

Knowledge of the mind and knowledge of the brain

Tim Crane

University College London

The problem of consciousness – the problem of how the matter of our brains produces perception, sensation, emotion and thought – is often described as one of the outstanding remaining problems for science. Although a lot is known in detail about how the brain works it is widely believed that the explanation of consciousness is something which still eludes us. According to a recent survey in (of all places!) *The Economist*, ‘consciousness awaits its Einstein’.¹ Consciousness researchers are looking for that missing piece of the jigsaw which will explain how the lived world of conscious experience arises out of the initially unpromising yoghurt-like matter of the brain.

Yet it seems to me that our predicament in understanding the relationship between mind and brain is not one of ignorance, but one of confusion. In other words, it is not the kind of problem which will be solved by finding out some more facts about the brain (or the body, or the external world). If there is a solution to this problem at all, it is something which we can only find by re-arranging what we already know. This is what I would like to argue today.

I’d like to start by reminding you of some of what we already know. The human mind might be the biggest mystery, but oddly enough it is a mystery we know

¹ *The Economist* December 23rd 2006, p.11.

a lot about. By this I do not mean the knowledge that neuroscientists and psychologists discover; although there is a lot of this too. I mean instead the knowledge which we all have as part of our common endowment: our self-conception and our conception of others. We know that other people think, perceive, remember, imagine, reason and desire; we know that they experience sensation, they feel emotion, they make plans and decisions and intentions. And we know a lot about what these various kinds of mental phenomena are. We know, for example, that if someone intends to do something, they must *think* or *believe* that they can do it; that if someone remembers doing something, they must remember that it was *they* who did it; that if someone feels a sensation, they must feel it to be located in their body. This is the kind of thing which forms the foundations of our knowledge of the mind; more controversial hypotheses employing these mental concepts (such as the psychoanalytic hypothesis that dreams are wish-fulfilments) are built on these foundations.

It might be thought that this knowledge is simply knowledge of the meanings of mental words like ‘sensation’, ‘memory’ and ‘intention’, and therefore not knowledge of substantial truths about the mind. But this objection rests on the dubious idea that there is a sharp distinction between knowing the meaning of a word and knowing substantial truths about the world. Certainly one does learn something what the word ‘intention’ means when one learns that one cannot intend to do something unless one believes one can do it; but one also learns something about intention itself. Suppose you are reading about wine and oenology and you come across the word ‘malolactic’ to describe a kind of fermentation. You don’t know what the word means so you look it up in a dictionary. You discover that ‘malolactic’ refers to the stage in the development of wine when the harsh malic acids are fermented into

softer lactic acids. Have you learned what the word ‘malolactic’ *means* or have you learned what malolactic fermentation *is*? The obvious answer is, *both*; and we can say the same about the meaning of the words ‘intention’ and ‘belief’.

This knowledge is sometimes called the knowledge of a ‘theory’: the theory of mind, or ‘folk’ or commonsense psychology. If any collection of principles, implicit or explicit which we use to understand something deserves the name of a ‘theory’ then we should not object to calling commonsense knowledge a theory (though I myself regret the pejorative connotations of the word ‘folk’, as if the theory were as significant as homespun folk wisdom or folk music). But there are nonetheless some dangers in thinking of our knowledge in this way, and we need to guard against them.

The philosophers Paul and Patricia Churchland have for many years argued that our commonsense psychological knowledge is theoretical, and that the theory is a kind of ‘proto-scientific’ theory, just as our primitive ‘folk mechanical’ theory of solid objects and their motions through space is a kind of precursor of fully-fledged mechanics. But the Churchlands have also argued that as a theory, commonsense psychology is woefully inadequate, and it should not form any basis for any future scientific study of the mind. They call this doctrine ‘eliminative materialism’ because its aim is to ‘eliminate’ the categories of commonsense psychology:

eliminative materialism is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of the theory will eventually be displaced ... by completed neuroscience.

Their reason for saying this is that commonsense psychology has been around for a long time, but has not explained very much about the mind. In particular, it has failed to explain

the nature and dynamics of mental illness, the faculty of creative imagination...the nature and psychological function of sleep...the rich variety of perceptual illusions...the miracle of memory...the nature of the learning process itself...²

This conception of the limits of commonsense psychology has not won many followers over the years; and it is not hard to see why. The idea that the knowledge which is embodied in our everyday self-conception should be thought of as a second-rate form of scientific knowledge is the kind of distortion which could only come from a myopic, over-intellectualised and Whigish conception of human knowledge in general. Human knowledge is of many kinds – practical, theoretical, experience-based, know-how, ability-based... Here I think we can say about knowledge what Edmund Husserl said about truth:

The trader in the market has his market-truth. In the relationship in which it stands, is his truth not a good one, and the best that a trader can use? Is it a pseudo-truth, merely because the scientist, involved in a different relativity and judging with other aims and ideas, looks for other truths – with which a great many things can be done, but not the one thing that has to be done in the market? It is high time people got over being dazzled, particularly in philosophy

² P.M. Churchland, , 'Eliminative materialism and the propositional attitudes' p.73

and logic, by the idea and regulative ideas and methods of the “exact” sciences – as though the In-itself of such sciences were actually the absolute norm for the being of objects and for truth.³

What all knowledge has in common is not that it is aspiring to the kind of knowledge we have in science, but that it is a form of cognitive achievement. And our knowledge of the mind is a different form of cognitive achievement from scientific knowledge, as I shall argue shortly.

But even when applied to scientific knowledge itself, the Churchlands’ ideas are wildly implausible. For psychology starts off with a conception of the phenomena it wants to explain which is thoroughly mentalistic and ‘folk-psychological’. A psychologist might start a project investigating, say, joint attention, and in doing so will construct experiments which may involve asking people (in ‘folk psychological’ terms) to attend to things in their environment. What they are interested in is the mechanisms that subserve joint attention: what makes it the case people and animals have that very capacity. Similarly with memory, perception, cognitive development, and all other areas of psychology. The central aims of the science of psychology are expressed in terms of phenomena which the eliminative materialist is committed to rejecting.

We should not take too seriously the idea of commonsense psychological knowledge as a theory, then, and we should take less seriously the idea that it is a proto-scientific theory. The title of Gopnik and Meltzoff’s book on the child’s theory of mind, *The Scientist in the Crib*, should be taken with a liberal pinch of salt. But we should take even less seriously the idea that it is a bad theory.

³ Edmund Husserl, *Formal and Transcendental Logic* (translated by Dorion Cairns; The Hague: Nijhoff 1969) p. 245

Nonetheless, the popular view remains that when it comes to consciousness, the solution to the problems must be *scientific* (in fact, it must be found in neuroscience), and that philosophy as a purely ‘a priori’ non-experimental discipline has very little to contribute to the real solutions of these problems. This is actually the impression one gets when reading the series of books produced over the last ten years or so by some of the world’s most distinguished scientists – from Roger Penrose to Francis Crick and Gerald Edelman – which attempt solutions to the traditional philosophical problems of mind and consciousness.

One clue that the problem of consciousness is not a straightforward scientific research programme is that there is no general consensus among these Nobel prize winners even about the basic methodological questions in this area. One question which has been very important in the philosophy of mind and science is the question of reduction: whether higher-level or complex phenomena can be explained in terms of, or identified with, more fundamental, simpler phenomena. The paradigm of reduction in the philosophy of science is the reduction of thermodynamic phenomena to mechanical phenomena. According to the reductionist, thermodynamics has been explained in terms of statistical mechanics, and thermodynamic properties like temperature can be identified with mechanical properties.

Reductionists about the mind say either that psychology can be explained in terms of biology or neuroscience, or that psychological properties can be identified with biological or neuroscientific properties; or both. Some consciousness researchers are straightforward reductionists. In his well-known book, *The Astonishing Hypothesis*, Francis Crick says that

“You”, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.

Gerald Edelman has a different view:

To reduce a theory of individual's behavior to a theory of molecular interactions is simply silly.... Even given the success of reductionism in physics, chemistry, and molecular biology, it nonetheless becomes silly reductionism when it is applied to the matter of the mind.

It seems, then, that even when taking a scientific approach to these problems, fundamental philosophical questions (like the question of reduction) are not settled. In science, reduction is as controversial as it is in philosophy. The reason, it seems to me, is that the question of reduction is partly a philosophical question.

Curiously, though, one thing that the Nobel prize winners do agree on is their conviction that it is only through science that these problems can be solved. The popular scientific books on mind and consciousness do not (unlike the Churchlands) aim to debunk or sidestep the traditional philosophical debates and problems. Rather they proclaim that they will solve them by the employment of rigorous science instead of philosophy (which is in turn accused of being both woolly and scholastic).

According to Crick, Antonio Damasio and many others, the traditional philosophical problem of consciousness has only become tractable since science got its hands on it.

We find a number of striking examples of this in Damasio's 2000 book, *The Feeling of What Happens*. Damasio tells us that ‘science can now successfully

distinguish among several components of the human mind' and offers the distinction between *consciousness* and *conscience* as an example. But we did not need science to tell us this; all we needed was a dictionary. Similarly, in his account of the latest research on animal minds, *Wild Minds*, Marc Hauser dismisses the questions, 'do animals think?' and 'are animals conscious?' as unhelpful because they are 'vague', preferring to replace them with more 'precisely specified' questions, such as whether an animal can 'understand its own beliefs'. But it's hard to see how an animal could understand its own beliefs if it were not a thinker, since whatever else they are, understanding and believing are surely kinds of thinking. The first question is no vaguer than its 'precise' replacement. Indeed, for my own part, I find it harder to get a grip on the question of how an animal can 'understand' its own beliefs, than I can to get a grip on the question of whether an animal can think. Whether, and in what way, an animal can think is a difficult and puzzling question; but we do not make progress here by replacing it with an even more complex one. There can be an illusion of rigour in these discussions, a spurious sense that now that scientists are involved, the traditional concerns of the philosopher and the non-scientific reflective thinker can be sorted out.

Some might object at this point that the reason why neuroscientists take this route is because of their frustration with the philosophers' traditional interest in questions of the immortality of the soul, the proofs of Cartesian dualism, the irreducibility of qualia, and so on – ideas that might be of some debating interest but not ones which are live options for working scientists. With a few famous exceptions, modern scientists working on the brain or the mind have never found the need to postulate Cartesian souls. The postulation of entities strikes some scientists as a paradigm of 'untestable' philosophical speculation, rather like the hypothesis of

determinism. In his recent book, *Mind Time*, Benjamin Libet writes that ‘there has been no evidence, or even a proposed experimental test design, that definitively or convincingly demonstrates the validity of natural law determinism’. And he concludes from this that if there is no actual evidence either way, then ‘one can propose anything without any fear of being contradicted’. I think this is the attitude of many experimental scientists to metaphysical questions.

I myself would defend metaphysics from this kind of attack, but I am not objecting to the purely scientific approach on these grounds today. The philosophical issues surrounding the relationship between mind and brain are not all ontological: they are also epistemological, and phenomenological. That ontology is not the issue can be illustrated by the fact that Crick’s ‘astonishing hypothesis’ is not so astonishing to anyone who has read 20th century materialist philosophers (and their 19th and 18th century predecessors). The claim made by Edelman and Giulio Tononi at the end of their book, *Consciousness*, that ‘consciousness arises from certain arrangements in the material order of the brain’ is something which would have been endorsed by Thomas Hobbes in the 1650s, and has been virtually a commonplace in late 20th century materialist Anglophone philosophy.

The issue is not that philosophers think that the real question is over whether materialism or dualism is true. Rather, it is that the problem of consciousness is not *solved* by asserting materialism; for in a sense, materialism is what gives rise to it. I can illustrate this by commenting briefly on the relevance of fMRI scans to the philosophical problems of consciousness. We are given pictures of which bits of the brain ‘light up’ when certain mental functions are exercised. (‘Lighting up’ means: the scanner detects brain activity because brain activity is associated with a higher

level of oxygenation in the blood in the brain, and the magnetic resonance signal of the blood depends on the level of oxygenation in the blood.)

It is impossible not to be impressed by this way of finding out about the brain. But how much does it help us in understanding the mind-brain relation that puzzles us? We are baffled about the relationship between the wet yoghurt stuff in our heads and our inner lives. The fMRI tells us that when certain mental functions are exercised, there is neural activity in some part of the brain. But in itself, this is hardly surprising: that the brain is the organ of thought and other mental activity is something which everyone knows, and the idea that some parts of the brain light up when you talk, and other parts light up when you walk, tells us little about how it is possible that the brain could be the organ of thought in the first place. Given that we are already committed to a very close kind of dependence of mental activity on the brain, the mystery of consciousness is not dispelled by finding out where exactly the brain activity corresponding to this mental activity is located. We already knew it would be located somewhere in the brain; why worry about exactly where it is located?

The same applies to other ambitious attempts to find what some call the neural correlate of consciousness, like Crick and Koch's suggestion that consciousness is correlated with the firing of neural assemblages at a common frequency of 40 Hz. The question I am interested in is not whether there is such a correlation, but that *if* there is, what would it tell us? Supposing we found the very same pattern of brain activity whenever someone was in a conscious state: how would this relieve the mystery? We are not helped by merely finding something that is *correlated* with consciousness; we wanted to know how the brain gives rise to consciousness at all. To say 'certain states

of the brain are correlated with states of consciousness' is not an answer: put in this way, we already knew that something like this would be true.

I am not saying: conscious states are not states of the brain because they don't *seem* like they are. The fallacy in this reasoning can be exposed by comparing it to a well-known story about Ludwig Wittgenstein. Wittgenstein asked his students why people used to think that the sun went round the earth. One of his students said 'because it looks as if the sun goes round the earth'. Wittgenstein responded: 'but how would it look if the earth went round the sun?'. The obvious answer is: exactly the same. We can make a parallel point about the idea that consciousness might be a brain state. Why did people think that consciousness was not a brain state? Because it seems as if consciousness is not a brain state! But how would it seem if it were a brain state? Exactly the same! My point is not that materialism is false; it is that we do not know how to fit all of our knowledge together.

In a famous essay on this subject, 'What is it Like to be a Bat?', the philosopher Thomas Nagel said that materialist theories of consciousness are in the position of an imagined ancient Greek philosopher who says that matter is energy: they have said something true, but they do not know how it *can* be true. If this is right, then the problem of consciousness is first and foremost a problem about our *knowledge* and how it is to be integrated. I cannot solve this problem in this lecture – you may not be surprised to know – but in the time that remains I will offer some constraints on any viable solution.

The essence of the problem derives from the fact that we have substantial knowledge of the brain, and also substantial knowledge of our conscious states of mind; and although this knowledge is not – it cannot be – inconsistent, we do not know how it hangs together, how it should be integrated. Some philosophers express

this in terms of the idea that there is an ‘explanatory gap’ between our knowledge of the material world and our knowledge of consciousness. Let me elaborate briefly on this conception of the problem.

I claimed earlier in this lecture that one thing we have to accept is that the knowledge in question is not all scientific, theoretical knowledge. Bertrand Russell once wrote that ‘it is plain that the sighted know things that the blind do not; but a blind man can know the whole of physics’. It follows from this that the knowledge which sighted people have and the blind lack is not part of physics. What kind of knowledge is this?

Russell’s idea can be illustrated by using a thought experiment of the Australian philosopher Frank Jackson’s. Jackson imagined a brilliant scientist, whom he christened Mary, who lived all her life in a black and white room. Let’s suppose that Mary knows all the scientific facts – everything that there is to know – about colour and what it is like to see colour. And then suppose that one day Mary comes to see something red for the first time. Intuitively, she learns something new. Yet what she learns cannot be part of science, since by hypothesis she already knew all the scientific facts. So there must be more knowledge than the knowledge which science gives. In Mary’s case, it is the knowledge of what red looks like, of what it is like to see red.

Jackson originally drew an ontological conclusion: that Mary must be learning about some non-physical aspect of reality. I don’t think this conclusion follows. All that follows is that she is in a new state of knowledge, and a difference in a new state of knowledge does not imply a difference in things known about. This new knowledge is the kind of knowledge that you can only have by having certain kinds of subjective experience: this is Russell’s point. For this reason, I call it ‘subjective

knowledge'. Part of the problem of integrating our knowledge of mind and our knowledge of brain is explaining how such subjective knowledge is possible.

The problem of the 'explanatory gap' arises for subjective knowledge because it does not fit the normal model of how our knowledge of the world fits together. For example: the superficial facts about water (transparency, liquidity etc.) can be explained by the underlying molecular facts and the laws of nature, in such a way that someone who knew these underlying facts and laws would be able to deduce that the superficial facts are the way they are: given full knowledge of the underlying facts and laws, it is impossible and therefore genuinely inconceivable that the superficial facts could be other than they are. But with the kind of knowledge that consciousness gives, it seems perfectly conceivable that someone could know all the underlying scientific facts and laws about the brain, and yet not be able to draw any conclusions about whether, or in what way, the creature was conscious. This is Mary's predicament. So even if consciousness is, as a matter of fact, physical, we do not yet explain how it is, since we cannot derive in any way the truths about consciousness from the underlying physical facts. This is the explanatory gap; it is the problem of how we integrate our knowledge.

The general solution to the problem of the integration of our knowledge requires two things. First, it requires that we remove the conceptual obstacles to our achieving an adequate conception of our knowledge of the mind. And second, it requires that we provide what Wittgenstein called a 'perspicuous representation' of our knowledge of mind. I will finish this lecture by making some comments on the right way to approach these two tasks.

First, we must remove obstacles to the correct understanding of the phenomena. Like any subject of scientific explanation, discussions of consciousness

need to start with a clear conception of the phenomena to be explained. But in the case of consciousness, there seems to be an especially acute danger of being captivated by an image or picture of the inner life, which can lead at best to dead-ends and at worst to confusion. In *the Feeling of What Happens*, Damasio describes the problem of consciousness as that of how we get a ‘movie-in-the-brain’. Although he notes some limitations of this metaphor, he does not mention its most obvious limitation: being conscious of the world is *nothing like* watching a movie. When we watch a movie, we are aware of something happening in a represented space, and we are aware of the boundaries of that space. No matter how absorbed we are in the movie, we are always aware in the background that we are sitting in a cinema surrounded by others. (Woody Allen’s mother apparently once said that she never went to the cinema: ‘what is the point in sitting in the dark with a hundred people you don’t know and spoiling a good outfit?’)

Ordinary states of consciousness, by contrast, do not involve awareness of a represented space, or of representations at all; we feel ourselves to be immersed in the world which we perceive. Descartes, often portrayed in these discussions (as he is by Damasio) as the source of many misconceptions about the mind, described things so much more convincingly: ‘I not lodged in my body, like a pilot in his ship, but I am joined to it very closely and indeed so compounded and intermingled with my body, that I form, as it were, a single whole with it.’ Trying to combine this important insight with the metaphor of the movie in the brain leads nowhere. Whatever Descartes’s error was, it did not lie in this description of the phenomena of consciousness; scientists of consciousness may still have much to learn from the philosophers – even from Descartes.

Descartes was drawing attention to the phenomenological intimacy we feel when we experience ourselves as embodied. The distinctive phenomenal character of bodily sensation is experienced not as a simple quality, as one might experience a flash of light before one's eyes. Rather, bodily sensation is experienced as having a structure, with its objects being both the felt qualities and the felt locations within the body. Where we can feel a sensation to be is also felt to be a part of our body; bodily awareness involves what my colleague Mike Martin calls 'a sense of ownership'. In this way, sensation differs from outer perception: the objects of outer perception are experienced as inhabiting a space independent of us.

These phenomenological differences cast some doubt, I think, on the ideas behind recent attempts to find a neural correlate of consciousness (called NCC in the literature). The 40 Hz proposal, for example, presupposes that the NCC is a single homogenous type of phenomenon with a relatively simple structure. But the phenomena with which this is supposed to be correlated are far from homogenous and not simple in structure. The sensation of pain, for example, is not a simple quality, but involves a structure of intentionality and affective response. And these aspects of the phenomenal consciousness of pain are very different from the phenomenal consciousness involved in outer perception, for example. While there is no *a priori* objection to correlating a range of complex things with one simple type of thing (the NCC), we should not hope to get much of a good explanatory relation out of this kind of correlation.

Philosophers have not helped here, in their tendency to use words which no-one really understands, under the pretence that we all know what they mean. The word 'qualia' is the prime example. Philosophers tell us that consciousness involves qualia, and when asked what qualia are, they say, 'you know! The smell of coffee, the

taste of chocolate, the look of something red, the sound of birdsong'. And how could anyone deny that there are such things? As Ned Block once said, if you have got to ask what qualia are, you ain't never going to get to know.

But when you think about it, there is nothing that smelling coffee, tasting chocolate, seeing red, or hearing birdsong have in common, other than the fact that they are all conscious experiences. And if qualia just means 'conscious experience' then nothing is explained by saying that conscious experience involves qualia.

In fact, those who talk in terms of qualia do think that it means something other than 'conscious experience': they normally mean intrinsic, non-intentional (non-representational) properties of conscious experience. By being non-intentional, qualia normally get to be simple and unstructured, and so the perfect correlates for simple, homogenous neural correlates of consciousness. The qualia theory and the hunt for the NCC fit together very smoothly: find the intrinsically conscious properties of states of mind, the simple qualities that make them conscious, and then the scientific project is to pair them up with things in the brain. But if there are no such simple qualities, then the project cannot really get started.

The story goes that the logical positivist philosopher Herbert Feigl, the author of an important 1958 monograph defending a materialist approach to the mind-body problem, once gave a visiting lecture on the problem of consciousness at UCLA, where Rudolf Carnap was teaching. Feigl argued that although there were good reasons for believing that the mind is fundamentally physical, the physical explanation of the 'qualia' of sensory experience was still a mystery to science. Carnap is supposed to have interrupted, 'But Feigl, there is something missing from your lecture. Science is beginning to explain qualia in terms of the alpha factor!'. Feigl, alarmed by this interjection from the great Carnap, replied 'But Carnap, please

tell me: what is the alpha factor?'. 'Well, Feigl' Carnap replied 'if you tell me what qualia are, I'll tell you what the alpha factor is'. Although I do not share Carnap's reductionist attitude, I agree with him about qualia.

That's enough about obstacles. Our second task is to provide a perspicuous representation of states of consciousness themselves. What we need, first of all, is an understanding of the right ontological categories in terms of which to formulate the subject-matter of the mental. Should we, for example, be thinking in terms of conscious mental states, mental events, mental processes, or mental objects? Or what?

We need an adequate idea of the ultimate subject of consciousness. Philosophers are accustomed to talking of conscious *states* and conscious *processes*, as if what is conscious in the first place are states and processes. But a little reflection should tell us that this way of talking can mislead us. What is it, after all for a mental event – say, the event of me noticing a bird flying into its nest – to be a *conscious* event? Some say it is for *me* to be conscious of the event itself: this is the higher-order thought conception of consciousness. But this account gets things the wrong way around: I can become conscious of my noticing the bird *because* it is an event in my consciousness, not vice versa.

The truth of the matter is rather that this is a conscious mental event not because I am conscious of the mental event itself, but it is a conscious mental event because I am conscious of the bird flying into its nest. People's mental states are conscious because people themselves are conscious. It is the person which is the fundamental subject of consciousness.

This is why we should not say that the *brain* is conscious: that the brain thinks, feels, remembers, understands and so on. We would, of course, know what is meant if someone says something like this, but if we are aiming for total clarity, we

should recognise that this is just a way of talking. The 20th century American philosopher Roderick Chisholm put it well, when he said:

‘the brain is the organ of consciousness, not the subject of consciousness – unless I am myself my brain. The nose similarly is the organ of smell and not the subject of smell – unless I am myself my nose’.⁴

The implication is that the fact that the brain is organ of consciousness no more implies that the brain *is* conscious, than the fact that the nose is the organ of smell implies that the nose smells.

The same can be said of thought. The ultimate subject of thought is the human person, not the human brain. Noam Chomsky – a thinker as remote from Roderick Chisholm as could be imagined – comes surprisingly close to what Chisholm says on this matter, when he says that ‘people think, not their brains, which do not, though their brains provide the mechanisms of thought’⁵.

Here it would be helpful to employ a distinction introduced originally by Daniel Dennett, between mental states or events which are attributed to the whole person, and those that are attributed to some part or subsystem of the person. Dennett called this the distinction between personal and sub-personal mental states.

Within the category of personal mental states, there are distinctions we need to make between consciousness and self-consciousness. This relates to Damasio’s account of consciousness, and another way in which a more careful interaction with philosophical tradition would have helped him. Damasio makes a distinction between what he calls ‘core consciousness’ which provides an organism with a sense of self

⁴ R.M. Chisholm, ‘Questions about the Unity of Consciousness’

⁵Noam Chomsky, ‘Language and Nature’ *Mind* 1995, 8.

about the ‘here and now’, and ‘extended consciousness’ which gives the organism an ‘elaborate’ sense of self. The distinction is suggestive, but it raises more questions than it answers. Philosophers have for a long time operated with a distinction between consciousness and *self*-consciousness: to be conscious is for the world to be present to one’s mind, while to be self-conscious is to be aware (in some maybe in some very minimal way) of oneself. There may be a reason to reject this distinction; but Damasio does not give us one. Rather, in assuming that core consciousness itself involves a ‘sense of self’ he builds the rejection of the traditional distinction into his starting point. The point is not that he is wrong to do this – maybe he is right, maybe all consciousness involves minimal self-consciousness – it is rather that he shows no awareness that he is doing it at all.

Let me draw these various lines of thought together. In order for neuroscience and psychology to be able to give an account of the mechanisms of consciousness, they have to know what it is they are giving an account of. They must therefore appeal to the subjective knowledge which we have of our own mental states, and not be misled by false pictures of the phenomena, which make it look as if consciousness is a single simple quality which all conscious states of mind have in common. To avoid this, we have to look carefully at the actual content of this knowledge, and take seriously the phenomenological distinctions which a proper self-conscious reflection reveals to us.

Some writers talk as if solving the problem of consciousness will be an achievement rather like getting a man on the moon. If I am right in what I say here, this is not right. It would be closer to the truth to say that the scientific explanation of consciousness will be more like finding a cure for cancer than it is like getting a man on the moon. For just as there no *one* thing which deserves the name of

consciousness, there is no one thing which deserves the name of ‘the’ cure for cancer, and by all accounts it is unlikely that there will ever be such a thing. But there is a complex network of treatments, diagnoses, therapies and preventative measures, some specific to the various forms of the disease, some based on well-understood correlations and some on poorly-understood but effective shots in the dark. The result of all this is that a person’s chance of surviving cancer are better than they have ever been. This kind of complex network of empirical correlations and hypotheses is perhaps, a better model for the scientific explanation of consciousness.

Thomas Nagel, a philosopher whose views on consciousness have dominated the current debate, has argued that we need to develop new concepts if we are to solve the problem of consciousness. If I am right in what I have said today, this is a mistake. We do not need to develop new concepts. We already have all the concepts we need, but what we need is a correct understanding of them, and how they actually do apply to our conscious mental lives. No new conceptual or empirical discovery is needed, only a perspicuous re-arrangement of our actual knowledge of the mind and the brain.

Copenhagen

17 April 2007