



Nothing but a pack of neurons?

Stuart J. Judge

Summary

This paper refutes the claim that we are ‘nothing but a pack of neurons’. Furthermore, the reductionistic notion that finding neural correlates of all aspects of our conscious experience would undermine the reality of our conscious experience and agency is self-defeating: if it did so then the whole scientific world-view would collapse with it. On the other hand, the discoveries of neuroscience do raise problems for classical interactive dualism. Dual-aspect monism provides a midway position between reductionistic materialism and interactive dualism that avoids the problems of the two extremes and is compatible with theism.

The sort of Cartesian dualism that sees us as disembodied souls piloting a brain that exists only to sense the external (and internal) world and to execute action has long been difficult to reconcile with knowledge from neurology of the extent to which many aspects of cognition depend on the brain, in that they are impaired or lost when it is damaged.

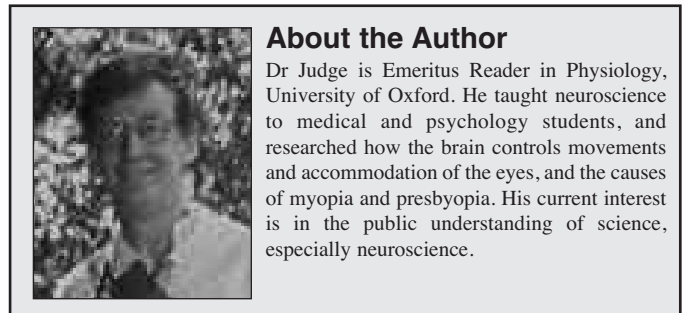
More recently a wide range of techniques has been used to investigate information processing in the intact brain, both in humans and animals, so that for some aspects of behaviour we now understand not only which areas of the brain are necessary but also a good deal about the pathways and neuronal mechanisms involved.

While there is certainly much that we do not know about the brain and cognition, it would be fair to say that where it has been possible to define a quantitative procedure for investigating a cognitive task, it has been possible to find neuronal activity that *correlates with* the cognitive performance.

By using the term ‘correlate with’ I intend to leave open the issue of the causal relationship between the two kinds of data – what we experience as conscious agents and the activity of neurons in various parts of our brain. I do that for two reasons. First, the data of experience are in a different realm of discourse from those of neurophysiology: it makes no sense to say that a neuron thinks, or that a photoreceptor sees. Thought and perception are predicates of an agent rather than a piece of hardware. Secondly, because it is not obvious, even if a mechanistic process underlies a conscious experience, that the experienced qualities, or ‘qualia’, of the conscious experience are captured by the description of the mechanism underlying it. To ask a related question, supposing one had a complete knowledge of the neuronal machinery in the brains of individuals of a number of different species, could one say anything at all about the existence or nature of their conscious experience?

Some have argued that it is reasonable to suppose that we will eventually have a complete knowledge of information processing in the brain, and that having this will a) allow us to reduce conscious experience to mechanistic terms and b) show that conscious experience is an epiphenomenon that has no causal role and is at best a shorthand way to refer to the underlying neuronal realities. This view is sometimes called ‘strong reductionism’. It is much vaunted by a small number of scientists and philosophers¹, but runs into a number of serious difficulties and I shall argue that

1 e.g. Dennett, D.C. *Consciousness Explained*, London: Penguin (1991).



About the Author

Dr Judge is Emeritus Reader in Physiology, University of Oxford. He taught neuroscience to medical and psychology students, and researched how the brain controls movements and accommodation of the eyes, and the causes of myopia and presbyopia. His current interest is in the public understanding of science, especially neuroscience.

it is mistaken. But first let us examine the scientific grounds on which it is based.

Neuroscience and artificial intelligence

What sorts of things has neuroscience demonstrated?

A considerable body of data demonstrates that the performance of different cognitive tasks depends on distinct regions or circuits in the brain. For example, it was known in the nineteenth century that damage to one part of the cerebral cortex caused difficulties with understanding language, while damage to another part of the cortex caused no such difficulties but made it difficult for the person to speak or otherwise express thoughts in a fluent way. In the middle of the twentieth century, the neurosurgeon Wilder Penfield² showed that electrical stimulation of certain parts of the cortex deep in the temporal lobe could elicit the recall of memories of events years before, of which the patient sometimes spoke as if they were in some sense re-experienced. More recently it has been shown that simply *imagining* the making of a sequence of finger movements is accompanied by activity in a region known as the supplementary motor cortex. We also know that activation of certain areas deep in the brain can be an extremely powerful motivator – animals will work, sometimes to exhaustion, to receive stimulation at some points. It has been shown that the nerve cells in such areas are normally involved in the processing of signals related to natural rewards – sight of food when one is hungry, for example³. Each of these examples (and many others) points to a close association between brain events and experience.

New methods of ‘imaging’ the intact human brain are adding detailed knowledge of which parts of the human brain are and are

2 Penfield, W. & Rasmussen, T. *The Cerebral Cortex of Man*, New York: Macmillan (1957).

3 Rolls, E.T. *The Brain and Reward*, New York: Elsevier (1975).

not active⁴ in a wide variety of cognitive tasks.

More invasive studies in animals are providing information about what sorts of neuronal transactions are carried out in different brain areas, by recording the signals carried by individual neurons in particular kinds of task. For example, consider the task of detecting whether a visual stimulus is moving left or right. It is a long-established principle of such experiments on sensory performance that one includes 'catch' trials in which there is no movement at all. It is well known that subjects sometimes make errors and deem some of these null stimuli to have moved. It is now known that in such situations the 'mistakes' in judgement are associated with slight increases in activity in neurons that normally signal motion in the direction (mistakenly) sensed. Artificial activation of a very small number of these neurons biases the (monkey) subject's responses to null stimuli.⁵ There is therefore a reasonable case for reckoning the activity of such neurons a correlate of the discrimination of visual motion.

Complex information-processing can be mechanised

While the intellectual groundwork for information technology was laid some sixty to seventy years ago by Shannon, Turing and others, miniature semiconductor technology has made it possible to incorporate these principles into devices, such as satellite navigation systems that direct every turn of one's journey, or computer programs that do the job of a shorthand typist producing a tolerable text of speech, that have made it clear to all that mechanistic processes can carry out tasks that were previously impossible without direct human involvement. It is, of course, still true that human intelligence is being used, albeit the intelligence of the designers of the machines. The point I want to make is that information technology has enlarged what we mean by a machine from something like clockwork, where one may see one cog pushing another round and so forth, to include devices with enormously complex internal structures that might contain elements (such as 'neural networks') whose detailed behaviour (as distinct from operational principles) is not known even to the designer. When assertions are made about neural mechanisms it is this kind of information-processing machine that is envisaged.

The goal of cognitive neuroscience

The issue is whether it is reasonable to extrapolate from the current findings of neuroscience to envisage a future in which a neural mechanism can be found underlying *every* aspect of our conscious (and unconscious) activity. Probably the first thing to say is that if this aim is achievable at all, we are very far from such a goal. To illustrate the difficulties, consider a kind of cognitive activity that it is not too difficult to imagine having neural correlates, because it is definable in operational terms – knowledge of elementary mathematics. Lay aside for the moment concern about whether it is appropriate to speak of knowledge at all without reference to a knower. We have some idea of which areas of the cortex are more critical for such tasks as spatial reasoning, but I don't think we have any idea how such a fundamental concept as that of an integer might be represented in terms of neural mechanisms and connectivity. Consider then how much more we need to know if we are to understand how an integer being a prime number is represented. One might conjecture that this involves the memory of how to carry out some procedure for finding primes such as Eratosthenes' sieve. But where is that remembered, and how? Now raise the stakes a bit. What might be the neural correlate of a rational conviction that there is no largest prime number? Presumably it would involve memory of the constructive proof that

the number formed from the product of all known primes, plus one, cannot be prime.

That is clearly a tall order and yet I have deliberately chosen an aspect of knowledge that is definable operationally, rather than one of the many aspects of conscious ability (say recognising an old friend, or writing an article) we have little if any idea how we accomplish. Nevertheless, the claim is made that sooner or later it will be possible to reach the goal of finding the neural mechanisms underlying each aspect of experience, and it is from this conviction that the radical conclusions about the absence of human freedom, and indeed the illusory or epiphenomenal nature of consciousness, have been drawn by some scientists and philosophers.

Strong reductionism

*Is neuroscience a threat to human dignity?*⁶

If it were true that modern neuroscience has shown that indeed we are 'nothing but a pack of neurons'⁷ (as the Nobel laureate Francis Crick put it) this would be a finding that would rank alongside any of the major turning points in the history of science. Several years ago a conference was convened in Washington DC to discuss the perceived threat of neuroscience to human values. As an editorial about the conference in one of the most prestigious neuroscience journals⁸ put it:

belief in free will and moral choice forms the basis for our concepts of responsibility and culpability, and hence our legal systems ... For most of us, they are among our most fundamental assumptions when dealing with other people. Yet, there is an uncomfortable tension between these widely held beliefs and the intellectual views of many scientists. In particular, the rapid progress of neuroscience has been interpreted by some as providing new ammunition for a materialist account of human nature and thus an attack on traditional belief systems.

Epistemological objection to strong reductionism

One presupposition of strong reductionism is the sort of naive realism that regards the stuff of the physical world (neurons and their patterns of connectivity and activity) as somehow more real than anything else. But consider, for example, how one knows that there are such things as neurons, the answer seems to me to be that such knowledge depends on a huge set of scientific conclusions established by reasoning from data. To see a neuron one needs a microscope. To know that a microscope produces valid results one needs some method of calibrating it, say by some sort of micrometer. But to know that the micrometer moves as it does rests on arguments about geometry, and so on. Ultimately one has to believe in the validity of observations by scientists and reasoning about them. In that sense, our knowledge of neurons and their structure presupposes the general validity of our conscious experience and reasoning. That we are able to make valid observations and reason about them cannot be less certain than the conclusions of those activities.

Strong reductionism is self-defeating

It is even worse than that. Suppose for the sake of argument that one accepted the view that consciousness is an epiphenomenon or illusion and that neuronal mechanisms are the only real thing about us as agents. Then we have no way of expressing that conclusion, because all such ratiocination is superfluous. In other words strong reductionism, like other forms of materialism, impales itself on its own sword. If we are 'nothing but a pack of neurons' then there is no one to make that assertion. Like other varieties of materialism

4 It should be realised that what is measured is energy consumption, or blood flow, rather than information processing per se.

5 Parker, A.J & Newsome, W.T 'Sense and the single neuron: probing the physiology of perception', *Annual Review of Neuroscience* (1998) 21, 227-277.

6 See MacKay, D.M. *Human Science and Human Dignity*, London: Hodder & Stoughton (1979) for a discussion of these issues.

7 Crick, F. *The Astonishing Hypothesis: The Scientific Search for the Soul*, London: Simon & Schuster (1994), p.3.

8 *Nature Neuroscience* (1998) 1, 535-536.

(and this is not of course an original point⁹) strong reductionism is self-defeating: it is only as conscious agents that we are able to design the experiments, make the observations and interpret the results that lead us to our neuroscientific conclusions. The mistake in strong reductionism is not in the positive assertion that we are embodied in neural machinery, but in the denial that we are anything else. Such denials have been aptly dubbed 'nothing-buttery' by MacKay¹⁰.

Cyberscopsocopy

One way of expressing strong reductionism is to consider a machine that could monitor, moment by moment, the state of all my neurons and their interconnections (rather a taxing task, to put it mildly, considering the enormous amount of data to be recorded and analysed). If conscious experience is an epiphenomenon, it is possible to 'translate' the neuronal data into the usual terms of conscious experience. But what or who could understand this translation? Suppose that at the time in question I was remembering my mentor and his importance to me. The output of the machine might be something like 'Thinking of someone who died two decades ago, was a father-figure as well as professional mentor, name pronounced MacKay ...'. I could understand that, as could one of my old friends or close family, but we do so by sharing a history. For the machine to have an understanding of this translation it would also need to have knowledge of the same history and to be endowed with many, if not all, of the attributes of a person. In short, it would need to be rather like another conscious agent. Something absolutely essential is missing from the reductive description.

Dualism

Opponents of strong reductionism have frequently adopted dualism. I refer here particularly to *interactive dualism*, the claim that there is a separate, non-physical, mind or soul *interacting* with the brain. This, too, has its problems.

Neuroscientific constraints on dualism

The dependence of cognition on the brain constrains the kinds of dualism that are viable or, to put it another way, what sort of role it is reasonable to attribute to a non-physical mind or soul. To take one example, memory is dependent on the integrity of the brain. The memories of someone suffering from severe dementia are lost. This raises the question of whether the mind has a memory of its own. If it does, then one has to allow that these memories can be inaccessible to us. That seems artificial, and it is more reasonable to suppose that memory has a physical basis, or relies on a physical basis, that is external to the mind. Similar difficulties arise with every other aspect of cognition known to depend on brain integrity; that so much of our cognition depends on the integrity of the brain at least greatly narrows the role for a non-physical mind. In a similar way, one might imagine the soul having a copy of memories that had been lost in dementia, allowing these to be recovered in eternity, but one would presumably need to accept that the memories came from the brain originally.

A doubtful argument for dualism

An argument that has been used by advocates of dualism that seems to me unsound goes as follows. To assert a physical cause for our action is necessarily to deny that what we did was responsible and rational. Therefore to say that I did whatever I did because of some sequence of neuronal transactions is inevitably to remove responsibility for the action. For example, if I strike my wife because I am suffering from a seizure in which my arms flail about,

then I am not responsible for that action. This seems to me to be an argument from an unrepresentative example – or at least an example that begs the question at issue. In the seizure a 'hardware' problem has arisen in which motor control is effectively uncoupled from all the normal considerations (which may or may not be completely represented by neuronal activity; that is the issue). Such examples cannot inform us about the general issue where there is no disruption of normal brain mechanisms.

Dual-aspect monism

There is a midway position between interactive dualism and materialism. I shall call this dual-aspect monism¹¹. This stresses the reality of both our conscious experience and our knowledge of the physical events in the brain that we have found empirically are correlated with our experience, and regards it as an empirical question how these two realms of discourse, or categories of experience (the insider or 'I' view and the observer or 'it' view) are related. Conscious experience is considered to be the inside view of the information processing that is going on in the brain.

Consciousness

Why should our experience as conscious agents be correlated with activity in the brain? At one level, this is not surprising. For our seeing an object to be veridical rather than illusory, granted that there are sensory mechanisms at all, there must be correlations between our experience and activation of the appropriate parts of the sensory mechanism. We may agree with this regardless of whether we believe conscious experience to be more real than the operations of sensory mechanisms, or vice versa.

If one starts from the standpoint of a naive realist materialist who considers that matter is all there is, it is indeed something of a puzzle why we (or indeed any other organism that may be conscious) should be so, but this seems to me to misconstrue the issue. We know ourselves to be conscious. It is one of the most certain facts. The issue is what sort of neural activity we would expect to be correlated with conscious experience.

This is certainly not a question to which we know the answer, but one may speculate that the sort of things one is conscious of might be those that can be communicated to one's fellow beings. If so, the unconsciousness of a dreamless sleep would be one in which the communicative mechanisms (including but not restricted to language; we can communicate by drawing or painting, for example) are either shut down, or at least disconnected from inputs supplying data from the senses, or memory, and whatever the substrate might be of imagining.

Free will and the soul

One of the long-standing attractions of dualism has been its support for the common-sense notion of the reality of our consciousness, and of such crucial aspects of who we are as having intentions, making decisions and so on. Dualists have been concerned about whether for us to be free and responsible agents, rather than automata, there must be gaps in the chains of cause and effect in our brains so that at these gaps our minds can exert control over the grey matter of our brains. The Nobel Prize laureate John Eccles, for example, was a dualist in this sense.

A detailed discussion of free will and brain determinism is beyond the scope of this paper, and I mention it here only in relation to the question of dualism. Most philosophers are either *compatibilists*, maintaining that free will is compatible with a deterministic functioning of the brain, or *libertarians*, maintaining that free will requires some degree of indeterminism. In general, dualists have tended to be libertarians, because a dualist requires a gap in physical causality for the non-physical soul to be able to

9 See e.g. Lewis, C.S. *Miracles*, London: Bles (1947), quoting Haldane, J.B.S. *Possible Worlds*, London: Chatto & Windus (1927).

10 MacKay, D.M. *The Clockwork Image*, London: IVP (1974).

11 See MacKay, D.M. *Brains Machines and Persons*, London: Collins (1980) for a similar view, though he does not use the term 'dual-aspect monism'.

influence the physical brain. In contrast, monists (including dual-aspect monists) have tended to be compatibilists; for them, gaps in brain determinism would imply dysfunction, and would work against free will, not for it. For a monist, free will requires freedom from external constraint, not from internal causality.

Most modern libertarians invoke quantum indeterminacy to permit covert steering of neuronal processes by the soul or mind¹². One difficulty with this view is that it is not obvious that the sort of free will one cares about – to make rational decisions to the best of one's ability, on the basis of the available evidence and according to one's convictions – can be safeguarded by such means. If every aspect of making a decision (the nature of the evidence, the principles to be applied, the weighing of the evidence) has a correlate in brain activity, then it is not clear what the extra input achieves.

Among the problems faced by compatibilists is the argument that we could hardly be free if a prediction of our behaviour existed that was binding upon us. A powerful counter-argument, developed by MacKay¹³, points out that there is a very curious logical status to predictions of our own future brain states – namely that they cannot be equally valid whether we believe them or not.

[If] all you think, believe, hope, experience, and so on is represented by the physical state of some part of your brain, ... it follows ... that there must be one part of the brain – namely the mechanism that represents what you believe – which must necessarily change if any change takes place in what you believe. Does there then exist a complete specification of that part of your brain, which you would be both correct to believe and mistaken to disbelieve if only you know it? Obviously not. Suppose I had the means of analysing your brain state and producing a complete description of it which is correct as I see it here and now; then obviously if you were to believe it, that state must change. By the same token even if I could calculate completely the immediate future of your brain from my description (without letting you know), my detailed prediction would have no claim to your assent. What I would be correct (secretly) to believe about your future is something that you would be mistaken to believe!¹⁴

This is just one of many arguments used by compatibilists, showing that free will can be compatible with monism, including dual aspect monism.

Brain and Soul

It may well be asked whether there is room in dual-aspect monism for the soul. The answer to that will depend on what one means by soul. Old Testament usage seems to be closer to the Aristotelian than to the Platonic concept of 'soul'. For example, when we read

12 Kane, R. *The Significance of Free Will*, Oxford: Oxford University Press (1996).

13 MacKay, D.M. 'On the logical indeterminacy of a free choice', *Mind* (1960) 69, 31-40.

14 MacKay, D.M. *Human Science and Human Dignity*, London: Hodder & Stoughton (1979), pp.52-53.

that 'The Lord God formed man of dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul'¹⁵, the word translated 'soul' is the Hebrew '*nephesh*'. There is no sense of *nephesh* as a separable part of man – indeed the word is also applied to animals. *Nephesh* is regarded as departing at death, but the word is never used to mean the spirit of the dead¹⁶. Similarly when the Psalmists speak of their soul (*nephesh*) the reference is not to a separable part of man but to man's whole being.

In the New Testament, the word used for soul (*psyche*) carries dualistic overtones from Platonic philosophy, but the New Testament writers emphasise the unity of the human person, and do not teach the idea of a disembodied soul. Most strikingly, the New Testament doctrine of the resurrection of the body is far from the Platonic concept of an eternal, immaterial soul with the potential to exist in isolation from the body, and indicates that embodiment is an essential aspect of our nature. St Paul is very clear that the new life to which those 'in Christ' look forward after death is not one in which we shall be disembodied souls, but one in which we will have new bodies¹⁷ – what Tom Wright, a masterful present-day expositor of Paul's thought, calls 'life after life after death'.¹⁸ Elsewhere Paul likens death to sleep, which would seem to imply that there is no consciousness (or at least only the possibility of an intermittent dreamlike consciousness) before re-embodiment in the world to come.

What then of Jesus's saying: 'be afraid of the One who can destroy both soul and body in hell'?¹⁹ This certainly reminds us that our Maker has the sovereign authority to do more than destroy our present embodiment. But does this use of the word soul (*psyche*) imply a separate kind of non-physical stuff that attaches to our bodies when we are alive and detaches from them at death, or can it be read as a way of indicating that from the point of view of the Creator our identity extends beyond space and time?

Conclusions

We are more than just a pack of neurons. Our conscious experience and agency cannot be undermined by science, because they are the ground on which science is constructed. In the opinion of the writer it is an open question whether it is possible to have a complete account of human activity at the mechanistic level, but in the current state of knowledge dual-aspect monism has many advantages.

Acknowledgements

The author is grateful to Prof. Peter Clarke, Lausanne University, Switzerland, for his input and advice on this paper.

15 Gen. 2:7 (KJV).

16 Anderson, R.S. 'anthropology, Christian', in McGrath, A.E. (ed.) *The Blackwell Encyclopedia of Modern Christian Thought*, Oxford: Blackwell, pp.5-7; Cameron, W.J. 'Soul', in Douglas J.D. (ed.) *The New Bible Dictionary*, London, IVP (1962), p.1208.

17 1 Cor. 15:34-46.

18 Wright, N.T. *The Resurrection of the Son of God*, London: SPCK (2003), p.31.

19 Matt. 10:28.

The Faraday Papers

The Faraday Papers are published by the Faraday Institute for Science and Religion, St Edmund's College, Cambridge, CB3 0BN, UK, a charitable organisation for education and research (www.faraday-institute.org). The opinions expressed are those of authors and do not necessarily represent the views of the Institute. The Faraday Papers address a broad range of topics related to the interactions between science and religion. A full list of current Faraday Papers can be viewed at www.faraday-institute.org from where free copies can be down-loaded in pdf format. Print copies like this one can also be obtained in bulk quantities of ten or more at £1.50 per 10 copies + postage. Secure on-line ordering details are at www.faraday-institute.org.

Publication date: March 2010. © The Faraday Institute for Science and Religion