Aping Mankind
by Raymond Tallis

Chapter Three
Neuromania: A Castle Built on Sand

Bold Claims

> [Neuro-talk] is often accompanied by a picture of a brain scan, that fast- acting solvent of critical faculties.[1]

It is surprising that the world has not wearied of stories of findings by neuroscientists that are supposed to cast light on our true nature. Popular articles - which are often heavily dependent on press releases provided by the public relations departments of grant-hungry laboratories - are usually accompanied, as we have noted, by a brain scan. These are seen as visible proof that those clever boffins have discovered the neural basis of love (maternal, romantic, unconditional), altruism, a propensity to incur toxic debts and so on. And that's just for starters. The sociologist Scott Vrecko has listed neurobiological accounts of (take a deep breath) in alphabetical order: altruism, borderline personality disorder, criminal behaviour, decision-making, empathy, fear, gut feelings, hope, impulsivity, judgement, love (see above for varieties of), motivation, neuroticism, problem gambling, racial bias, suicide, trust, violence, wisdom and zeal (religious). The extent of neuromanic imperialism is astounding. Before we examine the shaky general foundations of these claims, I cannot resist sharing some of my favourite examples with you, which you may wish to examine in more detail by looking at the original papers. They concern love, beauty and wisdom.

According to the neuroscientist Mario Beauregard, the truest form of love - truer than the interested love of those who hope to gain from their object, truer than maternal love, or truer even than romantic love - is the love that low-paid care assistants looking after people with learning disability feel for their charges. In a study entitled "The Neural Basis of Unconditional Love", care assistants were invited to look at pictures of people with intellectual disabilities first neutrally and then with a feeling of unconditional love. By subtracting the brain activity seen in the first situation from that seen in the second, the authors pinned down the neural network housing unconditional love. It was distinct from that which had previously been identified for romantic love and maternal love - although there was some overlap - and it included parts of the brain's "reward" system. This, Beauregard has argued, may be the link between reward and strong emotional links which (guess what?) "may contribute to the survival of the human species". Thus love (unconditional).

Next, beauty (aesthetic). You and I may feel that the impact on us of a work of art is deeply mysterious. Zeki and Hideaki Kawabata do not agree. A few years ago they reported that they had found the locus of our experience of the beauty of art. Their experimental design was marginally more sophisticated than the one that Beauregard used to peer into the souls of low-paid care workers. Subjects were scanned as they looked at pictures they had previously classified as "beautiful", "neutral" or "ugly". Their orbito-frontal cortex was more active when they were looking at beautiful pictures. *Voila!* The beauty spot.

Neuroscientists have also identified neural correlates of trust and of admiration but the big one, surely, must be the neural basis of wisdom and this, too, has revealed itself to the

neuroscientific gaze. “Scientists use brain scans to find the secret of what makes us wise”, Jonathan Leake reported in the *Sunday Times* newspaper. They did this by “pinpointing parts of the brain that guide us when we face difficult moral dilemmas”. This was the journalist’s take on an article published by Dilip Jeste and Thomas Meeks. The authors seemed a little more circumspect, noting that:

> the prefrontal cortex figures prominently in several wisdom subcomponents (e.g. emotional regulation, decision making, value relativism) primarily via top-down regulation of the limbic and striatal regions. The lateral prefrontal cortex facilitates calculated, reason-based decision making, whereas the medial prefrontal cortex is implicated in emotional valence and prosocial attitudes/behaviours. Reward neurocircuitry (ventral striatum, nucleus accumbens) also appears important for promoting prosocial attitudes/behaviours.[9]

This observation enabled them to construct a “speculative model of the neurobiology of wisdom”. It involves a large number of brain pathways but the key is an “optimal balance between functions of phylogenetically more primitive brain regions (limbic system) and newer ones (prefrontal cortex)”.

Following a familiar pattern, the "speculative” model was translated by journalists, with the help of a press release from the laboratory and rather optimistic interviews with the scientists, to an article headlined: "Found: The Brain’s Centre of Wisdom”. The tentative complex model in the original article was simplified to a matter of balance between “anterior cingulate cortex, linked with emotions” and the “prefrontal cortex”, which “governs conscious thought”. But we should not blame the journalists: they are not the only source of hype and journalism. “Knowledge of the underlying mechanisms in the brain”, Jeste said in an interview "could potentially lead to developing interventions for enhancing wisdom”.

It is easy to mock such BOLD aims. They seem like brochures from the Grand Academy of Lagado in *Gulliver’s Travels.* But we need also to specify what is wrong with them and why we should dismiss them as manifestations of what the professor of psychiatry William Uttal has termed “neo-phrenology": a recurrence of the claims of the eighteenth-century phrenologists we described in Chapter 1. They have two kinds of flaws. The first are technical: the limitations of fMRI, the design of the studies that use it and the way data are analysed. I shall discuss them in this section. Much more important, however, are flaws arising from conceptual errors, and I shall address these in the next section.

The first thing to remember when you come across headlines such as “Found: The Brain’s Centre of Wisdom” is that fMRI scanning doesn’t directly tap into brain activity. As you may recall from “You are your brain” in Chapter 1, fMRI registers it only indirectly by detecting the increases in blood flow needed to deliver additional oxygen to busy neurons. Given that neuronal activity lasts milliseconds, while detected changes in blood flow lag by 2-10 seconds, it is possible that the blood flow changes may be providing oxygen to more than one set of neuronal discharges. What is more, many *millions* of neurons have to be activated for a change in blood flow to be detected. Small groups of neurons whose activity elicits little change in blood flow, or a modest network of neurons linking large regions, or neurons acting more efficiently than others, may be of great importance but would be under-represented in the scan or not represented at all. In short, pretty well everything relevant to a given response at a given time might be invisible on an fMRI scan.

And then there is the almost laughable crudity of the design of the experiments that are used to support the conclusion that “This bit of the brain houses that bit of us”. They are mind-numbingly simplistic. We have already seen this in the case of studies looking for the “unconditional love spot” or the “beauty spot” or “the wisdom circuits". In a typical experiment, subjects are exposed to different stimuli, or asked to imagine certain scenarios, and the change in brain activity is recorded. Let me illustrate this with another

example: the work of Andreas Bartels and Zeki on love (romantic).

In these studies, they asked their subjects to look at a photograph of the face of someone with whom they were deeply in love and then at photographs of three friends. By subtracting the activity of the brain recorded when the subjects looked at their friends from that which was seen when they looked at their lovers, they claimed to be able to demonstrate the distinctive brain activity associated with love (romantic). On the basis of these experiments, Bartels and Zeki concluded that love (romantic) was due to activity in a highly restricted area of the brain: “in the medial insula and the anterior cingulate cortex and, subcortically, in the caudate nucleus and the putamen, all bilaterally". This caused them to wonder that “the face that launched a thousand ships should have done so through a limited expanse of the cortex”. I too feel wonder but for different reasons. How could anyone fail to see the fallacies in the experimental design?

What fallacies, you might ask. First, when it is stated that a particular part of the brain lights up in response to a particular stimulus, this is not the whole story. Much more of the brain is already active or lit up; all that can be observed is the *additional* activity associated with the stimulus. Minor changes noted diffusely are overlooked. Second, the additional activity can be identified only by a process of averaging the results of subtractions after the stimulus has been given repeatedly; variations in the response to successive stimuli are ironed out. The raw data tell a very different story from the cooked. If, to take a much simpler example, you offer a series of subjects the *same* spatial memory task, you will see enormous differences in the many areas that light up. Even simple actions are associated with highly variable responses. Jian Kong and colleagues found that when subjects were engaged in six sessions of a finger-tapping test, the test-retest correlation ranged between 0.76 and *zero* for the areas that showed significant activity in all sessions. For most of us, finger-tapping is less, rather than more, complex that being in love.

Which brings me to the third problem. The experiments looked at the response to very simple stimuli: for example, a picture of the face of a loved one compared with that of the face of one who is not loved. But as anyone knows who has been in love - indeed anyone who is not a Martian - love is not like a response to a simple stimulus such as a picture. It is not even a single enduring state, like being cold. It encompasses many things, including: not feeling in love at that moment; hunger; indifference; delight; wanting to be kind; wanting to impress; worrying over the logistics of meetings; lust; awe; surprise; joy; guilt; anger; jealousy; imagining conversations or events; speculating what the loved one is doing when one is not there; and so on. Likewise - to refer back to Beauregard’s study on what he calls "uncondi-tional love’’ - no one who has cared for someone with learning disability could see that reduced to a surge of emotion. That would hardly be sufficient even to support a surge of sentimentality at the *idea* of looking after someone who has special needs, never mind the 24/7 grind of actual hands-on care. (It is reassuring, perhaps, that only three out of the seven areas that Beauregard has reported as lighting up when carers looked at pictures of their potential charges coincided with those seen when romantic lovers looked at a picture of their beloved. If all seven had lit up, one might see recommendations for even more arduous Criminal Record Bureau checks.)

The same Martian tendency is evident in studies of the neurology of economic behaviour and, in particular, highly topical studies of the tendency to make unwise financial decisions. As we shall discuss in Chapter 8, neuro-economic researchers have determined that the reason subprime mortgages are so seductive, although the financial terms are so disadvantageous, is that they take advantage of our muddled brains. According to Samuel McClure and colleagues— there are separate “value” systems in the brain. How did they come to this conclusion? By looking at brain activity in individuals who were asked to choose between lesser but more immediate rewards and rewards that were greater but delayed. They demonstrated to their own satisfaction that the limbic system placed special weight on immediate rewards (even if they were smaller than delayed rewards), while the frontal lobes placed more weight on delayed rewards, if they were greater than immediate

ones: choosing two jars of jam tomorrow over one today. Sub-prime mortgages typically start with a very low interest rate, fixed for a couple of years, followed by a much higher (above the usual market) rate for the next quarter of a century or so. The first stage of the mortgage - in particular its immediate availability - appeals to the limbic cortex, while the second, much longer, stage should put off the frontal lobes. Unfortunately, in this competition within the brain the limbic circuit wins, because it houses automatic reward-seeking behaviour, which reflects evolutionary adaptations to those remote environments in which the human brain evolved as opposed to "the more recently evolved, uniquely human capacity for abstract ... reasoning and future planning".— As neuro-economist George Loewenstein (a collaborator on the McClure paper) has argued:

> Our emotions are like programs that evolved to make important and recurring decisions in our distant past. They are not always well suited to the decisions we make in modern life. It's important to know how our emotions lead us astray so that we can design incentives and programs to help compensate for our irrational biases.

The purchaser of "Chez Nous" is little different from Pleistocene man chasing a mammoth or, perhaps, requisitioning a cave with an en-suite midden.

As we shall see in "Neural political economics" in Chapter 8, this is but one of a whole raft of similar studies in neuro-economics. For the present, we note that only a behavioural economist would look to the fixed structures of the brain to explain a relatively new phenomenon such as the ready availability of mortgages to people who can't afford to pay them back. Its actual origins lie in a change of social attitudes towards debt, alterations in the financial regulatory system and political initiatives that began in the post-Pleistocene Jimmy Carter era. Only a behavioural economist would regard responses to a simple imaginary choice (between two relatively small sums of money - $5 and $40 offered immediately or in six weeks) as an adequate model for the complex business of securing a mortgage. Even the most foolish and "impulsive" mortgage decision requires an enormous amount of future planning, persistence, clerical activity, to-ing and fro-ing, and a clear determination to sustain you through the million little steps it involves. I would love to meet the limbic circuit that could drive all that.

The risible simplification of human behaviour seen in the studies of love, beauty, wisdom and (in the case of sub-prime mortgages) stupidity, reflected in their crude experimental design (which treats individuals as passive respondents to stimuli and then discovers that they are passive respondents to stimuli), is not the only empirical reason for treating fMRI with suspicion. A paper published a few years ago reported an extensive overlap between the neural circuits registering physical pain and those implicated in social pain; both pains seemed to "light up" the same areas. The authors (as have many others) have taken this as evidence that that the two are essentially the same, and have treated it as a great neuro-evolutionary discovery. For social animals like humans, so the story goes, the need for solidarity is served by making social exclusion painful and this requirement is met by employing circuitry that has already developed to register ordinary, physical pain. A more plausible interpretation, however, is that the failure to demonstrate fundamental differences between what you feel when you stub your toe and your feelings when you are blackballed by a club from which you are seeking membership is a measure of the limitations of fMRI scanning and, indeed, other modes of brain scanning.

I am not alone in questioning the validity of an approach that identifies activity in certain parts of the brain with aspects of the human psyche. In a controversial, but to me compelling, paper published in 2009 (originally provocatively entitled "Voodoo Correlations in Social Neuroscience"), the authors found serious problems with the localisations observed in such studies. The authors concluded that "in most of the studies that linked brain regions to feelings including social rejection, neuroticism and jealousy, researchers ... used a method that inflates the strength of the link between a brain region

and the emotion or behaviour".

One of the authors, Harold Pashler, is an experimental psychologist of the utmost distinction. He is the editor-in-chief of the major textbook of experimental psychology. The papers examined in the review had been published in top-rank journals, including *Science,* which is regarded as one of the two leading scientific publications in the world, the other being *Nature.* The authors observed that "a disturbingly large and quite prominent segment of fMRI scan research on emotion, personality and social cognition is using seriously defective research methods and producing a profusion of numbers that should not be believed".

So what problem had Pashler and his colleagues identified? They looked at the statistical methods used to derive correlations between activity in the brain and emotional states and found that the instruments used pretty well guaranteed high correlations between the variables observed. That such an elementary error should be allowed to pass on the nod is a measure of how the glamour of high science can disarm the most acute minds. Pashler and colleagues suggest that "the questionable analysis methods are also used in other fields where fMRI is used to study individual differences, such as cognitive neuroscience, clinical neuroscience and neurogenetics".

I first got wind of this article when *New Scientist* published a *mea culpa* editorial in 2009 about its own coverage of "breakthroughs" in understanding human beings arising from fMRI studies: "Some of the resulting headlines appeared in *New Scientist, so we* have to eat a little humble pie and resolve that the next time a sexy-sounding brain scan result appears we will strive to apply a little more scepticism to our coverage". And a more recent, admirably painstaking, review concludes that "the reliability of fMRI scanning is not high compared to other scientific measures"; moreover, there is no agreement as to what would count as a measure of reliability; and, finally, reliability is even worse in studies of higher cognitive tasks (experiencing beauty, deploying wisdom, being stupid) than in the case of simple motor or sensory tasks - in short, in the case of those papers that have made the popular press go pop-eyed with excitement.

The allocation of human faculties and sentiments to different parts of the brain is also being increasingly undermined by evidence that even the simplest of tasks - never mind negotiating a way through the world, deciding to go for a mortgage or resolving to behave sensibly - require *the brain to function as an integrated unit.* As David Dobbs has pointed out, fMRI scanning "overlooks the networked or distributed nature of the brain's workings, emphasising localized activity when it is communication among regions that is most critical to mental function". I shall return to this in a moment.

Although the spatial resolution of scanners is improving all the time, increasing the resolution does not solve the problems we have discussed. Normally fMRI scanning looks at cubes of tissue - three-dimensional pixels (called "voxels") - each of which comprises hundreds of thousands of neurons. It is now possible to examine finegrained patterns within voxels. Rees has used this technique to examine aspects of visual perception. You might recall from Chapter 1 that Hubei and Wiesel found certain cells in the visual cortex responding preferentially to lines presented at a certain orientation. By studying the fine grain of the fMRI in this area when subjects are looking at lines with different orientations, Rees and colleagues were able to infer the orientation of the presented line with 85 per cent accuracy: in other words, they were able to work out what the subject was looking at. This, however, is a far cry from examining the experience of an entire object, of an entire scene, of a changing scene, or of the changing meaning of a scene, never mind complex segments of people's lives as when, for example, they decide to take on a mortgage or fall in love. The claim that it is possible to look at a single fMRI image and see what the person is seeing, never mind what they are feeling, and how it fits into their day, or their life, is grossly overstating what can be achieved. Ordinary consciousness and ordinary life lie beyond the reach of imaging technologies, except in the imagination of neuromaniacs.

The technical limitations of fMRI are compounded by conceptual limitations. Some of

these are so fundamental that they are properly the object of philosophical treatment and I shall address them in the next section. Others, however, relate to the neuroscientific framework. The reader will recall the centuries-long debate, discussed in “You are your brain” in Chapter 1, about the modularity of the brain, triggered by the phrenologists in the late eighteenth and early nineteenth centuries. The same debate is hotting up again. You have only to read a few papers on the correlation between this function (e.g. mortgage-buying) and that structure in the brain (e.g. the frontal lobes) to start to notice that certain parts of the cortex appear again and again, serving quite disparate functions. You could be forgiven for thinking of the brain as being managed by a crooked estate agent letting out the same bit of real estate simultaneously to different clients. What is more, not only do certain brain regions serve multiple cognitive functions, but the same cognitive functions may activate the different regions of the brain. Not that this is surprising, given that the brain, ultimately, must work as a whole. Love (romantic), for example, involves a multitude of things: emotions, intentions, the motor activity necessary to buy flowers or to make a pass, and the long narrative with one’s self and real and imaginary conversations with the object of one’s affection. The point is this: the more you think about the idea that human life can be parcelled out into discrete functions that are allocated to their own bits of the brain, the more absurd it seems.

And it seems even more absurd in the light of what is accepted about something as seemingly simple as individual memories. According to Antonio Damasio:

> The brain forms memories in a highly distributed manner. Take, for instance, the memory of a hammer. There is no single place in our brain where you can hold *the* record for a hammer ... there are several records in our brain that correspond to different aspects of our past interaction with hammers ... and all... based on separate neural sites located in separate neural regions.

Even more telling is the observation made by Marcus Raichle and collaborators. They used another form of imaging called positron emission tomography (PET) scanning and found that learning something as elementary as the association of a word such as “chair” with “sits” involved not only the language centre in the left hemisphere but extensive stretches of the so- called “silent” areas of the frontal lobes and the parietal cortex. What hope is there, then, of locating something as global and untidy as my love for someone in a neatly demarcated area of the brain? None, I am pleased to conclude.

Observations of this kind have led some scientists, such as Karl Friston (who played a key role in developing neuro imaging techniques), to suggest that “the brain acts more as if the arrival of ... inputs provokes a widespread disturbance in some already existing state”, rather as happens when a pebble is dropped in a pond.— So we need to take the reports about “beauty spots” and “centres for unconditional love” from “leading” scientists with more than a pinch of salt. Neuroscientists who think they have found the circuits in the brain corresponding to wisdom seem to lack that very quality, as a result of which they are oblivious even to what the more critical minds in their own discipline are saying.

The current technical limitations of neuro imaging do not, however, support a *principled* objection to the idea that we can directly observe human consciousness - our experiences, motivations, intentions, emotions and propensities - in the brain. After all, some will argue, most imaging techniques are only a few decades old and they are improving rapidly. Soon we shall be able to track what is happening in the brain of waking, living subjects like ourselves, with a spatial and temporal resolution that will enable us to see precisely which neurons are firing, when and in response to what. Zeki and Goodenough anticipate the coming of "extremely high resolution scanners that can simultaneously track the neural activity and connectivity of every neuron in the human brain, along with computers and software that can analyse and organise these data”. Leaving aside the question of how the computers could be programmed to oversee the

infinite number of combinations of the activity of a brain with a trillion neurons and more potential connections than there are atoms in the universe, might it not be possible, although it may be difficult to imagine, to pick out the subsets of activity relevant to individual thoughts, or sharply define the boundaries of neural excitement implicated in particular feelings, or identify the particular weighting of different locations in the brain in determining character traits?

Let us suppose there were no limit to the precision of imaging. Let us suppose also that the kinds of localizations seen on fMRI scans, which have caused so much excitement, were robust. And let us suppose that the separate psychological states or functions to which the brain activity is supposed to correspond are real entities rather than *ad hoc* constructions. And let us, finally, suppose that we have explained how that which has been teased apart comes together in the conscious moment: something to which we shall return below, in "Brain science and human consciousness, II". What, then, would fMRI tell us? If we could obtain a complete record of all neural activity, and we were able to see the firing state of every individual neuron, would this advance our understanding in the slightest? Would the record of neural activity be as useless at telling us what it is like to be conscious as a complete print out of his genome at telling you what it is like to be with your friend? Would (human) consciousness be - to use Dennett's boastful term - "explained"? Would we be able directly to observe human consciousness and find out what is "really" going on when we experience the world, judge it and act upon it?

For this to be the case, one thing at least would be necessary: we would have to be sure that the neural activity we observed was in some strict sense *identical* with consciousness. Does the new neuroscience allow us to make that assumption and accept Hippocrates' conjecture as proved beyond reasonable doubt? To answer this question we need to move on from the technical limits and methodological muddles of scan-based cognitive neuroscience to the conceptual, indeed philosophical, problems that Neuromania ignores.

The Leap From Neuroscience to Metaphysics

I am now going to argue that neuroscience does not address, even less answer, the fundamental question of the relation(s) between matter and mind, body and mind, or brain and mind. If it seems to do so this is only the result of a confusion between, indeed a conflation of, three quite different relations: correlation, causation and identity.

Consider the research we have been discussing, based on fMRI. Typically, brain scanning reveals (rather wobbly and definitely loose, as we have seen) *correlations* between (say) the experience of seeing some item such as a loved one's face and activity in some part or other of the nervous system. Does this mean that what we see on the brain scan is either the cause of the experience or even identical with it? No, because a correlation is not a cause: even less is it an identity. Seeing correlations between event A (neural activity) and event B (say, reported experience) is not the same as seeing event B when you are seeing event A. Neuromaniacs, however, argue, or rather assume, that the close correlation between events A and B means that they are essentially the same thing.

The most obvious trouble, with the view that neural activity on the one hand, and experiences on the other, are the same thing is that they should appear like one another. But nothing could be further from the truth. The colour yellow, or more precisely the experience of the colour yellow, and neural activity in the relevant part of the visual cortex, however it is presented, look not in the slightest bit similar. There is nothing yellow about the nerve impulses and nothing nerve-impulse-like about yellow. If, however, they *were* the same thing, the least one might expect is that they would appear as if they were the same thing. Surely, it is not too much to expect that something should look like itself. As it is, nerve impulses seem required to have two sets of appearances at the same time that are profoundly different from one another: an appearance as electrochemical activity (of which

more below in “Why there can never be a brain science of consciousness”) and an appearance as an experience - of something other than themselves, such as the colour yellow belonging to an object.

The more philosophically astute neuromaniacs are not, of course, unaware of this difficulty and have found different ways of getting round it. The most popular tactic, and *primafacie* the most plausible, is to assert that experiences (such as the colour yellow) and the neural activity seen in the visual cortex in association with that experience are two *aspects* of the same item. This is the so-called “double-aspect” theory. While there is only one set of events - what we see in the brain - these events have two sides: a neural side and an experiential side. There are many objections to this ploy.

The first becomes apparent when we ask what is meant by "aspects” or “sides”. We know what it is like for an object, such as a house, to have one aspect when it is looked at from behind and another aspect when it is looked at from the front. But we cannot imagine any kind of entity that has an experiential (or mental) front end and a neural (or material) back end. The same objection applies if, instead of “front" and “back”, we speak of “top” and "bottom” or “inside" and “outside”.

We could summarize the failure of the double-aspect theory by saying that the difference between different aspects of a house - between the front and the back - is nothing like the difference between a material event such as a discharge of nerve impulses and a conscious event such as having the experience of yellow. What is more, the notion of two aspects of a house presupposes observers who see the house from different angles. The house does not, in or of itself, have two aspects or indeed any aspects. This touches on the most profound problem with the assumption of identity between neural activity and consciousness, and we shall return to this below in “Why there can never be a brain science of consciousness”. For the present, it is necessary only to note that we cannot invoke (implicitly conscious) observers to generate the two aspects of the events detected by neuro imaging - the neural activity and the experience - in order to explain how (material) neural activity is also (conscious) experience. To invoke doubled aspects is to cheat: it smuggles consciousness in to explain how it is that neural activity, which does not look like experience, actually *is* such experience.

This is a point that is overlooked even by the most thoughtful and sensible philosophers, for example John Searle, the scourge of much sloppy thinking in this area. Searle has his own version of the dual-aspect theory. Water, he says, is identical with H2O molecules and yet they appear quite different. H2O molecules are not shiny and slippery like water. And this is how it is, he says, with neural activity and consciousness: consciousness is made up of experiences, such as that of yellow, which are nothing like nerve impulses but are nonetheless the same as nerve impulses. Stripped to its bare bones, Searle offers us an analogy:

Water is to H2O molecules as conscious experience is to neural activity.

Or Water: H2O:: conscious experience: neural activity

In both cases, he argues, the large-scale phenomena (consciousness, drops of water) are identical with the small-scale phenomena (nerve impulses, molecules of H2O.)

This analogy is false. The reason it does not hold up is the reason we gave just now for the failure of all double-aspect theories: both shiny water and H2O molecules require *observation* in order to be revealed as one or the other. They correspond to two different modes of observation: one by our ordinary unenhanced senses (introspecting experience, sensing water); the other by means of complex equipment and representations and interpretations that render H2O molecules “visible” and brain activity recordable. The two aspects of water are two appearances, two modes of experiencing it, and this hardly applies to neural activity as electrochemical activity and as experience.

Searle’s error is interesting, not just because it is perpetrated by a philosopher who thinks hard, writes lucidly and does not lose sight of common sense (something, by the way, for which he has been criticized), but because he compounds it in a particularly

revealing way. He argues that molecules of H2O, as revealed through science, and water as we directly experience it are not only the same thing but that they stand *in a causal relation to one another,* and this is how it is with nerve impulses, which have the same kind of causal relation to conscious experiences. The molecules of H2O, he says, *cause* the appearances that we associate with water as we encounter it in our everyday lives; and, likewise, nerve impulses *cause* conscious experiences. This is, of course, incompatible with the notion that they are the same thing. We cannot say that A is the same as B *and* that A causes B, because cause and effect are, by definition, different items; and so, too, are the molecular and macroscopic appearances of water, respectively. (The only item I can think of as being the cause of itself is the God of monotheistic religions.) Nor can we see one aspect of an object causing another aspect: they are present, simultaneously, side by side, so one cannot be the product of another. The inside of a house cannot be caused by the outside any more than the latter can be caused by the former. Both, of course, require another cause: observers who see the house from different angles.

When a philosopher as gifted as Searle makes such an elementary mistake, it must be because he is in the grip of an intuition that is hidden from him, although it is directing his thought. The intuition is worth exploring because doing so should help to pre-empt its casting its spell on us. Searle thinks that H2O molecules cause the experience of dampness and shininess because he thinks of the dampness and so on as the macroscopic *appearance* of large aggregations of molecules. This is wrong for the reason we have already pointed out; namely that H2O molecules - as an array of triplets of atoms - are already themselves a kind of appearance, although one that is mediated by scientific instruments and measurements and theories in the way that the shininess of a pool of water is not. If we deny that the individual molecules have an appearance at all - arguing that they are simply inferred from measurements, for example - then we arrive at an interesting result. Water, as we see it in everyday life outside the laboratory, is the appearance of large quantities of something - molecules of H2O - that do not have an appearance in everyday life. Their representations in physics are a borrowed appearance. If this is accepted then we have to ask this question: what it is that gives the molecules an appearance at all? The answer to this will be the same as the answer to the question as to what it is that brings microscopic molecules together into a macroscopic patch or stretch of water of the kind that we see is shiny or feel as damp. And it is, of course, *a conscious observer,* or conscious "experiencer". The water looks as it does - indeed has a look - because someone is conscious of it.

What Searle has done is to move the relation between the water and a creature such as a human being aware of it into a causal relation between (a) what water is reduced to in the eyes of physical science - molecules of H2O - and (b) an appearance that it supposedly has *in itself.* This enables him, without being fully aware of it, to smuggle in the consciousness he needs in order to get from nerve impulses to experiences and hence to make nerve impulses plausible as the basis for experience. While molecules of H2O are of course necessary for the experience of the shiny stuff that is water, they do not of themselves create that experience. They are necessary but not sufficient. The shiny appearance, the damp or liquid feel, requires in addition a conscious observer. And so, also, does the appearance of water as an array of H2O molecules. What we are referring to when we talk about macroscopic pools of water that are shiny, and molecules that are not, are different ways of *experiencing* water: the direct, everyday experience and the molecular experience mediated through instruments. The relation between these two ways of appearing cannot be a model of the relation between nerve impulses and appearances, and even less an explanation of how nerve impulses can be both propagated waves of electrochemical activity and, say, the experience of yellow.

Searle, therefore, is not different from many other thinkers of a neurophilosophical persuasion, in taking the correlation between neural activity and reported experience to mean that there is an intimate causal relation between them: nerve impulses *cause*

consciousness. And, like many others, he also believes that nerve impulses (or some of them at any rate) *are* (identical with) consciousness. What makes his position particularly illuminating is that he holds both of these incorrect, and also incompatible, views at once. It is, however, possible to be a little more choosy and many writers opt for the idea that nerve impulses *cause* consciousness, period: experiences are distinct from nerve impulses but are the effects of them. Although this view runs at once into insuperable difficulties, to which I shall return, it is worth reminding ourselves why it seems so attractive.

I flash a light into your eye while I record activity in the visual cortex using my latest scanner. Following the flash I see a burst of impulses passing up the optic nerve and into the cortex. At some point, as this burst is spreading across your cortex, you report an *experience* of a flash of light. I note also a close association between the intensity of the light to which you are exposed, the amount of activity in the relevant neurons and the reported intensity of your experience. This seems to demonstrate beyond doubt that the light causes the nerve impulses and the nerve impulses cause the experience of light; in short that the nerve impulses are the means by which light energy is changed into experience of light energy. Two other kinds of observation, to which I have already referred, seem to place this conclusion beyond doubt.

First, it is possible to prevent the experience by various means. If I interpose a screen between your eyes and the source of the light, blindfold your eyes or damage the pathways taken by the nerve impulses into the brain and within it then you do not experience the light. This is indirect evidence of the causal chain; if the putative causal chain is broken, then the experience is not had. And for some, this is conclusive proof that mind and brain are one. The neuropsychologist Bruce Hood is speaking for most cognitive neuroscientists when he says: "We know that damage to certain parts of the brain produces characteristic changes in the mind. It's one of the reasons most psychologists are not dualists: they are very familiar with the idea that the mind is a product of the brain." The slither in the logic is plain. We shall return in the last chapter to the (incorrect) notion that the only alternative to accepting that the mind is identical with, or caused by, brain activity is dualism. But let us look a bit more closely at the claim that brain-damage studies should oblige us to conclude that "the mind is a product of the brain".

The correct conclusion from the evidence provided by brain damage, or indeed from less dramatic events such as closing your eyes, or covering your ears, or turning your head away, or indeed moving to another place, is that the brain is a necessary condition of experience and a brain in the right place is a *necessary* condition of experiencing that place. For example, it seems that provoking neural activity in the right place is a necessary condition of experiencing the light. A necessary condition is not, however, a *sufficient* condition. Now the difference between necessary and sufficient conditions and, indeed, between conditions and causes is very difficult to capture precisely, although it has stimulated a large philosophical literature. Let me, however, illustrate the difference with a simple example. In order for me to be knocked down by a bus in London, it is necessary for me to be in London. It is, however, not sufficient; otherwise I would avoid the place more than I do. If, however, nerve impulses in a particular part of the brain were *identical* with certain experiences then they would not only be a necessary condition but also a sufficient condition. You could not have the experience without the nerve impulses and, more importantly, you could not have the nerve impulses without having the experience. It should not matter how those nerve impulses arise. Now some observations do indeed seem to support the notion that the nerve impulses are a sufficient condition for the experience, and this would be consistent with the impulses being identical with the experiences. Here is an example from my own work as a clinician.

For many years, I was responsible for running an electromyography clinic. One of my tasks was to diagnose patients with damage to the peripheral nerves: the ones that go down the limbs to the toes and fingertips. The method consisted of electrically stimulating the nerves near the end of the limb and recording the response higher up, to see how big it was

and how fast it travelled. When the nerves were stimulated the patient felt a tingling. This might suggest that nerve activity alone could produce conscious experience. Even more impressive were the testimonies of some of my patients with epilepsy. Epilepsy, you may recall from Chapter 1, is a condition in which there are, from time to time, bursts of highly synchronized abnormal electrical activity occurring spontaneously in the brain. These cut right across the activity associated with normal function, and their usual effect is to disrupt consciousness (which may be lost or in some other way impaired) or replace voluntary activity with involuntary activity (so that the person falls to the ground, sometimes twitching, or engages in automatic behaviour). Some of my patients, however, had forms of epilepsy affecting the temporal and parietal association areas of the cerebral cortex. These resulted in very complex, formed images or indeed entire scenarios. Sometimes they are prolonged - so-called *status epilepticus* - and they may be mistaken for dreams. Removing the particular part of the brain affected by the abnormal neural activity gets rid of the hallucinations. Does this not suggest that the stand-alone brain has the wherewithal to generate at least fragments of consciousness on its own: that, in other words, its activity is a sufficient, as well as a necessary, condition of perceptual experiences; that experiences *are* neural activity?

Even more challenging are some observations made by the Canadian neurosurgeon Penfield that I mentioned in “You are your brain” in Chapter 1. Penfield, it will be recalled, pioneered neurosurgical techniques for treating intractable epilepsy by removing foci of irritable tissue in the parts of the brain where the seizures originate. Since it was vital not to cut out structures essential for speech and for other key functions, the operations were carried out in waking patients (the brain itself does not experience pain). Prior to the excision, Penfield mapped the location of different functions in the brain using stimulating electrodes. When he stimulated the temporal lobe and the hippocampus some patients re-experienced fragments of their past. A patient might feel himself eavesdropping on a familiar scene, for example, the voice of someone calling her child, or the arrival of a travelling circus in town. This, again, might seem to support the belief that the stand-alone brain could be the basis for complex consciousness.

Such observations - and others, for example the hallucinations experienced when the brain is affected by psychoactive drugs - underpin a famous thought experiment, which in turn inspired an even more famous film. The thought experiment was that of “The Brain in the Vat”, proposed by Hilary Putnam and the film was *The Matrix,* of which I have heard enough to know that I do not want to see it. Putnam’s thought experiment - which was designed in part to refute the idea that “meanings are in the head (or brain)”, something that need not concern us here - went as follows. Since neural activity seems to be sufficient for experience, and it does not seem to matter how the neural activity is triggered, is it not possible that we are deceived as to our true nature? If we were brains suspended in a vat of nutrient liquid, so that they could function adequately, and these brains were stimulated electrically under the guidance of supercomputers, would it not be possible to have the entire range of experiences that we have now? How could we tell that these experiences were not of a real world? Might not a computer regulate the activity of the brain such that I, the brain-owner, might have the impression of being surrounded by a world very like the one in which you and I are currently located? If this were possible, then all sorts of sceptical concerns about the world we are currently experiencing would be justified. Is this world, after all, a mere construct out of nerve impulses?

This thought experiment is valuable not just for the reason that Putnam introduced it. He wanted to argue that one could not have the thought "I might be a brain in a vat” unless there *were* external objects such as brains, vats, laboratories and scientists, and so, in short, a real world rather than one that was hallucinated by the brain. Well, I don’t think many of us needed persuading that words would not have meaning if no real referent corresponded to them and there was no world in which we were together with others. In other words, a brain in a vat would require a community of minds in a real outside world

for the experiment to be imagined, never mind to be set up. No, it is valuable because it demonstrates the absurdity of moving from the observation that neural activity is *correlated* with experiences to the conclusion that neural activity is not only a *necessary* condition of experiences but that it is a *sufficient* condition of them and may indeed be identical with them. This way lies the madness of concluding that a stand-alone brain could sustain a sense of a world. (The tendency to think of the brain as something stand-alone is reinforced by cognitive science, which imagines that what goes on in the brain are "representations' that are uncoupled from the world and are manipulated by the model-making brain.)

Be that as it may, neither the experiences of people with epilepsy nor Penfield's observations justify entertaining the possibility that we might be a brain in a vat or, more to the point, that the stand-alone brain can create a world and that neural activity would be not only a necessary but actually a sufficient condition of consciousness. Take the "memories" reported by Penfield's patients when they are stimulated (seen, by the way, in only 5 per cent of his subjects and not readily replicated by contemporary surgeons): they are essentially second-hand or recycled memories. No one who had not already had any experiences by the usual route, and had remembered them in the conventional way, would interpret what was happening as a *memory,* even less as a memory with a particular significance, meaning or reference to something other than themselves. The Penfield phenomena, like the pseudo-experiences of epilepsy, are simply *re-activations* of real memories of experiences had in the real world: had, by the way, not by an isolated brain but by a person. The electrical activity in the isolated brain appears to have the "aboutness" or intentionality of normal experiences (of which more presently) only because under all other circumstances (when the patient is not having a seizure or undergoing electrical stimulation) the experiences are genuinely of something that is really "out there", really happening, to a real person. As Sven Pfeiffer has pointed out to me, Penfield's patients are "awake, conscious and living before and while they are being stimulated". This existential and cognitive background is taken for granted but it undermines the claim that the neural activity in a stand-alone brain is, or could be, sufficient for consciousness: that brain stimulation is producing genuine stand-alone experience.

To look ahead somewhat, it is necessary to appreciate that our ordinary memories, and our ordinary current experiences, make sense because they are part of a *world.* Yes, we are located in this world in virtue of being embodied and we access it through our brains; but it makes sense to us, as a *world,* not solely on account of its physical properties but as a network of significances upheld by the community of minds of which we individually are only a part. The brain in the vat thought experiment helps itself free of charge to this world: a world, incidentally, in which, in addition to electrode-induced experiences, there really are material brains, electrodes, vats, scientists and the institutions, practices and know-how that support them. The hallucinations induced in the stand-alone brain by electrical stimulation or epilepsy also seem to make brain electricity a sufficient condition, or cause, of experience only because they, too, parasitize a real world experienced in the usual way.

The fact that neural activity is only a necessary and not a sufficient condition of consciousness is consistent with the observation that a person's behaviour becomes more completely explicable in neurological terms the more damaged they are. A seizure sits more comfortably within the neural model of mind than does *living with epilepsy,* which requires something to bring it all together.

And of the necessity for cerebral activity I have no doubt. My entire career as a doctor with a special interest in neurological diseases such as stroke and epilepsy has been a reminder of the extent to which our functioning as persons is vulnerable to the failures of our body. The distinction between necessary and sufficient conditions is a way of highlighting the fact that, even if the neuroscientific picture were complete, along the lines I have indicated just now, we would not have achieved an explanation of consciousness.

Nor should we expect to do so, since neuroscience is itself a late manifestation of consciousness. What is more, as we shall see below in “Why there can never be a brain science of consciousness”, physical science, to which neural impulses ultimately belong, does not have any place for consciously experienced appearances. A neural account of consciousness is *a contradiction in terms.*

We have some way to travel before we arrive at this conclusion. I want first of all to focus on different aspects or layers of human consciousness. I shall begin with sentience - the ground floor of consciousness, and something we may share with beasts - and then I shall examine higher-level or more organized aspects of consciousness, many of which we most certainly do not share with beasts. The sharpest division between the levels is signalled by the emergence of what I characterize as "full-blown” intentionality, or the "aboutness" of consciousness: something that is at once so simple and yet of such profound importance that it underpins the unique complexity of human life and our distance from all other living creatures.

Brain Science and Human Consciousness, I: Problems with Sensations

The errors of muddling correlation with causation, necessary condition with sufficient causation, and sufficient causation with identity lie at the heart of the neuromaniac’s basic assumption that consciousness and nerve impulses are one and the same, and that (to echo a commonly used formulation) "the mind is a creation of the brain”. There are, however, many other reasons for rejecting this belief and they apply to several distinct problems relating to physical explanations of consciousness: how matter became or relates to basic sentience; and how it is that certain material objects (such as you and I) are self- aware, how they are a subject of concern for ourselves. I shall begin, as I said, with the ground floor of consciousness: with qualia.

Qualia are the very fabric of consciousness: the material of experience, of the “what-it-is-like” feel of mental states. Although experience is gathered up into various kinds of wholes - objects, fields, situations, worlds - it is possible to pick out individual qualia to exemplify the notion. And so, somewhat at random, I pluck out from my rich sensory field the sound of a violin playing, the blackness of the letters growing across the screen, the feeling of pressure on my buttocks, the redness of a hat next to my computer, the sensations associated with a present anxiety. If *these* components cannot be identified with nerve impulses then no aspect of human consciousness can. So let me set out some of the problems that arise when one tries to identify qualia with nerve impulses.

The most fundamental and obvious problem is one that we have touched on already; namely, that nerve impulses *are not at all like qualia.* Those impulses in the visual cortex do not look like, say, the colour or shape or size of my red hat. We have seen how some philosophers have tried to deal with this by suggesting that what we see on a brain scan or an EEG is only one aspect of the neural activity and that consciousness is another aspect. This does not make the identity between neural activity and conscious experiences any more plausible because the very notion of "aspects” presupposes consciousness: an observer looking at something from a particular angle or in a particular way (as when it is examined through the lens of instruments, concepts and theories). But let us imagine it makes sense to think of a nerve impulse having an appearance in the absence of someone to whom it can appear. How would the intrinsic appearance of the nerve impulse relate to the experiences that it is supposed to embody? Not very well, it would seem. If we think of the nerve impulse as it appears to the observing neuroscientist, then we are really stumped. You will recall from Chapter 1 that it consists of sodium and other ions fluxing in and out of semi-permeable membranes. These do not seem like anything that is revealed in our ordinary experience of the world. And yet, if Neuromania is correct, they have to be the

intermediary through which the world - for example my red hat - is revealed to me. More generally, those sodium and other ionic fluxes have to be the appearance of the world to me.

This brings me to another problem. The trigger for the nerve impulses in virtue of which I am supposed to be aware of my red hat is not the hat itself: or not directly anyway. It is *light,* whose spectral frequency and patterns of distribution have been altered by running into my red hat. The neural activity is a response to this interfered-with light, and from this neural activity I can infer what it was that was interfered with by the light. This reaching back from the light to an object that interfered with it is something I shall come to in the next section when I talk about "intentionality". For the present, however, let me focus on the light itself.

Physics tells us that light is electromagnetic radiation and this does not in itself have a colour or, necessarily, visibility. Yellow-in-itself is not actually yellow; and electromagnetic radiation outside a very narrow bandwidth is actually invisible. Only an appropriately tuned perceiver can confer brightness, colour and beauty on light. Neurophilosophers have to believe that it is in nerve impulses, which have no appearance in themselves, that light energy acquires an appearance. Let us consider something else very elementary: heat. An increase in the rate of jigging of atoms (heat as seen by physicists) is not itself a hotting up: the transformation of jigging into an experience of *heat* requires something else - again, a conscious subject. A dispassionate examination of nerve impulses would not lead one to the conclusion that they could carry out this miraculous transformation: that they are capable of conferring the appearance of warmth on faster jigging; that electrochemical waves in nerve fibres, despite being items in the material world, are nonetheless able to confer appearance on the environing material world.

One way of getting a handle on the difference between nerve impulses and experiences is to try out the following well-known thought experiment. Imagine there was a device called an autocerebroscope that enabled us to see our own neural activity online as it occurs. Supposing I were able to look at the part of the brain where the neural activity corresponding to my seeing my neural activity through the autocerebroscope was happening. I would be seeing the neural activity and at the same time having the experience of seeing the neural activity. My experience would be that of someone seeing the activity from the outside and yet the activity would simply be itself, not itself seen from the outside. Or the activity would have to be both the experience *and* the experience of seeing itself from the outside: it would have to be at once subjective experience and an objective experience of the basis of the subjective experience. This would of course be impossible: it has to be one or the other but not both. What's more, the activity I can see through the autocerebroscope could be seen by someone else, whereas my experiences could be experienced only by me. Clearly an item cannot at the same time be something that can be visible to others as well as myself and something that cannot be experienced by others.

At the risk of making you dizzy, let me pursue this a bit further. Someone might object by saying that the nerve impulses I am looking at are not the same as the nerve impulses associated with my seeing the nerve impulses, which is something else that someone might share. Perhaps not: other nerve impulses are involved in my experience of seeing the nerve impulses. This, however, only moves the problem on, because those other impulses are also in principle visible to other people, while the experience they are supposed to be identical with is not. What this illustrates is that there is a gap, which cannot be closed, between experience and that which neuroscience observes; between experiences and nerve impulses. *Touché.*

All right, someone might say, mysterious and even paradoxical though the idea of the neural theory of consciousness might be, this is how things are. Get over it, accept it, believe it. Well, there are other problems that make me disinclined to just "get over it", most strikingly this one: there is a monotonous similarity about neural activity throughout

the cerebral cortex and yet it is supposed to underpin the infinite richness of phenomenal consciousness. How is this possible? There have been two kinds of explanation of how the nervous system creates or reconstructs the variety of the experienced world in the monotonous language of nerve impulses. The first appeals to differences of *location* in the brain; and the second to *patterns* of activity. Let's deal with location first.

Neural activity associated with the experience of different colours, or sounds versus colours, or with sense experiences versus memories, is located in different places in the nervous system. Although nerve impulses look the same, they are not the same when they are located in different places. Now, I don't know how it strikes you, but different locations don't seem to me to deliver what is needed. Why should the fact that a shower of nerve impulses is located two centimetres from another shower be sufficient to explain how one is the basis of a sense of disgust when faced with a bad smell and the other the feeling of pleasure given by contemplating that one's child has got into university. Why does this look even remotely plausible? Is it because we already know that there are certain functions partitioned in the brain: there are sense organs, nerve pathways, and sections of the brain devoted to particular aspects of the experienced world - say sight as opposed to hearing? This makes us inclined to say that the reason that neurons in the ventral visual pathways (interacting with the prefrontal and parietal cortex) give rise to visual awareness (as opposed to sounds or smells) is because these fibres are ultimately connected to the eyes. This is Muller's "doctrine of specific energies" that we referred to in "Neuroscience" in Chapter 1. Any stimulus to the eyes results in visual experiences; so I have sensations of light even when I stimulate my retina mechanically by pressing my eyeball, a stimulus unrelated to light.

This seemingly common-sense response is actually circular. Or it restates, rather than explains, the problem. The nerve impulses originating from the eyes give rise to visual consciousness because they are linked to central structures associated with visual consciousness and these centres experience visual consciousness because they are linked to the eyes. The fact that neither the light, nor the nerve impulses that are triggered by it, has an intrinsic appearance (of any sort, including that of visible light) shows how empty this circular explanation is. The eyes may respond primarily to light energy but this does not explain how it is that electromagnetic energy is translated into experienced light, into colour and brightness and so on. Different wirings - to the eyes or the ears or the nose - do not explain different experiences, particularly since, whatever energies land on sense endings, they are all translated into the same kind of energy: the electrochemical energy of nerve impulses. While each sense organ may be tuned to a different kind of energy - the eye to electromagnetic radiation and the ear to vibrations in the air - each translates those "specific energies" into the same language of propagated electrochemical disturbances. So much for the appeal to location.

Others have suggested that the differences that underpin the difference in experiences are to be found not in individual nerve impulses simply added up but in the hugely varied *patterns* of neural activity. There is potentially an infinite variety of patterns of nerve impulses: their numbers are not restricted, like the numbers of locations in the nervous system. It is in different patterns that we must find the difference between the experience of the red of a red hat, the experience of the hat as an object, the sense that to wear it would be a good idea, the emotional investment in the hat, and so on. But this "explanation" fails for the same reason as the supposed explanation by location: why should particular patterns correspond to experiences of material events - such as the interaction of electromagnetic radiation with material objects - that do not themselves have anything in them corresponding to those experiences? And there is another problem with explaining the variety of subjective experiences on the basis of the variety of patterns; this is the assumption that patterns somehow pick out themselves, add themselves up, know themselves. However, patterns of material objects or events, like aspects, have to be picked out by something else: *by a conscious observer.*

Let me illustrate this point with a simple example. Take a square consisting of nine letters:

T T T
T T T
T T T

This could be seen as three vertical rows each of three letters, three horizontal rows each of three letters, a group of six letters plus a group of three letters or a single group of nine letters. There are many other possibilities. What this variety tells us is not that the array left to itself contains all of these patterns inherently but that it contains only the *possibility* of these patterns, and not, for example, other possibilities such as a pattern consisting of two groups of six letters. The possibilities will be actualized, however, only by a conscious observer. In the case of patterns in the brain, such a conscious observer is not available, unless you imagine a little Cartesian ghost observing the activity in the brain and picking out the patterns.

There is another problem encountered at the most basic level of consciousness: awareness itself. Just as we cannot find any kind of basis in the uniform electrochemical gray (and actually “gray” is a bit flattering) of neural activity for the multicoloured world of sense experience, we cannot find any basis for the fact that we are aware of our sense experiences. We cannot, to use the jargon, find ‘‘the neural correlates of consciousness” (NCC): more precisely, identify an adequate basis for the difference between neural activity that is, and neural activity that isn’t, associated with consciousness. Anyone who believes in the identity of consciousness and brain activity has to deal with the fact that most brain activity is *not* associated with consciousness and the small amount that is associated does not look all that much different from the large amount that is not. There is not sufficient, or the right kind of, difference.

The NCCs have been sought most carefully in the visual system. The NCC-seekers agree that the primary visual area (VI), where the neural activity in the visual pathways first reaches the cerebral cortex, is not itself the seat of consciousness of sight, although it is necessary for there to be visual awareness. Visual consciousness, it is claimed, requires supplementary activity in the extra-striate visual cortex and the frontal and parietal cortex. The question then arises how all these disparate areas, in play at once, come together: how, that is to say, they *sum* their scattered activity to something that amounts to awareness, to a whole that is unified in itself. It is easy to see how an external observer could bring them together as a whole, just as I, looking at the brain, can see it as a whole as well as a collection of connected parts. But we don’t have an observer within the brain to *bind* the different parts into the kind of whole that seems to be required for consciousness: such an observer has to be constructed in the brain out of nerve impulses according to the neural theory and so the problem returns. (We shall come back to the binding problem in due course.)

At this point, it is important to keep asking questions that tend to get overlooked or discarded because they seem naive or even childish. One such question is this: if consciousness is identical with neural activity, which consists of travelling waves, is this activity to be considered as consisting in the travelling or the arrival? Only in certain areas of the brain, distant from where most nerve impulses originate, is neural activity associated with consciousness. This suggests that travelling is necessary, but only to ensure arrival. But what does arrival consist of? Well, as we know, when nerve impulses reach the end of a neuron, they may trigger activity in a connected neuron via synapses.

So "arrival” seems to correspond to more activity in certain central areas, presumably. But this, in turn, consists of travelling: nothing stands still; propagated impulses trigger other propagated impulses. If travelling remains essential and there is no real arrival in the sense of standing still, then the difference between what is happening in those places where consciousness is located and what is happening where consciousness is not located isn’t at all clear. Nor is it clear what localization actually consists of, given that nothing

keeps to a particular place.

Perhaps consciousness resides not in a place of putative arrival of impulses and in the moment of arrival but in the history of the journey they have undergone. Unfortunately, this would require the nervous system to step out of its present moment in order to reach into an (admittedly recent) past and an (admittedly short-term) future and integrate over time. This reaching out of the present tense, which means reaching out of the present (that is to say actual) state, is not possible for a material object; the physical world does not have tensed time, in which present, past and future exist side by side. It is, as we shall discuss below in “Brain science and human consciousness, III", unique to conscious creatures for whom time is explicit and whose lives have temporal depth.

We therefore have great difficulty with making sense of the notion of NCCs: that is to say, of neural activity, in a certain place, or a set of places, that is extraordinarily privileged, being (supposedly) the basis of consciousness in a brain that is overwhelmingly the site of unconscious processes. There is not enough difference between the kind of activity that is associated, and the kind of activity that is not associated, with consciousness plausibly to account for this absolutely fundamental difference. What’s more, it seems very odd that nerve impulses should have to *travel* in order to qualify to become consciousness or that a particular journey to a particular place would deliver the metaphysical transformation from electrochemical activity to subjective experience. For a start, the place they are coming from and the place they are going to does not seem different enough to carry the difference between events that are and events that are not associated with consciousness; or between events that are and events that are not consciousness itself. Given that nerve impulses never stand still, and have no clear point of arrival, the very notion of travelling to a location is problematic. And the idea that summed activity at several places is required for consciousness raises the question of how, or in what, it is summed. It does not exist in itself as its sum: to do so would require that it should somehow demarcate itself and then add up everything inside the boundary of demarcation.

Of course, no neuroscientist would suggest that location alone is sufficient to ensure that neural activity should be conscious. The other requirement is that the activity should be intense enough to break a notional threshold of awareness. The assumption that the more quantitatively impressive the activity, the more likely it is to do this - that more neural activity means more consciousness of something other than the neural activity - is not at all self-evident. The fact that it seems indisputable is due to transference of observations *within* the field of consciousness to the relation between consciousness and neural events. The fact that I am more likely to see a bright light than a dim one is translated into the assumption that I am more likely to have an experience when there is a lot of neural activity than when there is a small amount of neural activity; or that a lot of neural activity is more likely to amount to an experience than a smaller quantity. This is based on a false analogy illegitimately identifying the contrast between dim and bright lights with the contrast between less and more activity in the visual pathways. The difference between a bright and a dim light, what is more, is not the same as the difference between a light of which one is conscious and a light of which one is not conscious. Only the assumption that the difference is the same or analogous could make the assumption that more electrochemical activity means consciousness, or more intense consciousness, seem self-evident. Otherwise it would seem very odd that more nerve impulses would not only add up to the greater total but also, having done so, be more likely to.

This is why dedicated neuro-maniacs, most notably Dennett, have taken the desperate measure of denying the existence of qualia altogether, suggesting that they are spurious items left over from a “folk psychology” still haunted by Cartesian dualism. He argues this most thoroughly in *Consciousness Explained:* a book title that should have landed him in court, charged with breach of the Trade Descriptions Act, for what this, his most famous, book offers is not *Consciousness Explained, but Consciousness Evaded.*

Brain Science and Human Consciousness, II: Problems with Intentionality

Nothing I have said so far will cause neuromaniacs to change their minds. They will simply reiterate that this is how things are: the brain is mysterious but then so is matter. If you dismiss the neural theory of consciousness because it is baffling then, to be consistent, you ought to reject quantum mechanics, which demands that you set aside many more of your common-sense intuitions, even such fundamental ones as that things have a definite location.

In response, it is necessary only to point out that if you believe that the brain, or some small part of it, is the seat of consciousness then you are going to have to grant this bit of matter properties that no other material object - including most of the human nervous system, and perhaps all of the nervous system of some lower animals - possesses. You cannot be a materialist and ascribe to the brain the capability of making the material world present to itself. More specifically, you cannot deal with two features of consciousness that are connected, although I shall address them separately: *intentionality* (which I shall discuss in this section); and the *ability* to *make other items appear* (which I shall leave to the final section of this chapter because it is the most fundamental objection to the neural theory of consciousness).

So what is "intentionality"? This is a philosophical term that has a long history (a "sordid history", according to Searle), but its use in modern philosophy is traceable to Franz Brentano and a landmark book that he published in 1874, *Psychology from an Empirical Standpoint.* In this book he reminded his readers what it was that distinguished mental items from physical items. Mental items had the property of "aboutness": they were directed on, or about, other things. This was most obvious in the case of what Bertrand Russell later called "propositional attitudes": items such as hopes, desires, fears and, more broadly, beliefs, which are directed on objects or parts of the world or real or virtual entities or clusters of possibilities that are felt to be other than the subject. But they are also present in all knowledge and, indeed, all perception. It is perception that I want to focus on here because it illustrates most clearly what Brentano meant.

Consider a very simple example: my perception of the red hat next to my computer. The standard story is that I see the red hat because the hat interferes with the light in a certain way and some of the light bouncing off the hat enters my eyes. Changes in the retina result and these changes trigger impulses in the optic nerve and, eventually, in those parts of the visual cortex that have been identified by neuroscientists as the seat of visual awareness. This chain of events is very similar to causal sequences seen elsewhere in the material world. Physicists, physical chemists, biophysicists and so on would be entirely at home with the processes I have just described. But that, of course, is not the end of the story. I am *aware* of the red hat; and I am aware of it as being separate from me, at some distance from me, as having properties and a reality all of its own, some of which I cannot currently see. My awareness, that is to say, is of or *about* an entity that is located *causally upstream* from those events in virtue of which I am aware of the hat. The causal chain points in one direction, from the hat to my cerebral cortex, with the light being translated into electrochemical events as the key step; but the aboutness of my experience points in another direction, from my cerebral cortex back to the hat.

Actually, it's much more complicated than that. For, although I see a hat, I see it in virtue of events that involve it: the interaction between the hat and the light incident on it. That is primarily what I see, although I interpret it as the-hat-in-a-certain-light. For it is events, not objects, that count as causes of events. Nevertheless, it is an *object* that I see and (an added twist) I see it in a certain light, or that it is in a certain light, part of which is the light in virtue of which it is seen. The key point, however, is that intentionality - my awareness of the hat - points in the opposite direction to the arrow of causation. It points

from effects (nerve impulses in the higher levels of the visual pathways) *backwards* to their causes (the interference between the object and the light). And then it points further backwards to the partners producing the effects: the red hat and the light it is bathed in.

How the object and the light are unpacked, or inferred, from the events has been the subject of a huge research effort in the philosophy, psychology and physiology of perception. Some of this has focused on what is called "object constancy": that in virtue of which an object looks the same size and shape irrespective of the distance and angle we see it from. Object constancy is puzzling because the image cast on the retina will diminish as the object recedes. Other research has investigated depth perception: my ability to infer, from a two-dimensional image on the retina, that the object has three dimensions. There are other and even more intractable problems but our main concern here is with this fundamental property of perception: intentionality - namely, that perception is *about* something other than itself. The irony is that it is the neural accounts of consciousness that highlight just how mysterious this aboutness is. Giving perceptions a definite location in the brain (say neural activity in the visual cortex), makes the separation between the perception and that which it is about a literal, that is to say spatial, distance. The relevant neural activity is, say, a yard away from the hat that it reveals or is about. This is shown in Figure 1.

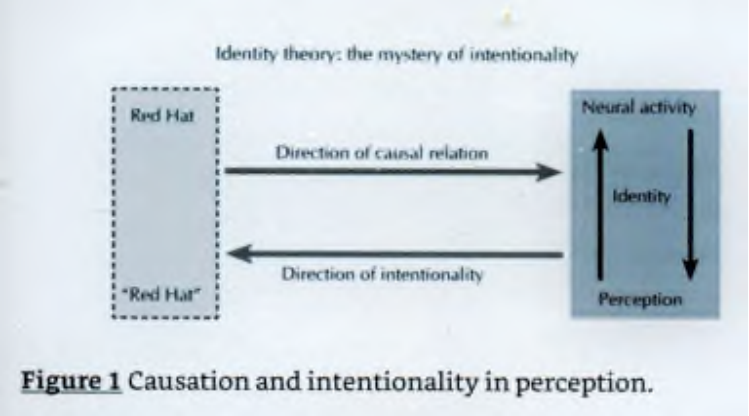

Figure 1 Causation and intentionality in perception.

There is nothing elsewhere in nature comparable to intentionality. It will prove, as we shall see, to be the key to our human differences: our subjectivity; our sustained self-consciousness; our sense of others as selves like us; first- and second-person being; our ability to form intentions; our freedom; and our collective creation of a human world offset from nature. For the present, I want to focus on the phenomenon itself. In Figure 1, the two arrows correspond, respectively, to the light getting into the brain (upper arrow) and the gaze looking out (the lower arrow). Physicalist neuroscience has no problem with the light getting into the brain through the eyes and triggering nerve impulses. The gaze looking out is another matter entirely. It is different from causation and *it is in the opposite direction.*

Nothing in physical science can even seem to provide an adequate explanation of why, or how, some (although not most) neural activity would reach causally upstream to events that led up to themselves; why or how a burst of impulses in my visual cortex should refer itself back to the interaction of the light with the hat and, out of this, construct a hat-in-the-light "out there". It is not merely a case of "registering" those events, as a photoelectric cell might register light from any source. For we not only register events, but also register them *as* belonging to something other than our self: we are *aware* of them and aware of them as "over there". It is a *revelation:* of an object to a subject in which object and subject are kept separate and distinct, with the subject (me) being here and the object (the hat I am looking at) being over there.

This difference between physical registration (if one can truly speak of something being "registered" by an entity, such as a photometer, that is not conscious of that which is registered) and perception is absolutely fundamental but quite elusive. It is easy to lose sight of it, particularly if one is a neuromaniac and has a vested interest in concealing it. It

is even easier to conceal it if one treats the brain both as material object and as a quasi-person. Normally one wouldbe inclined to say that the light impacts on the brain while it is the *person* who looks out and this would highlight the inadequacy of accounting for the gaze in neural terms. The habit of describing brains in terms that properly apply only to people (something we shall examine in detail in Chapter 5) makes it easy to think of the brain doing the looking and (more importantly) to imagine that the looking consists only of brain activity. Most importantly - and this is the pillar of unwisdom on which Neuromania rests - this makes it possible to conceal the outward arrow of intentionality or (more usually) to bury it in the inward arrow of causation.

This act of assimilation is most clearly expressed in the causal theory of perception. According to this seemingly common-sense theory, perceptions are caused by the things or events that we perceive; indeed, if they are not so caused, they are not true perceptions but hallucinations. We can now see how causation does not on its own deliver perception: that perceptions are more than effects of that in virtue of which perception is possible. *Something has to be done with the effects for them to reach upstream to their causes and become perceptions of the objects or states of affairs that are implicated in their causation.* This is overlooked, so much so that the causal theory has been extended to encompass more complex modes of awareness: propositional attitudes such as beliefs, expectations and so on; and verbal and non-verbal meanings and linguistic reference. My beliefs are, so the story goes, effects of the material world on my brain. The meaning of a word or a sentence is the effect it has on me. A word has reference in virtue of its creating an effect in my brain that stands proxy for the object that would have a similar effect in my brain. And so on.

You can see where this might lead: the brain (and hence the mind) becomes a mere causal way station, linking inputs into and outputs from the body. Perceptions, beliefs, meanings and reference are simply the intermediate neural steps between experiential inputs and behavioural (in the broadest sense) outputs.

The assimilation of consciousness to the causal net in which the organism is located has been the central pillar of materialist theories of mind, in particular of a highly popular theory called "functionalism". Functionalists argue that mind is not importantly about the phenomenal aspects of consciousness: actual awareness. No, its job - and consciousness, according to them simply is its job - *is to* refine the connection between inputs and outputs in such a way as to optimize the survival of the organism or the group to which the organism belongs. Any particular element of consciousness is constituted entirely by its functional role: its causal relations to sensory inputs, to other mental states, and to behavioural outputs.

This is as close to missing the point that one can get. At the most basic level, it ignores the lower arrow in Figure 1: that in virtue of which experiences, memories, beliefs and other propositional attitudes are *about* something other than themselves. And so it is easy to understand why those who wish to defend a materialistic account of consciousness have either dismissed or marginalized intentionality. The lengths to which they are prepared to go to achieve this are illustrated by the writings of Dennett.

Dennett argues that intentionality is not an intrinsic property of mental phenomena; rather, it is a product of "the intentional stance", an attitude that ascribes intentionality from without. Intentionality is not something in itself but a level of abstraction at which we view or describe the behaviour of a thing in terms of mental properties. This, Dennett says, gives us greater computational power when we are concerned to anticipate or understand their behaviour and hence is of adaptive value. Trying to make sense of my behaviour by seeing me as a collection of atoms - *the physical stance* - and predicting the future behaviour of that collection of atoms would place an impossible burden on my cognitive capacity. Even adopting a *design stance,* which would see me as an artefact or organism designed to achieve certain goals, would make working out what I might do next very difficult; I am, after all, more complex than an artefact such as a thermostat. *The*

intentional stance alone has sufficient power. This stance assumes that you are a self who acts according to beliefs, thoughts and intentions and on the basis of that I can make a pretty good guess at what you are, or are likely to be, up to. It does not, however, mean that you truly are such an item or that beliefs and other propositional attitudes are anything other than artefacts postulated by "folk psychology". The inner life we ascribe to others is merely an interpretative device and nothing in reality corresponds to it. And the assumption that we are related to the world by perceptions, beliefs, reasons is just such an interpretative device.

It is difficult to know why this argument has been taken seriously. While we might need to use a very sophisticated intentional stance to make sense of, and predict, the behaviour of a robot primed to behave just like me under all circumstances, there is an irreducible difference between myself and such a zombie. And, what is more, the intentionality ascribed to the zombie is real but mislocated; it lies within the team that designed it and had its functions in mind and within anyone, such as myself, who tries to anticipate the zombie's "behaviour". But it is not out of mere interpretative convenience that we ascribe all sorts of intentional phenomena - perceptions, feelings, thoughts - to people; it is because these intentional phenomena are real, as we know from our own case.

Overlooking the aboutness of perception and other conscious experiences means that we shall overlook many other things - which is very convenient for neuromaniacs but disastrous for anyone who is serious about capturing human consciousness. The intentional relation lies at the root of the distinction between the subject and the object, as a result of which human beings are not simply organisms but rather are *embodied subjects.* (We shall discuss this in Chapter 6.) While the material light gets into the brain by physical means, the gaze that looks out is not a continuation of that chain of physical events. It is a person that looks out, not a brain. The person is aware of herself as other than, as confronting, the object. While perception connects us with the material world it also asserts our distance from it and, more broadly, our difference. This uncoupling is most evident in vision among the modalities of perception but it is elaborated in the infinitely complex mediations of experience that are afforded by the signs - signals, gestures, codes, languages, words - that fill our lives.

For humans, perception is not simply a means by which, as organisms, we are wired into the world; it is also the basis of the distance that is opened up between ourselves as conscious agents and the world we can operate on as if from an outside: a virtual outside that is built up, as we shall see in Chapter 6, into a real, but non-physical outside that is the human world. Our perception yields objects that transcend our awareness; we are explicitly aware that the object is more than our perceptions - it is not exhausted by our perceptions - and that it is other than our self. This transcendent object, which is seen as something only partly revealed, is related to a transcendent self that is other than it. There is no room for this kind of thing in a causally hard-wired universe of material objects, which would include material organisms and material organs in those organisms, such as the brain. That is why Dennett, in common with many other mind-brain identity theorists are intentionality-deniers (or intentionality reducers); it enables them to see the mind entirely in terms of the function of a material brain evolved through material processes. Hence his claim that intentionality is just the product of an intentional stance that enables us to make a quicker assessment of the likely behaviour of a predator than, say, using an atomic or design-based approach. To ascribe intentionality to others is simply to deploy a conceptual tool to promote survival. Against Dennett, we would argue that intentionality is not simply something that is ascribed; it is a fundamental feature of human consciousness and it begins with perception. How, anyway, should we ascribe it to anything else unless it was something we had experienced in the first place in our selves?

By a nice irony, those who try to be hard line about consciousness and see it as simply an *effect* of the material world on a material brain end up in a position that is far from hard line. The claim that my experience of the red hat is a set of nerve impulses in parts of my

cerebral cortex raises awkward questions. The first is, given that those impulses really *are* about the hat, why does their aboutness stop at the hat? Once there is a reaching causally upstream, then there is no reason why it should not continue right back to the Big Bang. This may seem to be a silly suggestion but let us stick with it for a moment and examine the actual things that are thought to trigger the nerve impulses that are in turn supposed to reveal the object. It is not the object that causes the perception of itself but its interaction with the light that results in my seeing it. This is a bit messy: the interaction is a fizz of events, not just a few neat straight lines of light connecting the object with the eye. The object has to be constructed from the interference with the light: a challenging task, to put it mildly. Indeed, it is so challenging that many neuropsychologists argue that the object that we experience is not really an object that is out there at all: it is a construct put together by the brain. This leads to the idea that the world we inhabit is *a mental model* that has only a tangential relation to what is "out there", an idea that has dominated cognitive psychology for many decades. Frith has gone further and argued that the contents of the mind are not real.

If the objects we experience are actually constructed out of data that may mislead us, although they may be corrected by subsequent experience (otherwise we would not survive to be further deceived), then we have an interesting case of the pulled rug. The brain, which is supposed to be the passive recipient of energy from an outside world, now suddenly becomes something that actually constructs that outside world rather actively. Such activity seems to be at odds with the notion of the brain as a material object helplessly wired into the material world that surrounds it via causal interactions guided by the laws of physical nature. One would like to know where, out of the electrochemical activity of the cortex and other bits of the nervous system, the ability to construct an illusory or approximate world arises. The brain, it seems, has the power to fight back and shape the world by which it is shaped. This, of course, relies on counter-causally directed intentionality.

Those of us who are not brainwashed into thinking that they are brains washed by the laws of physics might be tempted to hazard a daring suggestion: that it is a *person,* or something like a person, that looks out at, peers into, interprets and shapes the world. And that person is prefigured in the counter-causal direction of intentionality: the very "bounce back" that some neural theorists of consciousness find so awkward they wish to deny it. Indeed, neuro-talk often dismisses reference to persons and their beliefs and conjectures and volitions as belonging to a pre-scientific "folk psychology" that it has itself grown out of. But we shall find, again and again, that we cannot make sense of what the brain is supposed to do - in particular postulating an intelligible world in which it is located - without appealing to talk about people who are not identical with their brains or with material processes in those brains.

The neurophysiology of the visual system falls short of explaining the mystery of the gaze for many reasons but, most fundamentally, because it cannot deal with intentionality. Intentionality highlights the mystery of what brains are, ultimately, supposed to do; namely, *to make other items, indeed worlds, appear to someone.* This presents an insuperable, ground-floor problem for neural accounts of consciousness, and we shall return to it in the final section of this chapter. In the meantime, let us look at other aspects of consciousness that elude neural explanation.

Brain Science and Human Consciousness, III: Problems with Pretty Well Everything that Matters

The problem with neural theories of consciousness becomes clearly evident when we consider full-blown perceptions; but it is already there, if less prominent, in the case of smaller fry, such as isolated sensations. Consider an itch or a tingle. The neural theory

would have to explain why, if the tingle or itch is actually in the brain, it seems to be located, it is felt, in the arm: where the cause of the neural activity arises rather than where the neural activity is located. There seems to be no way of explaining, if the experience and the nerve impulses are the same, how something can be in two places at once: in the brain and in the arm. The fact is that the brain is not aware of itself; even less are collections of nerve impulses in parts of the brain aware of themselves. They always refer any awareness elsewhere. When I cut myself I feel the pain in my finger, although the neural activity that is supposed to *be* the pain is in my brain. Identifying brain activity and experiences, far from explaining the latter, seems to make them more difficult to understand.

One way of trying to get round this is to argue that the brain "represents" what is in the arm, so that the itch is, at it were, the object, and the neural activity is the representation of it. Unfortunately, this way of recasting the relation between the itch in the brain and the itch in the arm is not acceptable for a very simple reason: representation presupposes prior presentation. For example, my face in a mirror counts as a representation of the visual appearance of my face only because my face, courtesy of consciousness, already has an appearance. In short, as with Searle's example of water and molecules of H2O, we require consciousness to be already in place in order to make the concept of sensation as "representation" or "re-presentation" work - or seem to.

So even lowly sensations cause problems for the neural theory of consciousness, but you ain't seen nothing yet. Other aspects of human consciousness are much further out of reach. Spoiled for choice, I am going to focus on features that are relevant to my larger aim of highlighting those ways in which humans are distant from the natural world. The features in question are all connected with first-person being: (a) the sense of (being an) "I" at a given time; (b) the unity of the self at a time and over time that also accommodates a sense of multiplicity; and (c) the sense of explicit time - of a timetabled future and an explicit past revealed in memory.

The sense of being an "I"

It is tempting to say that the material world is third-person, while human consciousness is first-person. This does grasp half the truth. But the world in which we live is also in some respects first-person: it is set out, in the first instance at least, in what Russell called "egocentric space", where near and far, here and there, are defined with respect to one's own location, as defined by one's body, and, in a more complex sense, by one's interests. There are no inherent centres or nears and fars in *physical* space. The material world is without viewpoints that arrange items along a gradient of proximity and distance. This viewpoint-less world is strictly no-person, rather than third-person. What *is* third-person is the objective, scientific view arrived at by suppressing individual viewpoints and favouring an imaginary viewpoint that gathers up all possible points of view. It remains a view, however, and is not inherent in matter that is noperson rather than third-person. The no-person view, a "view from nowhere" (to use the philosopher Thomas Nagel's poignant phrase) in which all appearances are summarized in the abstract, quantitative account of possible experiences, had by no one in particular and consequently by no one at all, is the ultimate goal, or at least the regulative idea of natural science. In order for this view from nowhere to be achieved, the third-person view must give way to a paradoxical viewless no-person view, which is the material world seeing itself but from no particular point of view. (It would not be a *world,* however, since that is a gathering together of many items in a centred whole infused with significance.)

It is easy to fall into the trap of thinking that a viewpoint, and the basis of first-person being, is bundled into the starter pack of any organism, and, since the organism is a material entity, to imagine that viewpoint could be present in the material world. After all, all organisms have inputs and outputs, an organic being and an environment around them: they seem objectively to lie at the centre of a world. However, in the case of organisms that

are not conscious or self-conscious, these contrasts are borrowed, imputed, honorary. Yes, there is a sense in which it is correct to invoke the opposition between “the organism” and “the environment” in the case of an insentient creature such as a bacterium. The organism does not, however, itself lie at the centre of its environment, creating an organism-centred space. The centre-surround distinction belongs only to the observer, just as that which counts as the surroundings of a pebble belong to the observer rather than to the pebble. It is the observer who posits the organism as being related to an environment centred on it. "Surroundedness” does not come free along with, say, a membrane marking the boundary between the organism and the rest of the material world any more than it comes free with an entity such as a pebble that has a continuous surface marking its limits. The boundaries visible to *us* do not transform the organism’s objective location into a point of view that stipulates that which is physically around it as *its* surroundings.

The brain, seen through the eyes of neuroscience, is a material organ within a material organism. It will be evident from what we have just said that there is nothing in the material transactions it has with the material world that would form the basis of the sense of a centred world, of "me”, or of the ownership that makes a brain *my* brain, *a* body *my* body, a portion of matter *my* world. There is nothing, in short, to underpin the sense of self: the feeling that I am and that certain things are addressed to me. Many neuromaniacs, as we have seen, would happily accept this and argue that the absence of a neural basis for the self is evidence that this is an illusion. I shall return to this in due course, but for the present I would argue that there are some fundamental elements of selfhood that cannot be denied without self-contradiction. It is not possible to deny viewpoint, the sense that one is (the feeling of “am”), and the feeling that one is in a setting that is centred on one’s self.

There are two other more prominent aspects of selfhood that cannot be denied: the feeling of being a *unity at a given time;* and the feeling of having some kind of *unity or coherence over time.* Let’s take a look at these.

The unity of consciousness: here and now

Consider, first, unity at a given time: the unity of my conscious moment. As I sit here I am aware of many things: my action of typing and all the movements, sights and sounds associated with this; the pressure of the seat that is supporting me, and other sensations arising from my body; several conversations in the background; thoughts coming into my head; memories; and so on. These are all distinct - otherwise I could not specify what they were - and yet they are also together. They belong to the present moment of my life. This capacity to keep things separate and *at the same time* experience them as together is evident at every level of our experience. To see the problem it is not necessary to look to complex experiences, such as entire visual fields, or scenes impregnated with meaning and memory, or a sense of rejection, or a hope for the future, which must involve many layers of integration without loss of the separate identity of the components. Unity, and the problems it poses for Neuromania, begin at a very basic level, as when we see a material object as the unitary bearer of many distinctive properties: as a subject with many predicates. Think of an experience as simple as my seeing my red hat.

According to standard visual physiology, this involves the stimulation of neurons that are responsible for detecting edges and synthesizing them into a perceived shape; for sensing colour; for determining location in space (the "where” of my hat); and for seeing the kind of thing something is (the “what” of my hat). I experience all this at once; my awareness of colour, shape, location and meaning of the hat are not presented separately. I see a recognizable red hat at a particular place, which, in addition, I might note, is likely to be damaged by someone who has just entered the room and wants to borrow it. There must be some place, according to the neurophysiological story, where the inputs into the various specialized groups of cells converge, the basis of that *sensus communis* which has

haunted the project of developing a neurophysiology of mind discussed in "You are your brain" in Chapter 1. For the organism to be successful in its "million- sided environment", it must in its reactions be many- sided. There will therefore have to be a mechanism that summates the signals from the senses, and from other sources, in a pathway to a common destination.

The usual putative mechanism, as we have seen, involves joining up neurons at synapses into networks and connecting those networks into other networks that ultimately summate the entire activity of the nervous system. At the microscopic level, this was described (in the wake of research using single-cell recording) as being carried out by "higher-order" cells. But this solution creates more problems than it solves.

If the inputs *do* converge, one would expect them to lose their individual identity, just as the mixing of colours results in a composite in which individual colours are lost. The higher-order cell would be a point where, instead of a red-hat-at-a-particular-place, one would have some unholy mixture of redness, hat-shapeness, location and meaning. It is as if the higher-order cell - or the region of convergence - has to deliver the hat simultaneously as its constituent features *and* as an integrated whole. This is, of course, impossible if one thinks of what happens at synapses: a kind of adding up and subtraction, so that what comes out of the higher-order cell is the sum of its inputs. It is as if in the equation 2 + 2 = 4 the right-hand side had somehow to hang on to, or be, the left-hand side; that the 4 had to keep the two 2's separate within itself while being 4.

The neurophysiological explanation of the unity of things that are also experienced separately is so evidently flawed that there must be some undeclared intuition that is making it seem right. The undeclared intuition is that the lower-order cells and the higher-order cells add up *themselves* to a different kind of whole that has two parts: the lower part where the component features of the hat are kept apart and the upper part where they are together. This is cheating, of course, because the higher-order cells are required to integrate the features of the hat as a whole and there cannot be implicit a prior integration of what the higher-order and the lower-order cells report, unless one imagines there is yet another viewpoint - higher still - from which both ways of experiencing the hat, as separate features and as a whole, can be seen.

The undeclared intuition, although invalid, gains apparent support from the anatomical fact that the lower cells and the higher cells coexist physically, side by side in the nervous system, so that what goes on in the latter does not obliterate what goes on in the former. This coexistence would not, of course, translate into the explicit co-presence of the activity in both, or the side-by-side presence that the summed activity is supposed to stand for: or not, at least, without a third viewpoint to gather up the lower and higher cells together. And, given that unification takes places at many, many levels - single object, single visual field, single sensory field, the unfolding of events in a sensory field, the unfolding of events in life - one would require an endless multiplication of higher viewpoints to retain the unity and separateness of the viewpoints sustained by cells lower down the hierarchy.

This rather tortuous argument can be summarized very simply; there is no model of *merging* of activity in the nervous system that would not lead to *mushing* of the merged components and a loss of their individual identity. The fact that the neural pathways supposedly dealing with the different aspects of an object, of a scene, of a life, are anatomically distinct does not solve the problem because it is the anatomical distinctness that creates the need for integration in the first place. Consequently, there is no neural explanation of how I see a visual field as an integrated whole and yet can still appreciate its component objects, and the relations between the component objects, and the constituent features of the component objects. And this insuperable problem is replicated at many levels all the way up to my feeling of being in a world that makes sense to me on the basis of past experience.

Our experience of being located in a sensory field that is at once unified - it hangs together as a field, so that the things in it are all related to one another - and at the same

time populated with a myriad of distinct items has another feature that resists explanation in terms of neural integration. Just consider for a moment your awareness of the visual field that is surrounding you now. The light arising from that field has two fundamentally distinct components: one is what we may call "the background lighting"; and the other is the array of illuminated objects we see in the light. We see the objects, so the story goes, because of the way they interfere with the light. Their presence, that is to say, is derived from an analysis of this interfered-with light. Now consider this. All that interference of objects with the light enters together through the narrow portal of the pupil; even so, you are able to fasten back on to the individual objects their own share of interfered-with light. In other words, the arrow of intentionality is very precise. *At the same time,* however, it can also be global, seeing the light in itself as a background illumination that is making the objects visible. The theories of integration that are on offer - which appeal simultaneously to anatomical separation (localization of function within the nervous system) and functional convergence - would have even greater difficulty explaining how *this* is possible. And we have already seen how the models of integration that are on offer would, if taken literally, generate objects, sensory fields, indeed lives, that would be an unholy puree of colours, feels, distances, meanings, memories and so on.

This, then, is the heart of the problem: consciousness at any given time is manifestly unified but also explicitly multiple. Models of integration, even if they did explain how it is that my experience of a million leaves amounts to an experience of a tree, or my experience of a red and round and distant object becomes the experience of "a rubber ball over there", could not *at the same time* explain how it is that I am still, nonetheless, aware of the tree as being composed of millions of leaves or of the ball as being red *and* round *and* distant. Our sensory/perceptual/cognitive fields are simultaneously unified *and* divided. This mystery - greater to me than that of the Trinity, of the three-in-one, that exercises theologians - is insufficiently appreciated, even by those aware of the so-called "hard problem" of consciousness. The appeals by Kantians to the notion of "synthesis' — and by neuroscientists to "integration" do not explain how we get merging without mushing. We have the same unanswered questions that dominated the debate in the nineteenth century - which we discussed in "You are your bra in" in Chapter 1 - between the unifiers and the localizers over "the parliament of little men".

Just how desperate things are is illustrated by the mechanisms that have been invoked to explain the physical basis for the unity of consciousness. One favourite ploy is to appeal to quantum physics. Sometimes this is mere hand-waving but some serious work has been done. Steven Hameroff and Roger Penrose have suggested that the unity of consciousness may be underpinned by a phenomenon called "quantum coherence", which they believe could be generated by the special properties of the folded membranes in axons. This doesn't persuade me for many reasons. The most obvious objection is this: the kind of structures that are supposed to house quantum coherence are widely distributed throughout the nervous system, and are not confined to those areas that are associated with consciousness. It might be argued (somewhat tendentiously) that quantum coherence does not make you conscious but unifies your consciousness if you have it already. We should, however, be suspicious of thinking of consciousness as a kind of stuff that is potentially dissipated but can be called to order by what, after all, are microscopic physical forces. Besides, there is no reason why the unification that quantum coherence supposedly

imposes should translate into subjective or experienced unity, even less into a unity in which multiplicity is retained. The brain itself, after all, is at one level a single, unified material item, and so should provide all the coherence that is needed - if the *physics* of the system were going to provide it - and, what is more, has the added advantage of being the right kind of size.

The appeal to quantum physics is deeply flawed for another reason; according to the Copenhagen interpretation, the ultimate constituents of the material world have definite properties (as waves or particles and possessing a definite location or velocity) only in the

presence of measurement - that is to say an observer. In other words, quantum phenomena *require* consciousness and so cannot generate it.

Those who look to classical, as opposed to quantum, physics for an explanation are even more obviously on a hiding to nothing. It has been suggested that electromagnetic phenomena may bind neural activity into a coherent whole. This falls foul of all the objections we made to using quantum theory. The truth is, no theory of matter will explain why material entities (e.g. human beings) are conscious and others are not. The phenomena described in physics are present equally in conscious and unconscious beings; indeed, they are universally distributed through the material world. So they provide no account of the difference between, say, a thought and a pebble, which is the kind of difference that any theory of consciousness worthy of the name must be able to capture.

Crick and Christof Koch thought they had solved the problem of the unity of consciousness by invoking the synchronous rhythmic activity of large number of neurons which act as a reference that binds all the activity together. One reason this is wrong touches on something that is central to consciousness: that it (unlike the material world) has tensed time, so we shall look at it in more detail later. However, it is also daft for another more obvious reason: it assumes that the rhythmic activity will bind itself in a unity; or that an objectively observed synchrony will automatically translate into subjective unity. It also fails to explain the property we have just been talking about: how that which binds the contents of consciousness together also keeps them apart. Anyway, a few years later they ditched this theory, which (for reasons that must have had more to do with Crick's justifiable reputation as a molecular biologist than with its merits) had attracted a huge amount of largely sympathetic attention. They looked instead to structures in the brain where things come together. As we have already noted, they focused on a little entity called "the claustrum" as the leader's office where "the parliament of little men" would be called to order.

In a paper that Crick was correcting on the last day of his life, they wrote of:

> The notion of the *dynamical* core - a shifting assembly of active neurons throughout the forebrain that is stabilized using massive re-entrant feedback connections. Its representational content, highly differentiated and yet integrated, corresponds to the unitary and yet amazingly particular content of phenomenal consciousness.[57]

This structure, whose representational contents are both "highly differentiated and yet integrated", apparently would answer to "the need to rapidly integrate and bind information in neurons that are situated across distinct cortical and thalamic regions". The claustrum, in virtue of its enormous reciprocal connectedness, is in "an ideal position to integrate the most diverse kinds of information that underlie conscious perception, cognition and action". The discussion in the past few pages should be sufficient to expose phrases such as "integration of information" as a smokescreen hiding the real nature of the problem. The idea that unification could occur at some point of physical convergence in the nervous system is empty because, to repeat, it gives no model of merging without mushing.

The unity of consciousness: being one over time

So much for the unities of our consciousness at a particular time. However, we are also unified *over time.* In order to pre-empt the objection, of which we have heard much, that we are *not* unified over time, or at least that our sense of being enduring selves is an illusion that neuroscience should disabuse us of, let us just think of any everyday activity and see how it is dependent on our being intricately internally connected from one day to the next, or indeed one week, month or year to the next. Consider an ordinary commitment: say a plan to meet for an important dinner in a couple of weeks' time. The

commitment knits together a multidimensional lace of moments. These include: those in which we discussed the dinner, the where, when and why; the time we spent clearing a space for it, making sure that we got there punctually; and those moments in which we deployed all sorts of implicit knowledge in order to find our way via car and foot to the right restaurant at the right time, while in the grip of a thousand other preoccupations, and floating in a sea of sense data. This is just for starters. There are also those moments in which we remind ourselves of the dinner, in which we check our other commitments, in which we think about its purpose or purposes or lack of explicit purpose, in which we consider what we are going to say, itself rooted in a complex sense of who we are, and so on. The fact that this ordinary arrangement comes off at all is a striking manifestation of the inexpressibly complex inner organization of our lives and its extendedness across time. And it also shows how the favoured solution to the problem of the complexity of our lives - the appeal to localization in the brain - would just make things worse. For keeping things tidily apart would obstruct the process of bringing them together in a way that is infinitely more complicated than is required to bring together the aspects of an object such as my red hat.

The troubles that the dinner date presents to the neural theory of consciousness go deeper than this. If you think of all the things that would have to be going on in my brain in order to ensure that I turned up at the right place at the right time, you could be forgiven for entertaining the image - based on conventional neuroscience - of a vast number of overlapping electronic microcircuits supporting a huge ensemble of different functions, and it is difficult to see how they could be kept apart so as not to interfere with one another. You will recall the suggestion by Friston that "the brain acts more as if the arrival of ... inputs provokes a widespread disturbance in some already existing state", rather as happens when a pebble is dropped in a pond. Well, let's build on that notion and think of everyday consciousness as a million set of ripples in a pond created by the impact of a dense shower of hail, compounded by all sorts of internal sources of ripples. How are we to explain how each ripple or set of ripples - such as those supposedly corresponding to my complex plan to have dinner with you - could retain its separate identity? It hardly seems possible. It seems even less possible if we remember that, ultimately, the nervous system has to allow everything to merge in the moment of present consciousness, steeped in meaning, but retaining its relation to a highly structured near and distant past and reaching into an equally structured future of expectation, responsibility, time table, ambition and life plan. This moment (unlike the present moment of a computer, even a Cray supercomputer with 10^{12} operations per second) has to bring everything together, so that I know where (in the widest sense) and who (in the deepest sense) I am. So again we have the nineteenth-century problem highlighted by Flourens, who lost the battle against modularity and Lange's "parliament of little men" in the late nineteenth century.

What makes the problem insoluble in neural terms is that if there were a neural mechanism for bringing everything relevant together, it would simply exacerbate the problem of keeping everything apart. For, while the events in the brain are required to be bound into some kind of unity, something must *at the very same time* keep distinct vast numbers of projects, actions, micro-projects, microactions. Moreover, to make things even more difficult, those distinct projects must connect with a thousand others as each provides the others' frameworks of possibility. My keeping this important engagement explains my refusing other invitations; rearranging the day so that I arrive on time; being more than usually concerned to keep my distance from someone who had a cold a couple of days ago because I know that you can't afford to catch a cold as you have a crucial lecture to give the following week. The distinctiveness of the patterns of ripples has to be retained, although the patterns have to be open to one another. And worse, moment-to-moment consciousness has to retain a *global openness* in order that I can fulfil the multitude of activities adding up to attending the dinner date in a sea of unplanned events, so that, for example, I avoided the cyclist who might have killed me as I crossed the road to

the venue, or took account of the fourth step outside the restaurant on my way to accomplishing this timetabled complex task.

The unifying organization necessary to complete even the simplest task - such as keeping an appointment - reaches down to the smallest details. I look at my watch and am shocked to see what time it is. If I do not hurry I will be late for the dinner. I therefore lock the door rather hastily and speed up the conversation I have with a neighbour who has hailed me from across the road, thinking of courteous ways of escape. The pressure of time, which requires the modification of both these actions and many others, means that I have to re-set the motor programmes of which they are composed, without the harmony between the subroutines being disturbed.

The amazing feat of unification that is exemplified in every voluntary activity is all the more amazing for being accomplished while all the components of the unified action retain their distinctness and are accessible to observation and individual modification: and even more because they are deeply interrelated in the hours, days, weeks and years of a connected self. And this connectedness is a personal, long-term, inner connectedness that cannot be downloaded to the impersonal connectedness of synapses.

When we see the unity of consciousness for what it is, we should be able to resist the temptation to reach for easy analogies, as for example with a computer: a model we shall dissect in "The computational theory of mind" in Chapter 5. Yes, a computer has numerous modules in which various inputs are kept separate and, yes, it has a central processor where they all in some sense come together so that an output can be fashioned that has been influenced by them all. There is, however, no place or time at which that which is separate is *also* unified; whereas, by contrast, every moment of our consciousness has precisely this characteristic of being unified and multiple.

Now you will note that I haven't talked about my sense of continuing personal identity, which some neuroscientists dismiss as an illusion. No, I have made my case for such a unity over time on the basis of aspects of behaviour that even neuromaniacs cannot deny.

Memory in a dish?

The kind of integration over time that I have just talked about often, although not always, makes time explicit. There is one mode of integration over time where the dimension itself - its passage, its length, the distance of things from the present - most definitely is clearly made explicit. I am referring to memory. Contrary to the claims of many neuroscientists, full-blown memories, such as you and I experience all the time, and more broadly the explicit temporal depth of our lives, cannot be captured by a neural account of the mind. To appreciate this and, more generally, to grasp how memory cannot be found in matter, however configured it is, we need to remind ourselves what memory is and, after this, what matter is.

When I remember something I have experienced, the memory is not merely a recurrence of the experience. Nor is it, as the philosopher David Hume suggested, a "pale" or less vivid copy of the experience. No, when I recall something that is past I am aware *that* it is past; remembered red is not just like a present experience of faded red. I have a sense of a place in time, outside the present, in which what was experienced, what the memory *is* about, took place. Supposing I remember that yesterday you asked me to do something. Although my memory is necessarily a present event it is aware that it is about something that is not present. The memory is not only the presence of something that is absent but also the presence of something that is *explicitly* absent. When I remember your request, however clear my memory, however precise the mental image I might have of you making the request, I am not deceived into thinking that you are now making the request. Your request is firmly located in the past. As for the past, it is an extraordinarily elaborated and structured realm. It is layered; it is both personal (memory) and collective (history); it is

randomly visited and timetabled; it is accessed through facts, through vague impressions, through images steeped in nostalgia. *This realm has no place in the physical world.*
The physical world is what it is. It is not haunted by what it has been (or, indeed, by what it might become): by what was and will be. There are, in short, no tenses in the material world. This is beautifully expressed by Albert Einstein in a letter, written in the last year of his life, to the widow of his oldest friend Michael Besso: “People like me”, he said "who believe in physics know that the distinction between past, present and future, is only a stubbornly persistent illusion”. Tenses are not, of course, illusions, unless the only reality that is accepted is the world as revealed to physics. But they have no place in the *physical* world. And they therefore have no place in a *piece* of the physical world: a material object such as the brain. The only presence that the past has in the material present is in virtue of the contents of the present being the effects of the past. As we shall see, being an effect of past events does not of itself amount to being the presence of the past.
Just how completely memories elude translation into neural activity is illustrated by comparing them with perceptions, which, as we have seen, are not, in virtue of their intentionality, explicable in terms of the causal relations seen in the material world. Memories, too, have intentionality or aboutness, but they have a double dose of this. They reach through time to the experience on which they are based. That is the first dose of intentionality. But those experiences were in turn about the events that they were experiences of. This is the second dose of intentionality. This double dose reflects how memories are both in the present (they are presently experienced) and in the past (they are of something that was once experienced). They are the presence of the past.
Needless to say, neuromaniacs imagine they can deal with this. Indeed, there have been recent claims that the neural mechanisms of memory are close to being cracked. One researcher - Eric Kandel - received the Nobel Prize in Physiology or Medicine in 2000 for research that led him to claim that he could capture "memory in a dish”. It is worth looking at his studies in some detail because they demonstrate very clearly how it is possible to deceive oneself into thinking that memory can be explained in neural terms.
Kandel’s studies were carried out using the giant (almost 30 cm long) sea snail *Aplysia. Aplysia* has two features that make it attractive to neuroscientists. First, it has relatively few neurons (a mere 20,000 compared with the hundred billion in your cerebral cortex alone and its hundred trillion synaptic connections); second, the neurons are strapping cables of a millimetre or more in diameter, and uniquely identifiable, so it is easy to see what is happening inside them and, more importantly, what is happening inside their connections, the synapses. The snail has the additional advantages of being (a) ugly and (b) dim and so it is unlikely (a) to attract the protection of the Animal Liberation Front or (b) to seek legal advice. This is relevant because of the unkind nature of the experiment that Kandel used for his investigations, which involved a defensive withdrawal reflex. When the animal received an electric shock to its tail, it demonstrated a gill withdrawal reflex. They had been weakened by habituation to repeated stimuli. After the shock, it would withdraw even after an innocuous stimulus. This was a form of learned behaviour, which lasted longer the more shocks it received. Kandel saw this as a model for memory.
Because of those giant neurons, he was able to identify the changes that occurred in the electrical and biochemical properties of their synapses as the snail learned to be jittery. It was this that he described as “memory in a dish”. His own, and subsequent research by others, on a variety of species, such as young chicks, showed that when an organism is trained on a novel task there are increases in the size and strength of certain synaptic connections in particular regions of the brain. The synapses enlarge and the effectiveness of the neurotransmitters within them is increased. But why should we think this has anything to do with memory as we humans know and value it? Kandel thinks it has because, he argues, there are no fundamental functional or biochemical differences between the nerve cells and synapses of humans and those of a snail, a worm or a fly. From this he concludes that similar changes not only underpin your memories and mine but are

what memories amount to. Human memory, like that of *Aplysia,* is stored in, is identical with, the modifications of the connections between nerve cells. Experience leaves a biochemical imprint on the neurons and this alters their excitability. This altered excitability is the *trace* of experienced events: the presence of the past.

Memory is, of course, a little more complicated with you and me than with *Aplysia* but, many neuroscientists would argue, the principles are the same. My memory of the smile on your face when we last met at London Waterloo railwy station is more sophisticated than the learned flinch of the unlucky sea snail. Even so, my memory is stored in the form of the altered connectivity of the neurons associated with the smile. Those neurons are primed to fire off in response to present cues, prompting me to recall the smile. The experiments on *Aplysia* and other animals supposedly show how this rewiring takes place and hence how memory works. Memory, so we are told, is “encoded” in changes in the minute structure, and consequently the responsiveness, of neurons. Irrespective of whether it is a matter of learning to behave in a certain way or acquiring factual knowledge, there is the same underlying mechanism: facilitation of the transmission of nerve impulses across synapses due to long-term enhancement of their reactivity.

You might be resistant, as I am, to this idea. *Aplysia,* for all its altruistic commitment to advancing the science of memory, does not, as far as I know, have any of the following: memory of *facts,* such as that there is a London Waterloo station (this is what psychologists call semantic memory); explicit memories of events, such as the meeting at the station, that it locates in the past (so-called "episodic memory”); or autobiographical memories it ascribes to its own past (corresponding to my sense that it was I who saw your smile - an "I” that lay at the centre of the circumstances, of the self-world, in which the experience was had). Nor does it have an explicit sense of time, of the past, even less of a collective past where a history shared with one other person - two, ten, a thousand, a million, a billion other people - is located. Nor can one imagine it *actively trying* to remember past events, racking its meagre allocation of 20,000 neurons to recall the shocks that now make it twitchy, any more than one can think of it feeling nostalgic for the time when it had confidence in a benign world free of electric shocks. In short, the altered behaviour of *Aplysia* has little, perhaps nothing, in common with memory as I understand it.

Neuromaniacs will not be impressed by my objection. The difference between the shock-chastened sea slug and my feeling sad over a meeting that passed over so quickly is simply the difference between 20,000 and 100 billion neurons or, more importantly, between the modest number of connections within the sea snail’s nervous system and the unimaginably large number of connections (said to be of the order of a 100 trillion) in your brain. Should we accept that the difference between Kandel’s “memory in a dish” and actual memory is just a matter of the size of the relevant nervous system or the number and/or complexity of the connections in it? I don’t think so. What we noted earlier about tensed time should be enough to show that numbers of neurons and the mind-boggling complexity of their connections will not deliver the difference between Kandel’s “memory in a dish” and the kind of thing we think of when we talk about our memories.

Let us return to that smile. It is supposed to be "stored” or "encoded” in the form of a changed state of excitability in part of my neural circuitry resulting from my being exposed to the smile. A present experience reminding me of the smile is one that stimulates the part of my nervous system whose activity corresponds to the experience of the smile. The present event are cues or triggers. The memory, that is to say, is a present state of a part of my nervous system: a physical state of a physical entity, namely my brain. Somehow, this has to be about, or refer to, the smile by referring to an experience that was itself about or "of” the smile.

This is the "double intentionality” that we noted above. One arrow of this double intentionality explicitly refers backwards in time to something that is no longer present: indeed, no longer exists. A remembered smile is located *in the past:* indeed in a past world,

which is, as John McCrone has put it, “a living network of understanding rather than a dormant warehouse of facts”. Thus we see intentionality elaborated: it opens us up to a present world that exceeds our experience; and it opens up the present world to the absent, the actual to the possible. As a result, as we shall discuss in Chapter 6, we have our being in a world that is an infinitely extended space of possibilities; we are not simply "wired in” to what is. And this, as we shall see in Chapter 7, is the basis of our freedom. And the failure to see this is the reason why Kandel’s claim of seeing memory in a dish is not only wrong but importantly so.

Scientifically, Kandel’s work has been hugely influential, as recent work bears witness to. For example, the observation of the emergence of new proteins in the synapses of *Aplysia* in response to stimuli has been described as "watching memories being made”. A paper by Hagar Gel-bard-Sagiv and colleagues, published in *Science* in 2008, claiming to solve the problem of memory, inadvertently underlines why this claim is without foundation. The authors found that the *same* neurons were activated, and in the same way, when individuals remembered a scene (from *The Simpsons')* as when they actually saw it. But seeing and remembering seeing are (as you don’t need me to point out) different. The neuroscience that can’t capture this absolutely fundamental difference, in which lies the very essence of memory, cannot claim to have an account of memory. And this difference eludes it because it is unable to separate that which is activated now from that which happened *then,* as both are present as consequences of past events. And this is why it is not possible to get even a conceptually clear account of the difference between the memory and the act of remembering: that which is presently stored as the memory and the processes by which memories are actively remembered or spontaneously recalled.

We have already seen that making present something that is past *as* something past, that is to say *absent,* hardly looks like a job that a piece of matter, even a complex electrochemical process in a piece of matter such as a brain, could perform. There are, to repeat, no tenses in the physical world; no realms of “what was" (or “what will be”) outside “what is”. Material objects are what they are, not what they have been, any more than they are what they will be. A changed synaptic connection is its present state; this changed state does not hold on to the causes of its present state. Nor is it “about” those causes or its increased propensity to fire off in response to cues. Even less is it about those causes *explicitly located at a temporal distance* from its present state. For a real memory not only reaches back to its cause, but also maintains the temporal distance between itself, the effect and its cause. If it didn’t, it would be confused with a perception. Reference to the experience-based behavioural changes that are not associated with any sense of the past, such as those seen in *Aplysia,* as “implicit memory” is simply a fudge.

So how did anyone ever come to believe that memory could be a “cerebral deposit” (to use Henri Bergson’s sardonic phrase in his classic *Matter and Memory)?*— In a sense, Kandel’s account of memory is the latest version of Socrates’ suggestion (as reported in Plato’s *Theaetetus)* that memories are analogous to the marks left in a wax tablet by the impress of events. This is the intuition that leads us to imagine that an altered state of something is, or even could be, *about* that which caused its altered state. How do we allow that obviously dodgy idea to pass? I think it is because we smuggle consciousness into our thoughts about the relation between the altered synapse and that which caused it to be altered, so that we imagine that the one can be "about” the other: that the altered synapse or the alteration in the synapse can be about that which caused the alteration. It is easier to see what is wrong with this if we look at a more homely example of alteration: a broken cup.

A broken cup can signify to me the unfortunate event that resulted in its unhappy state. But this requires my consciousness. If you allow that the present state of the cup can signify its past state, or the events that took it from its past to its present state, without importing consciousness, then you should be prepared to accept that the present state of *anything* can be a sign of all the past events that brought about its present state and that

the sum total of the past can be present at every moment. From this it would follow that all matter could claim to be blessed with memory in virtue of having being changed; and the present state of the universe would be a delirium of all its previous states, present side by side. Fortunately, such a claim is without foundation. Yes, a pebble is in a sense a record of its past, just as a battered suitcase is a record of all the vicissitudes it has undergone and, indirectly, of the journeys in which it has accompanied me. But the pasts are not housed in the pebble or the suitcase. It is I who make the present state of the pebble or the suitcase a sign of its past states and of elapsed time. The footprint is not the memory of a foot, except to an observer.

This point was made indirectly by William James when he remarked that “a succession of feelings is not a feeling of succession. And since to our succession of feelings, a feeling of their own succession is added, that must be treated as an additional fact requiring its own special elucidation”. This remark applies with even greater force to the succession of the states of a synapse - or a pebble. None of those states carries the sense of succession, or of the one being past and the other present: not unless, of course, we smuggle in consciousness by thinking of an observer who sees both states of the synapse or the pebble.

Smuggling in consciousness like this is, of course, inadmissible because the synapses are supposed to supply the very consciousness that reaches back in time to the causes of their present states. But, as we have seen, they don’t. So they cannot be memories or the basis of them. This is connected with the fact that in the physical world no event is intrinsically past, present or future. It becomes so only with reference to a conscious, indeed self-conscious, being who provides the reference point, the “now”, that makes some events past, others future and yet others present. The temporal depth created by memories, which hold open the distance between that which is here and now and that which is no longer, is not to be found in the material world.

We must assume that neurophysiologists and others who think of memory as a material state of a material object - as "a cerebral deposit” - also believe what physicists have to say about matter. In this case, they ought not to believe that tensed time could be manufactured in a material object such as the brain or, more specifically, in a particular state of synapses, irrespective of whether they are located in the spinal cord (which has little to do with memory) or the hippocampus, which is supposed to be a key memory structure. Only homeopaths believe that material substances remember their past states. A synapse no more contains its previous state than does a broken cup. Nor does it retain, as something explicitly present, its previous state, the event(s) that caused it to be changed, the fact that it has changed or the time elapsed between its present and one or more of its past states, so that the latter would be present in all its “pastness”. All this would be necessary, however, if synaptic alteration were truly to be the stuff of memory.

Another reason we might be persuaded into thinking that the present state of a piece of matter such as a synapse could be a memory of the past events that have impinged on it is linguistic. The word “memory" is used very loosely and covers a multitude of phenomena, ranging from an acquired habit (which may not even be conscious) to an explicit recall of a unique event. Neurophysiologists of memory trade on this profound ambiguity. They slither from memory as you and I understand it (as when I recall your smile last week at London Waterloo) to learning (as when I get to acquire expertise or knowledge); from learning to altered behaviour (as when a sea slug acquires a conditioned reflex); from altered behaviour to altered properties of the organism (as happens in the synapses of a sea slug conditioned to withdraw into its shell when water is disturbed); and (Bingo!, there we have it) the materialization of memory. But with Einstein’s help we can see that sincere materialists - those who believe in neural accounts of consciousness - must acknowledge that they have no explanation of memory. Instead of thinking that it can be located in the brain, even less captured "in a dish”, they ought to hold, along with Bergson, that “memory [cannot] settle within matter” even though (alas) “materiality begets oblivion”. (This is an illustration of the difference between necessary and sufficient conditions.) In short, they

should take off their dull materialist blinkers and acknowledge the mystery of memory: the presence of the past, and the temporal depth this implies, which does not exist in the material world.

The inadequacy of the neurophysiological account of memory should be obvious from the fact that it can be applied equally well to a *Aplysia,* whose behaviour is changed by an electric shock, as to a human being reminiscing about past days. The lowest common denominator between us and sea slugs is low indeed. And yet the changes in the properties of synapses have been invoked not only as the basis of memory but also (where the self is reduced to neuronal states rather than denied to exist outright) as the basis of the self: that feeling that I am, that I am such-and- such, and that I am the same such-and-such over time, so that I am responsible for actions that I carried out many years ago. Indeed, renowned neuroscientist Joseph LeDoux has even published a book that argues that the synaptic connections between our neurons, modified by our past experience, are what make us who we are.

Tensed time, change, endurance and the nervous system

Our discussion of memory has led us to think about the nature of time: more particularly about physics of time. It is important to appreciate that, in the absence of an observer, time has no tenses; not only does the physical world not have past and future in which events are located but (and this may seem less obvious) it doesn't have the present. For an event to count as being present, there has to be someone for whom it is present, for whom it is "now" as opposed to "then" or "not yet". The mere fact that something is does not generate a present tense: matter does not turn back on itself and become "That it is (now)". The complex consciousness of self-aware human beings brings tenses into the world and makes the happenings of the material world the contents of the present tense. Only by overlooking this human basis of tensed time can memory as we experience it be assimilated to learning, learning assimilated to behavioural changes and behavioural changes reduced to altered properties of a piece of matter such as a brain. We could put this another way by saying that matter cannot entertain *possibility:* that which may exist or turns out not to exist; the contents of the remembered past or anticipated future. We may take an even more radical stance. The great philosopher of time Adolf Griinbaum does. He has argued that there is no such thing as *becoming* in the physical world: in the absence of an observer. Unfortunately he concluded from this that "becoming" was unreal. This is clearly the kind of absurd position you get into when you assert that the sum total of reality is *physical* reality. But the challenge he presents to those who would reduce consciousness to physical events is valid: "How can temporal becoming be intrinsic to mental events but not to physical events (such as events in the brain) with which these mental events are correlated and upon which they are, for a naturalist, causally dependent?" There is one way to deal with this: to conclude that mental events are not physical events in the brain.

At any rate, it is arguable that both explicit change and endurance (persistence) require more than matter as conceived by the physicist. Change requires someone who will connect state A of an item with state B of the same item by making them both present at the same time. Endurance requires linking an earlier phase of state A with a later phase of state A. In neither case can the things that are related in the perception of change or of endurance physically exist at the same time; even less can they exist at the same time as separate *and* be connected.

The lack of tensed time in the material world is relevant not only to memory but also, less obviously, to the issue we dealt with above: the unity of consciousness at a particular time. We found that it was difficult to account for even a small component of this unity; namely, my experience of the different features of an object as the features of a single

object. How do I get the redness, the shape, the location and the meaning of my red hat together and yet still keep them apart, so they can be noted separately? The appeal to different locations of the brain, such that the component characteristics of the hat were experienced in one place *and* its unity as an object in another, was found to be unhelpful. It required the same collection of nerve impulses to work *twice:* as part of a smaller crowd - say neural events taking place in a particular area - and also as part of a larger crowd, of neural events taking place in several locations that are linked. We are now in a better position to see how the explanation offered to deal with this “binding” problem - as it applies to the visual field as a whole, the sensory field as a whole, and indeed the unity of the self - by Crick and Koch doesn’t work.

It will be recalled that they proposed a background rhythm of about thirty - five cycles per second - “a common neural oscillation” - which engages large quantities of the brain, and that it is this *synchronous* activity that "binds” together all that is happening in the present moment into a unity. Although Crick and Koch no longer subscribe to this, it remains relevant to us, however, for overlooking the oversight that gives it what little plausibility it has. Namely, it assumes that synchrony is something inherent in physical events. But the material world does not bind separate events together and label them as parts of “now”. What’s more, according to physics - more specifically the special theory of relativity - synchrony depends on the perception of events. Two events are synchronous if and only if they are *observed* to be synchronous. This requires picking them out and seeing the temporal relation between them, something that is not possible without an observer. So the unifying effect of the synchronous activity Crick and Koch talk about depends on an observer to synthesize them into a unity. But if an “I” or something like it is needed to confer unity on the very elements that are supposed to bind the moment and the “I” together, we may as well cut out those elements.

So, the present tense - which gathers together all those things that are “now” - does not exist in an observer-free material world, and hence must be absent from the brain understood as a material object. Nor does the past or, indeed, the future. The future, after all, does not yet exist. It is the notional location of possibilities, which we humans have mapped in a multitude of complex ways into boxes populated with events anticipated or planned. Matter can house only actualities. While there are indeed sequences of events in the material world, the relation in virtue of which one event is "past” compared to another, or “future” compared to another, has to be established by an observer.

It will be obvious that if neuroscience cannot capture the unity of the present moment, nor the sense of the past or future, it will not be able to deal with the unity of an enduring self. The closest science can come to an enduring self is a succession of events that are bound together only in virtue of the objective facts (not available to it as facts) that they are housed in the same brain and have cumulative effects on the structure or functioning of that brain. The neurophysiological self is at best the locus of “one damn thing after another”, which hardly comes near to the self of a human being who *leads* her life, who is a person reaching into a structured future with anticipations, aims and ambitions, that are themselves rooted in an almost infinitely complex accessible past that makes sense of them.

While tensed time does not correspond to anything in the physical world, it is indubitably real and a ubiquitous presence in human life. This is why biologists, who are determined (as we shall discuss in the next chapter) to find human characteristics in non-human animals, try to find a sense of tensed time in beasts. We need, therefore, to be on our guard when animal models are used to explain the basis of human memory. We should likewise treat talk about animals having a sense of the *future* with similar scepticism. One example is particularly relevant here: a study by Nicola Clayton and her colleagues published in the prestigious journal *Nature.*

Clayton and colleagues claim that western scrub-jays, a member of the crow family, have an explicit sense of the future as evidenced by their apparent ability to plan for it. Like

many other animals, these jays store and recover food caches. There is nothing unusual about that. But what has excited attention are Clayton's studies of "caching" of food under different experimental conditions. Jays (apparently) make provision for *a future* need, both by preferentially hiding food in places in which they have learned they will be hungry the following morning and by differentially storing a particular type of food in a place in which that type of food will not be available the following morning. Clayton and colleagues believe this pattern of caching cannot be attributed to conditioning or cue-driven behaviour.

In fact, there is no need to ascribe a sense of the future to an animal whose present behaviour optimizes the servicing of future needs. Even if the jays did have a sense of the future, would this be evidence that this sense of the future, or even of explicit time, was similar to that which rules our lives? Not at all. Our human sense of the future is that of a densely populated open space of possibility that is structured according to anticipated seasons or (in recent history) numbered days. This is hardly mirrored in the behaviour of crows choosing between caching more of one type of food rather than another, even if this does seem to indicate a sense of future need. We could imagine the jays' behaviour, which would ensure the optimal allocation of food between serving present hunger and meeting future needs, being "hard-wired". We cannot imagine this of an explicit, fully developed sense of future such as we humans have. Behaviour addressed to singular future possibilities that we anticipate is not something that seems to correspond to fixed neural wiring, for, being material substances, neural wires do not deal in explicitly entertained possibilities. At best, they can be tuned to objective probabilities.

If scrub-jays truly had a fully developed sense of the future, this sense would be addressed to a *field* of possibility and it would not refer only to food caches. While our human sense of the future is not entirely gathered up in timetables, or some kind of formatting of the not-yet, timetabling is the only sure evidence of a extended, spacious sense of future. What is more, it would be astonishing if the scrub- jay's future were personalized: in other words, permeated by a sense that it is its own future, a future that it can influence, for which it will be in part responsible. If, as seems likely, it were no such thing, it would not count as a true future in the sense that we humans have a future to which we orientate ourselves both individually and collectively.

Why There Can Never be a Brain Science of Consciousness: The Disappearance of Appearance

> It is [Bishop] Berkeley's merit to have realised that the Cartesian/Newtonian philosophers, seeking to account for a *seeable* world, succeeded only in substituting a world that could in no sense be *seen.* He realised that they had substituted a theory of optics for a theory of visual perception.

There are many other aspects of consciousness that elude any kind of conceptually coherent explanation. For example, it is not clear how, within the population of nerve impulses, we could find the basis for the absolutely fundamental difference between the *level* of consciousness (alert, drowsy, comatose) and its *content:* between background lighting and that which is lit. And what about the active directing of attention or racking one's brains to remember something? But I won't pursue these problems because I think I have already given enough reasons for maintaining that not only are current neural explanations of consciousness inadequate but also neurally based stories are wrong in principle and their inadequacy won't be amended by technological advances enabling ever more complete accounts of what is going on in the brain. For even quite profound inadequacies are themselves only symptoms of a yet more fundamental problem: a contradiction at the heart of neural theories of consciousness that I want to discuss now.

This contradiction rules out *the very idea* that certain material events in the brain could make a world around the person appear to that person. The materialist account of mind requires us to confer on brain events properties that actually run contrary to the physicist's notion of the matter of which they are formed. I want to dwell on this because it addresses the following objection to my critique of the neural theory of consciousness: that neural theory does not aspire to be an explanation; it simply reflects empirical truth, and the fact that it is mysterious does not make it untrue. It is this seemingly perfectly reasonable response that requires us to dig a bit deeper and to ask some more fundamental questions about the kind of activity that is supposed to be identical with consciousness.
Let us go back to what Dennett (accurately, I believe) calls "the contemporary orthodoxy" in a passage quoted in "You are your brain" in Chapter 1:

> There is only one sort of stuff, namely *matter* - the physical stuff of physics, chemistry, and physiology - and the mind is somehow nothing but a physical phenomenon. In short, the mind is the brain ... we can (in principle!) account for every mental phenomenon using the same physical principles, laws, and raw materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition, and growth.

It is when we examine this, the clearest possible statement of the metaphysical framework of Neuromania, that we shall see why it is a castle built on sand. Neuromania has to look for consciousness in material events (neural activity), located in a material object (the brain), while holding that the final truth of material events and material objects is captured in the laws of physics. The trouble with physical science, however, is that it is committed to seeing the world in the absence of consciousness (at least prior to quantum mechanics); indeed, at its heart is the *disappearance of appearance.* This presents not one but three insuperable problems for Neuromania. They are inextricably connected but it is helpful to address them separately: the first concerns the nature of nerve impulses; the second is about the things nerve impulses are supposed to make appear; and the third relates to the supposed capability of nerve impulses to make those things appear.

Nerve impulses don't have an intrinsic nature

Let us get back to basics. If I claimed that consciousness was identical with neural activity then you might reasonably assume that I had a clear idea what I meant by "neural activity". We have already seen that there are serious ambiguities in this concept, which leaves it unclear how we should think of what goes on in those parts of the brain that are supposedly associated with subjective experience. Is "neural activity" something that is delivered to a certain place in the brain? Or is it the sum total of what is happening in several places of the brain? If so, where is the summing and the totalling taking place? Does consciousness reside in the travelling of nerve impulses along neurons or its arrival at a synapse? These questions invite us to look more closely at what we think a nerve impulse is in *itself.*

You may think this had been spelled out in (fairly piti-less) detail in Chapter 1. It will be noticed, however, that the nerve impulse could be described in different ways. Here are some:

- a *cycle* of events taking place at a particular point on the membrane that occurs *over a time* (of the order of milliseconds) represented by a wave or a spike traced on an oscilloscope screen; in other words, the *sum total* of the changes in the potential difference across a particular point in the membrane;

- the overall journey of the wave along the length of the nerve axon; a propagated

displacement of the alterations in the potential difference along the length of the axon; a displacement of a displacement;

- part of a summed total of many millions of nerve impulses as seen on an EEG or inferred from an fMRI scan.

Since the nerve impulse may be represented with equal validity as being any of these things, it is *intrinsically* none of them. Which properties are ascribed to it are observer-dependent.

To put it slightly differently, there are different "takes" on a nerve impulse. It could be seen as an influx of sodium ions at a particular point in the neuron followed by an efflux of positive ions; or as a change of the potential difference between the inside and the outside of the membrane at a particular place; or as a succession of events, lasting about a millisecond, at a particular point in the neuron; or as a wave of activity at that point; or as a wave moving along the neuron; or as a wave arriving rather than travelling; or as one of a crowd of waves, several thousand, several million or several billion strong, occurring in a particular place in the brain. There are many other candidates, for example patches of coloured pixels in brain scans or brain maps. But I hope the point will have been made: the nerve impulse is not *in itself* a local passage of sodium ions or *in itself* part of a billion-strong crowd of waves; otherwise it would have to be both of these at the same time.

And it is not just a matter of how the impulse appears. What a nerve impulse is depends on how it is viewed. A micro-pipette recording from a single neuron will deliver a different account of a nerve impulse compared with an EEG recording large-scale activity through the skull. Or, to draw the conclusion that should be obvious to anyone who is not ideologically wedded to Neuromania, the nerve impulse does not have any intrinsic determinate character. It depends how it is looked at, on how it is teased apart or put together. We are deceived if we think that scientific instruments reveal what it is "in itself". It is easy to overlook this when we confuse the representation(s) of the nerve impulse with the thing in itself. We are less likely to do so if we remind ourselves that there are many competing ways of representing a nerve impulse. The nerve impulse requires a viewpoint (provided by a highly mediated consciousness involving sophisticated scientific instrumentation) to be either an instantaneous displacement in potential difference at a particular point in space and time; or a spike extended over a short time at a particular place; or a spike moving over space and time; or a member of a crowd of spikes moving over space and time and spreading over space and building up over time.

Anyone who still thinks that neural activity has an intrinsic appearance that is independent of observers might want to reflect on the following final twist. Some of the ways we may represent nerve impulses to ourselves can be analysed into two or more takes that correspond to incompatible viewpoints. For example, seeing the impulse as a travelling spike requires an observation *over time* at a particular place (this generates the image of the spike) *and* observation at successive places. But temporal depth, as we discussed in the previous section, is not to be found in matter - or in material events such as nerve impulses.

Material objects do not have (phenomenal) appearances when viewed through the eyes of physics

Nerve impulses are not uniquely impoverished in having no intrinsic appearances. This lack characterizes the entire material world as seen through the eyes of physical science. This was noted early in the history of modern science. Galileo - and subsequently philosophers such as Descartes and Locke - marginalized most of the things that make up the *appearance* of material objects as being (mere) ''secondary qualities". Colours, tastes,

smells, sounds and so on exist only where there are observers and they do not correspond to what, according to physical science, is objectively there. As Galileo said, “If the living creature were removed, all these qualities would be wiped away and annihilated”. The material world has only primary qualities such as solidity, extension, motion, number and shape. These by themselves would not, however, amount to a full-blown appearance. You couldn’t imagine an object without a colour (and “colour” here includes black and white). Primary qualities by themselves don’t really amount to much. An object such as a cup reduced to its primary qualities would not only lack colour, but also features such as being near or far, looking small or large, and being related to this object rather than that. Indeed, it would boil down to naked *numbers* that capture (abstracted) shape, motion, size and so on. This is what lies behind Galileo’s famous assertion that the book of nature is written in mathematical language. One manifestation of this view is connected with the centrality of measurement in all sciences, the reduction in physical sciences of the phenomenal world to numerical quantities and the unfolding of events to the relations between quantities, ultimately expressed in equations. The output of measurement is a number: of abstract units, or patterns of numbers of abstract units or general laws connecting numbers of abstract units.

Let’s illustrate with a simple example of what happens when we progress from immediate (subjective) experience to (objective) measurement. Imagine you and I are looking at a table. Because we are looking at it from different angles it seems square to you and oblong to me. What’s more, I think it is bigger than another table and you think it is smaller. We decide to settle our disagreement by taking a measurement, and discover that it measures 100 cm x 75 cm. End of argument; but also end of the appearance of the table. It is no longer “square”-looking or "oblong”-looking, nor “bigger” or "smaller”. It loses these qualities and, in addition, it lacks position and relation to us. We have replaced its appearance by two numbers. You might want to argue that there is a *residue* of appearance: the appearances that are necessary to make the measurement; for example the appearance of the ruler next to the table. But of course, these appearances are set aside once we have the result: "100 cm x 75 cm” gives no hint of the appearance of the devices (the tape measure or ruler) by which the measurement was made or of the processes that led up to the measurement. They are as irrelevant as a quarrel over which side of the tape measure to use. And the actual appearance of the measurement as written down - "100 cm x 75 cm” - is equally irrelevant. It would not matter whether the result was recorded in blue ink or black, was written as "1 m x 0.75 m” or “1000 mm * 750 mm” or "one hundred centimetres by seventy-five centimetres”, or whether it was spoken or presented on a screen.

We seem, therefore, to have a *disappearance of appearance* as we move from subjective experience towards the scientific, quantitative and ultimately mathematical account of the world as matter. This loss of appearance is strikingly illustrated by those great equations that encompass the sum total of appearances, such as *"e = me^2".* But it is also present at a more homely level when we try to envisage material objects as they are in themselves. Think of a rock. I can look at the rock from the front or from the back, from above or below, from near or far, in bright light or dim. In each of an (innumerable) range of possible circumstances, it will have a slightly or radically different appearance. *In itself,* it has no definite appearance; it simply offers the possibility of an appearance to a potential observer (although those possibilities are constrained - the rock cannot look like a sonnet). So we can see that, as we get closer to the material world "in- itself”, as a piece of matter, so we lose appearances: colour, nearness or farness, perspective. (The history of science, which is that of progress towards greater generalization is a gradual shedding of perspective - a journey towards Nagel’s “view from nowhere”.)

You might want to say that it still has *primary* qualities. Weight, size and shape may exist independently of any consciousness, as is evident from the fact that the rock

may have an impact irrespective of any perceiver. It may provide shelter to grass, stop

the dampness in the soil underneath it from drying out so quickly, arrest the path of another rock rolling down the hill, cast shadows and so on. Primary qualities, however, do not add up to an appearance. A rock does not have the wherewithal to generate the way it would appear in consciousness, even less “from a long way off” or “from close to”. It is, of course, potentially, all these things, but the potential will not be realized unless it is *observed.* If those appearances were intrinsic rather than merely potential, if they were in the rock itself, then the item would be in conflict with itself: trying, for example, to look as it does from far off and from nearby at the same time. Like the nerve impulse, the rock - or indeed any other material object considered, in the absence of an observer, as matter - does not have an appearance.

To summarize, such appearances as material objects do have are the "takes” that external observers - or an entire community of scientific observers coming to a conclusion about the appropriate way(s) to represent them - have on them. While the object provides certain constraints on takes, it does not of itself deliver takes; takes require consciousness; indeed, consciousness is made up of takes. Matter has to have an angle, a viewpoint, a perspective, to support awareness of a world. It has none of these things intrinsically. Material objects as viewed by physics "in themselves”, as matter, have no appearances. The very notion of a complete account of the world in physical terms is of a world without appearance and hence a world without consciousness.

Nothing in appearance-less nerve impulses suggests that they have the ability to make appearanceless material things acquire (phenomenal) appearances

So far we have arrived at two conclusions: first, nerve impulses do not have definite appearances or phenomenal character in themselves; and, second, they share this lack with all material items when the latter are considered independently of an observer, most obviously when they are seen through the eyeless mathematical eyes of physics. We are now in a position to see the *inherent contradiction* of trying to find consciousness in nerve impulses or, more broadly, to see consciousness as a property arising out of certain events in the material world, where matter is as defined by physics. Consciousness is, at the basic level, appearances or appearings-to, but neither nerve impulses nor the material world have appearances. So there is absolutely no basis for the assumption, central to Neuromania, that the intrinsically appearance-less material world will flower into appearance to a bit of that world (the brain) as a result of the particular material properties of that bit of the world: for example, its ability to control the passage of sodium ions through semi-permeable membranes. We cannot expect to find anything in a material object, however fashioned, that can explain the difference between a thought and a pebble, or between a supposedly thoughtful brain and a definitely thoughtless kidney. And there is even more obviously nothing in the difference between a spinal cord and a cerebral cortex to explain why the former should be unaware and thoughtless and the latter (in parts) aware and thoughtful. This makes more obviously barmy the idea that nerve impulses can journey towards a place where they become consciousness: that, by moving from one material place to another they are mysteriously able to be the appearance of things other than themselves. If this is physics, it is not the physics to be found in textbooks.

The difficulty of seeing how nerve impulses could confer appearance on the material world has led some to suggest that we do not experience the material world as such, only nerve impulses. Iain McGilchrist, whose extraordinary *The Master and His Emissary* represents Neuromania at its most extreme, asserts that "one could call ‘the mind’ the brain’s experience of itself”, and many others have suggested that consciousness is our perception of some physical processes in the brain: in short, that consciousness and appearance are made of the appearance of nerve impulses to themselves! Leaving aside

what we have already established, that nerve impulses do not have a definite appearance apart from a viewpoint that has a certain take on them, there is no reason why they should be riddled with a self-awareness that is, mysteriously, awareness of the material world that is their immediate or remote cause: that their unique self-awareness should be awareness of a world that is other than them.

It is no help moving away from matter and appealing to the energy of mass energy. Just as matter itself, by definition, ex *officio,* as it were, does not have an appearance corresponding to the kind of things we experience in consciousness, no more does energy. There is nothing in either corresponding to my seeing a rock. The light-mediated rubbing together of an appearance-less object (my brain) with appearance-less light arising from an appearance-less object (the rock) is hardly going to explain the appearance of the rock to me, the owner of the brain, even less my sense that the rock is independent of me (a foundational intuition of physics and the folk metaphysics of everyday life) or that it has the potential to yield an infinity of other different appearances to ourselves and other people (the foundational intuition of the public world we humans live in).

So, the neural theory of consciousness is at odds with the very notion of matter that lies at the heart of the "orthodoxy" - to use Dennett's word - that underpins it. The objects that surround us analysed as elementary particles are remote from the phenomenal world experienced and lived in by conscious beings. As the scientific gaze goes beyond ordinary objects, perceived in the ordinary way, to their underlying material reality, so it progresses from things that have qualities to things that are characterized by numbers. It is not by accident that atoms are colourless, odourless and so on, and are defined by numbers that capture their size, speed and quantities; that experiences and experienced phenomena are replaced by numbers, patterns, and laws; that the progress of physical science is characterized by a progressive disappearance of appearance.

Further empirical research, therefore, within the current way of understanding the problem will not take us any closer to a neural explanation of consciousness. What is needed is a revolution in the way in which we approach the problem. This may require us to see that it is more than a problem, or even to see that it is more than "a hard problem". It is a mystery.

1. Crawford, "The Limits of Neuro-Talk", 65.
2. Vrecko, "Neuroscience, Power and Culture".
3. Beauregard *et al.,* "The Neural Basis of Unconditional Love".
4. /bid., 96.
5. Whether unconditional love is what low-paid care workers feel is not at all clear. Because they do it for less money may not mean that they do it for more love. They may be able to find no better-paid employment. At any rate, I would be surprised if the ideal of unconditional love would survive the removal of the salary as a condition of caring.
6. Zeki & Kawabata, "Neural Correlates of Beauty".
7. Leake, "Found: The Brain's Centre of Wisdom".
8. Jeste & Meeks, "Neurobiology of Wisdom".
9. /bid., 355.
10. *Ibid.*
11. Leake, "Found: The Brain's Centre of Wisdom".
12. Bartels & Zeki, "The Neural Basis of Romantic Love".
14. Kong *et al.,* "Test-Retest Study of fMRI Signal Change".
15. McClure *et al.* "Separate Neural Systems".
16. *Ibid.,* 506.
17. Quoted in Lehrer, "The Psychology of Subprime Mortgages".
18. Eisenberger *et al.* "Does Rejection Hurt?"
19. Vul *et al.,* "Puzzlingly High Correlations in fMRI Studies".

20. *Ibid., 285.*
21. *Ibid.*
22. *Ibid.*
23. *New Scientist,* "What Were the Neuroscientists Thinking?"
24. Bennett & Miller, "How Reliable are the Results?".
25. Dobbs, "Fact or Phrenology?", 24.
26. Haynes & Rees, "Predicting the Orientation of Invisible Stimuli".
27. Damasio, *The Feeling of What Happens,* excerpted in Harvey Wood & Byatt, *Memory: An Anthology,* 282.
28. Raichle *et al.,* "Practice Related Changes in Human Brain Function Anatomy".
29. Quoted in Le Fanu's profound *Why Us?,* 195.
30. Zeki & Goodenough, *Law and the Brain,* 218.
31. Searle, *Intentionality.*
32. Hood, *Supersense,* 231.
33. This is accessibly summarized in the admirable and admirably gen erous internet *Stanford Encyclopaedia of Philosophy:* Brennan, "Necessary and Sufficient Conditions".
34. His findings and the excitement of being present in the operating theatre when he was doing his work are beautifully described in a commemorative article by his very distinguished *protege* Brenda Milner: "Wilder Penfield".
35. See Putnam, *Reason, Truth and History.*
36. Again the *Stanford Encyclopaedia of Philosophy* is an excellent guide. See Brueckner, "Brains in a Vat".
37. Pfeiffer, pers. comm.
38. It is a healthy sign that the doctrine of "disjunctivism", originally proposed by my friend Howard Robinson, is now catching on. According to this doctrine, hallucinations have nothing in common with genuine perceptions apart from the fact that they seem the same to the person experiencing them. Too right.
39. For an excellent discussion of this concept and the scientific hunt for such correlates, see Rees & Frith, "Methodologies for Identifying Neural Correlates".
40. Under certain circumstances, a very large amount of neural activity can be associated with loss of consciousness, as seen in an epileptic fit, where giant waves of synchronized activity across the cortex blot out awareness.
41. Coleridge, Notebook 21, *Coleridge Notebooks* ii, 2370.
42. For those who are tempted to entertain the idea that qualia are not real, Searle's savage review of Dennett's book (reprinted in *The Mystery of Consciousness)* will convince them that they are real, and give them a lot of pleasure *en route.*
43. Searle, "Biological Naturalism", 327.
44. For a detailed account and critique of this theory, see my "The Causal Theory of Perception", in *The Explicit Animal.*
45. Dennett has expressed this view over many years but his most comprehensive statement of it is in *The Intentional Stance.*
46. Frith, *Making up the Mind.* The reader might be interested in my critique: "Not All in the Brain".
47. Even those who are sympathetic to the argument I have presented here may be uncomfortable with my dealing with perception as if it were independent of action, and argue that this separation is artificial. It is of course true that, biologically, perception is the servant of action and that even when we have loosened the bonds of biology what we perceive is a world that requires us to act or gives us an opportunity to act or shapes our ongoing actions. Nevertheless, the fact that we can perceive without acting, that we have the capacity to postpone and plan action, that we are able to contemplate the world without any intention of acting, is what ultimately leads to knowledge, to the know-

that that uniquely informs human know-how, that sets us off from the world and enables us to act on it so effectively. That perception is not an irresistible trigger to action is in part the basis of the sense that our actions are *ours* and that it is *we* who are acting. This in turn underlines the sense of self and our sense of leading our lives rather than merely living or suffering them.

48. "Egocentric space" is discussed in my *The Kingdom of Infinite Space.*

49. For a profound exploration of the implications of conceiving of a world without viewpoints, see Shand, "Limits, Perspectives and Thought".

50. Materialist neuroscience is equally impotent when it comes to dealing with second-person being.

51. This problem of perceiving the light that is the condition of our seeing, that is, of seeing background illumination separate from that which is illuminated, is analogous to another profound problem, to which neuroscience offers no solution: the difference between the level and content of consciousness; between that which we are conscious of and the state of wakefulness that allows us to be conscious of it. The various solutions on offer - mass activity in certain pathways (the thalamo-cortical pathways) - do not address the problems of the unity of consciousness.

52. Alhough (very) hard going, Kant's *Critique of Pure Reason* is worth struggling with because it addresses the problem at the right depth.

53.1 discussed and criticized this in my first foray into the field *(The Explicit Animal),* but to no avail.

54. Hameroff & Penrose, "Orchestrated Reduction of Quantum Coherence".

55. This is a view put forward by McFadden," The Conscious Electromagnetic Field (Cemi) Field Theory".

56. The key paper is Crick & Koch, "Towards a Neurobiological Theory of Consciousness". This notion comes back again and again. In a recent paper that purported to explain memories by "brain entanglement", the authors note that the voltage of the electrical signal in groups of neurons separated by up to 10 mm sometimes rose and fell with exactly the same rhythm and adopted the same amplitude: a phenomenon variously called "coherence potentials" or "phase-locking" (Tharagarjan *et al.,* "Coherence Potentials"). The precision with which these new sites pick up the activity of the initiating group is extraordinary; they were perfect clones. Since the coherence potentials seemed unique, they argue that each could represent a different memory, as if that distributed signature could be gathered up into a unified memory. Even if this claim were not riddled with conceptual problems, an empirical observation reported in the same paper does it in completely. The coherence potentials were not unique to those cells associated with memory as we understand the term: they were also seen in dish-grown neural cultures - hardly the site for nostalgia.

57. Crick & Koch, "What is the Function of the Claustrum?"

58. *Ibid.,* 1272.

59. Ibid.

60. Quoted in Le Fanu's profound *Why Us?, 195.*

61. Hume, *A Treatise of Human Nature,* I, pt I, § 1.

62. Quoted in Isaacson, *Einstein: His Life and Universe,* 540.

63. Quoted in Rose, "Memories are Made of This", 61.

64. Kandel's work and his philosophy are accessibly summarized in his Nobel Laureate lecture, "The Molecular Biology of Memory Storage".

65. We may recall that memory has been found in another dish by Tharagarjan *et al.,* "Coherence Potentials".

66. The intentionality of *autobiographical* memories is arguably more than double: perhaps five-way! After all, the relevant brain activity would have to be "about" a current memory; the latter "about" a past event. It would also be "about" the time elapsed from the remembered event to the present; "about" a past

world that the memory belongs to; and "about" the "I" to whose world it was. And there are, of course, memories belonging to m ore than one individual and each may support the other in recalling what happened. The French philosopher Maurice Halbwachs, writing in the first half of the twentieth century, referred to the "social frameworks of memory" (On *Collective Memory,* pt 1).

67. McCrone, "Not So Total Recall".

68. Wang *et al.,* "Synapse- and Stimulus-Specific Local Translation".

69. Gelbard-Sagiv *et al.,* "Internally Generated Reactivation of Single Neurons".

70. Bergson, *Matter and Memory, 176.*

71. James *The Principles of Psychology,* 628-9.

72. Bergson *Matter and Memory,* 177.

73. LeDoux, *Synaptic Self.*

74. A similar view was first suggested by Parmenides, who argued that being could not change and time was unreal (see my *The Enduring Significance of Parmenides').*

75. As summarized by Richard Gale, Introduction to "Section IV: Human Time", in *The Philosophy of Time,* 300.

76. Crick & Koch, "Towards a Neurobiological Theory of Consciousness".

77. Crick & Koch, "A Neurobiological Framework for Consciousness".

78. Consider a succession of events El to E100,000. Event E3, although it occurs early on, is not in itself intrinsically in the past nor is E85,OOO intrinsically in the future, although it occurs late on. We require a *viewpoint* to locate an event in the present, the past or the future. The viewpoint establishes the relation between present events and past events. From a viewpoint simultaneous with event E3, E3 is in the present, El is in the past and E85,OOO is in the future. From a viewpoint simultaneous with El, E3 is in the future, as is E85,OOO. From a viewpoint simultaneous with E100,000, El, E3 and E85,OOO are in the past and E100,000 is in the present. The relationship of the many hundreds of thousands of discrete events in the nervous system as past, present or future therefore requires a viewpoint that experiences or observes them as occurring simultaneously or in succession. According to Crick and Koch, "Towards a Neurobiological Theory of Consciousness", however, that viewpoint is precisely what the synchronous activity is supposed to construct. It will now be obvious that supposed binding activity will not bind multi-itemed consciousness into the moments of a unified conscious self unless it has itself already been bound into a unity by a unified conscious self. The factual simultaneity of the neural activity does not translate into a unity, in part because it would still have to be *observed* for the translation to be made and in part because they are not, for reasons already given, intrinsically simultaneous. What is more, because the elements are spatially separate, even though the separation is a matter of centimetres, the limits of the synchronous activity would have to be observed at different times by an observer. The belief that they *are* intrinsically simultaneous is the result of inserting into a material world the idea of an observer.

79. Raby *et al.,* "Planning for the Future".

80. Stebbing, "Furniture of the Earth", 78.

81. Dennett, *Consciousness Explained,* 33.

82. Galileo Galilei, *The Assayer,* 274.

83. The idea that we get closer to the essence of something as we progressively abstract from it towards mathematics most certainly does not apply to consciousness. This does not stop neuromaniacs such as Paul Churchland suggesting that sensations really boil down to spiking frequencies in different vector spaces of the brain. See Churchland, *Matter and Consciousness.*

84. Nagel, *The View from Nowhere.*

85. McGilchrist, *The Master and His Emissary,* 19. Indeed, this "solution" makes things worse for the neural theory of consciousness. Consider my consciousness of

this rock in front of me. There is no such thing (within the rock) as "what it is like to be" that rock. And there is no such thing as what it is like to be my body *qua* organism. And there is no such thing as what it is like to be my brain understood as a material object. The McGilchrist version of the neural theory requires all three things, if I am to be aware of a rock. Most mysteriously, it requires that “what it is like to be a brain" should be the revelation of what it is like to be a body and what a rock is like.