structure of the brain:

The white matter, lying under the cerebral cortex contains innumerable axons derived from nerve cells located in the cortex, connecting all the different parts of the cerebral cortex together. Under the microscope, what does that white matter look like? Look at my last image: It uses not the light microscope, but the electron microscope, which can magnify far more powerfully. Look at the scale bar for one micron – One thousandth of a millimeter.



What you see is a large number of axons – nerve fibres – cut in section. In addition, there are clearly two types of fibre – some quite large in calibre with a thick sheath round the outside, other less conspicuous because they are much finer in calibre, and do not have that prominent sheath – but if you count them, they are more numerous. What does all this signify?

*First,* the calibre of an axon determines the conduction velocity of the signals it carries - the larger the calibre, the faster the conduction.

*Second,* those sheaths, the ‘myelin sheaths’ as they are called, greatly increase the conduction velocity.

*Third,* regardless of the myelin sheath, the calibre of the axons varies to a remarkable degree. Altogether, this white matter contains axons whose conduction velocity spans a very wide range.

What does this mean for ***conduction time*** between the beginning and the end of the axon? Not many people know about this. I am one of them. I am not very good at the experimental side of brain science, but in the years 1973 and 1974, I was able to study this, in the brains of deeply anaesthetised animals. We could go into the technicalities of that if you like, but I want to focus on my conclusion. For a distance of about 1 cm, the conduction time in different single axons could range from perhaps 1 or 2 millisecond for the fastest, up to about 50 milliseconds for the slowest – a huge range So, this physiological finding matches very well the conclusion you can draw from those electron micrographs. Scale this up to brains the size of the human brain, where distances can be much longer: Conduction time in different axons might range from a few milliseconds up to several ***hundred*** milliseconds. Each axon is, as it were a ‘delay line’, because conduction is not instantaneous. The broad conclusion I reach is that when two regions of the cortex are connected by a population of axons, they make up what I call ‘*a repertoire of delay lines’*, some very fast, some quite slow.

Back to my problem of how we recognise speech sounds. When a signal is carried along an axon it cannot activate the nerve cell it contacts, unless several signals arrive there within a very short period of time – about 10 milliseconds. However, if signals start off at rather *different* times, but are conducted along axons with different conduction velocities, they might still arrive at their destination within that narrow window of 10 milliseconds. In this case, the firing of the receiving nerve cell depends on events occurring in sequence which are *not* synchronous, but might be as much as a few 100 milliseconds apart, as is the case with our consonant speech sounds. The nerve cell is then responding quite specifically to a pattern spread over time.

Now this was the start for me of a large body of theory based on the premise that the connections around the brain are not uniform in conduction velocity, but make up a repertoire of delay lines. As a result, at least for small time intervals the cerebral cortex - the organ of association - can make association not only across space, as in face recognition, but across time. This was just the beginning. It is certainly not an adequate basis yet for either deductive inference or for personhood. But even as an organ of association, there are problems.

*Representation in the brain*: So, let me ask another basic question: How do the nerve cells in the cerebral cortex represent the outside world (or anything else for that matter). When I was a student, top-class neuroscience researchers in England and America were promoting the idea that different entities in the outside world were represented one-by-one, by activation of corresponding nerve cells, one-by-one. So, there was the idea floating around at that time that somewhere in the brain there might be a cell whose activation was uniquely able to evoke the memory of one’s aged grandmother – one of innumerable memories acquired by the process of association.

This manner of representation is implausible for many reasons, but perhaps most basically because of some numerical facts about the cerebral cortex.: In the human cerebral cortex, the total number of nerve cells is about 15 Thousand Million – 15 \* 10 to the power 9. The number of junctions for each of those cells in humans is anything from 20,000 to 50,000, that is far less than the number of cells. The point is that any single cell has nothing like enough incoming or outgoing connections to connect to all the other cells in the cerebral cortex directly. One cell *could* make connections with all the others *indirectly*, via relay in one or two or three other nerve cells. But then a problem arises. If the connections are indirect, the connection is no longer at all specific to the other cell: At each relay, signal transmission could spread to a huge number of other cells. The gist of the argument then is that, as an organ of association, coding of associations in the cortex is beset by a huge degree of ambiguity. To generalise, no associations can ever be uniquely determined in the network. No stimulus can ever uniquely address any specific response. In deductive inference, if you like, no assumption can ever uniquely address a specific conclusion. It all becomes a bit of a mess. However, if the brain had a method to limit what connections might be used in each circumstance, the ambiguity could be resolved. So to shift back to a term I used a moment ago, we need to find a method to discover a ‘context’ to ‘disambiguate’ those ambiguous signals so that they start to make sense, and become meaningful.

Well, in the late 1980s, I did develop a theory of how the brain could discover ways to disambiguate those ambiguous signal. In any situation, the range of connections available for connection then becomes much more restricted. Stimuli could uniquely address responses; assumptions could uniquely address conclusions. I’m not going into the details of that theory except to say a few things.

* An essential part of the theory is my finding about conduction time in axons, that when brain regions are connected by a population of axons, they constitute a ‘repertoire of delay lines’
* My first paper on this got me invited to make a visit to Russia, which I made in November 1991, a very critical time in the history of that nation, because the USSR had just fallen apart, and this vast country had no name!
* I regard that theory as the best thing I have done scientifically. I have no idea of what its impact, if any, might be.
* At the time, I had no idea that it might help resolve that question about personhood, or more widely about psychiatry. Now I think it does!

*Personhood and Episodic Memory:* So, let me lead on from there to personhood: That mechanism for discovering the contexts which can best resolve the ambiguity which besets our ‘organ of association’ can be used in many ways. Not only to enable deductive reasoning, or to make stimuli address responses in a specific way; not only to make associations more specific; but - most important - to enable a very special type of memory, which is essential to give us that sense of being a person.

Let me talk a bit about types of learning and memory. In many forms of learning, what we learn depends on repeated experience. In experimental situations, a trial is repeated over and over again, and the experimental psychologist demonstrates gradual learning as a ‘learning curve’. But this does not correspond at all to our everyday experience of recalling memories. We can recall, sometimes in considerable detail, things which happened only once, and perhaps as long as fifty or sixty years ago. Each of us has our own talent for different sorts of memory. Some people have a talent at recognising faces – they might become policemen or teachers. For me, it is sounds, music. I can remember tunes I heard just once as a teenager, which made an impression on me. This is something quite different from repeated-trial learning, and the memories so acquired. Moreover, most of us can remember episodes in our life. This is what is called ***episodic memory.*** It is mainly studied in humans. It allows each of us to accumulate quite a detailed memory of our personal life-story. As such, it is essential to our sense of being a person; and therefore it is essential to the human capacity to build complex societies.

How is the remarkable fact of episodic memory achieved? Is it unique to the human species? Probably not. It is likely to exist at least in rudimentary form in other mammals (eg herbivorous animals, trying to work out what food had made them feel ill). Given this, the fact that it is possible probably arises from some feature in common to all mammalian species.

I suggest that it becomes possible because of the principle I have already referred to, that the brain can discover and use representations of context to disambiguate signals circulating in the cortex, which are otherwise utterly ambiguous. Specifically, I suggest that our memories for episodes in our life can be encoded, for precise recall long after, simply because they are linked to the context representation in which they were first acquired. Memories can be recalled, *provided the correct context can be reactivated. Then, the ambiguities are resolved.*

Obviously the capacity for episodic memory is essential to develop a sense of personhood. But I also suggest that the context representation which enables THIS function, is most essential for that sense, the most important of all contextrepresentations. I call that most fundamental context, the ***‘context-for-living’;*** and then - juxtaposing terms originally developed separately, I suggest that the ‘context-for-living’ is exactly the same as the concept of personhood.

Well, I’ve talked long enough; but I’ve left a lot of issues unaddressed, a writhing can of worms, if you like. Before I hand it over to you, let me just list some of them:

* ‘Cerebral localization of function’, and the history of this concept – a concept which is very simple, and therefore easily misunderstood. It is as much a philosophical as a scientific issue.
* More holistic alternatives to ‘cerebral localization of function’. The story of Kurt Goldstein
* Manner of representation in the brain. Is it item-by-item to each of the many single nerve cells (the ‘grandmother cell hypothesis). I think not. It is done in a different way, more subtle, flexible and powerful.
* Do we *perceive* things bit-by-bit, or as unbreakable wholes. Actually part of a very old philosophical debate of great generality.
* How many different context representations does a person need to use?
* How long does it take to acquire a new context? . . .including the ‘context-for-living’ in a newborn infant?
* Implications for psychiatry: When severe psychological trauma disrupts our context-for-living, how long does recovery take?
* Role of sleep – perhaps to reorganise our episodic memories, so that their recall is less dependent on reactivating the original context.
* Relationship between the possibility of personhood and the size of mammalian brains.
* Is it correct to draw an analogy between digital computers and brains. I think not – a dangerous and profound fallacy in my view.
* Is the concept of ‘information’ as defined in information technology, useful for understanding brains. I think not. We need a different fundamental building block to understand the brain, one rooted in biology rather than mathematics.
* What has gone wrong with the neurosciences?
* Is my approach a form of determinism? Does it therefore undermine the element of choice in moral decisions?
* What happened to notion of ‘the soul’?