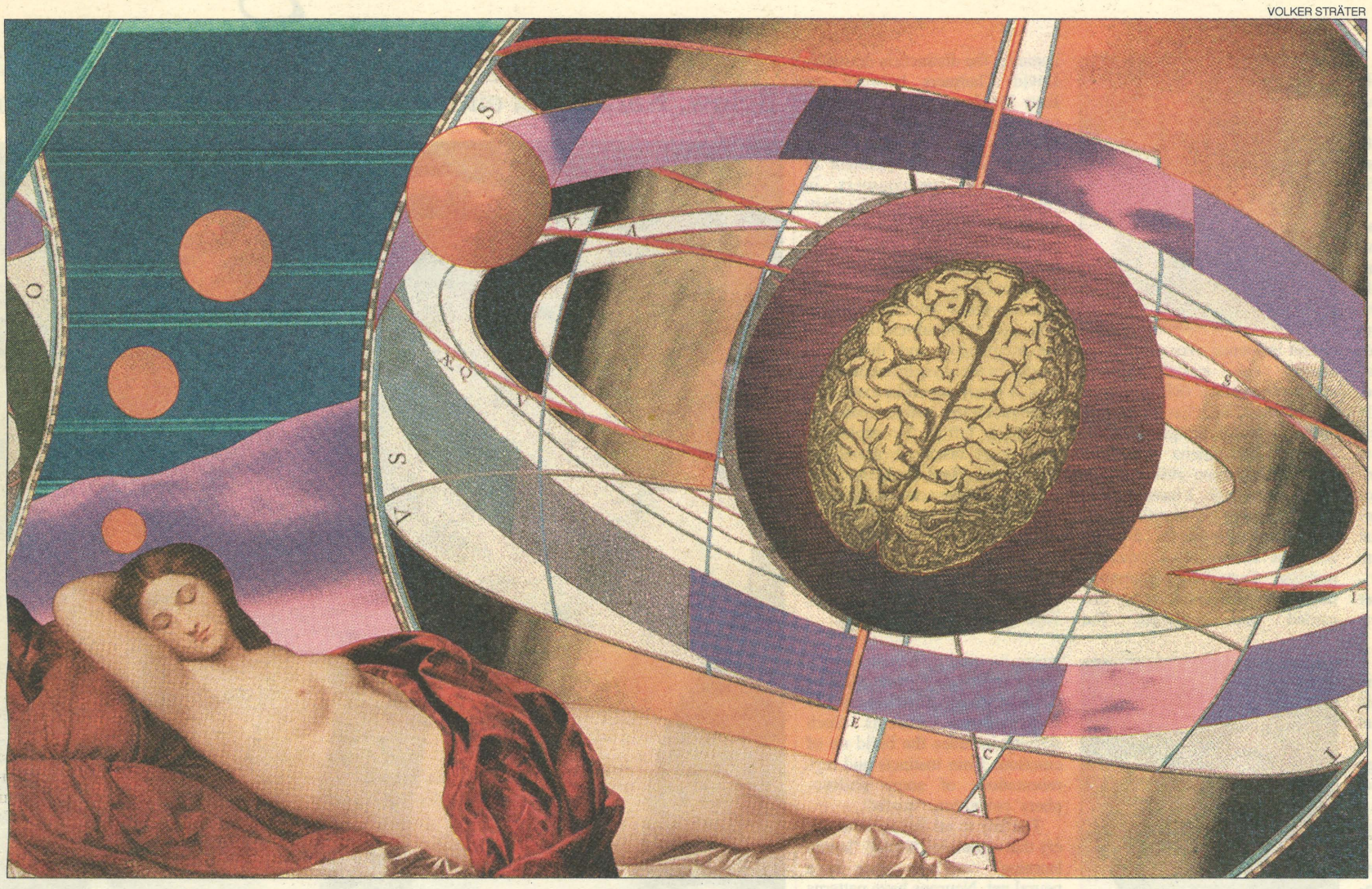
TOWARDS A NEW SCIENCE OF THE MIND: TUCSON 1996

consciousness



VOLKER STRÄTER

APRIL 5 1996

contents

THESIS

The Times Higher Education Supplement Internet Service, is broadcasting the Tucson II plenary sessions from its web site, thesis. newsint. co.uk. You will need a sound card to hear the lectures, which you will be able to do after downloading Real Audio software from the site. The site also contains a wealth of background material for the conference and will have text papers and reports as the conference unfolds.

Science's inner frontier

BY ANDREW ROBINSON

When you fall into deep sleep, what happens to your consciousness, to "you"? Does it hibernate or dissipate or what? Does it still exist? Your body continues to exist, your brain continues to control your body's rhythms — otherwise you would never wake up — but do "you" exist in oblivion?

The celebrated physicist Richard Feynman became interested in this question while he was a student at the Massachusetts Institute of Technology in the 1930s. He asked himself, do my thoughts suddenly stop as I fall asleep, or do they move less and less rapidly, or what? Feynman decided to study himself falling asleep. He observed that "the ideas continue, but they become less and less logically interconnected. You don't notice that they're not logically connected until you ask yourself 'What made me think of that?' and you try to work your way back, and often you can't remember what the hell did make you think of that!"

After four weeks of self-experiment, Feynman concluded that, while it was possible to watch himself falling asleep, "I don't really know what it's like to fall asleep when I'm *not* watching myself". He composed a short verse on the problem of introspection:

I wonder why. I wonder why.
I wonder why I wonder.
I wonder why I wonder why
I wonder why I wonder!

How can "I" observe "I"? What are we really seeing when we stare into the depths of our own eyes in a mirror? Another great physicist, Erwin Schrödinger, one of the founders of quantum theory, pondered on the problem in the tantalising epilogue of his classic *What is Life?* "Consciousness is never experienced in the plural, only in the singular. Even in the pathological cases of split consciousness or double personality the two persons alternate, they are never manifest simultaneously. In a dream we do perform several characters at the same time, but not indiscriminately: we are one of them; in him we act and speak directly, while we often eagerly await the answer or response of another person, unaware of the fact that it is we who control his movements and his speech just as much as our own." Schrödinger encapsulated the problem of consciousness in the form of two premises:

"(i) My body functions as a pure mechanism according to the laws of nature.

(ii) Yet I know, by incontrovertible direct experience, that I am directing its motions, of which I foresee the effects, that may be fateful and all-important, in which case I feel and take full responsibility for them."

To avoid a contradiction here, he said, "The only possible inference from these two facts is, I think, that I — I in the widest meaning of the word, that is to say, every conscious mind that has ever said or felt 'I' — am the person, if any, who controls the 'motion of the atoms' according to the laws of nature." And this would lead you to say, Schrödinger provocatively suggested: "Hence I am God Almighty."

Though even today to many western ears such a statement sounds both "blasphemous and lunatic"—and in 1943 it caused the rejection of *What is Life?* by its original (Catholic) publisher — the idea is hardly new. As its author noted, this "grandest of all thoughts" was recorded in the *Upanishads* more than 2,500 years ago, and has long been considered the deepest insight in Indian philosophy. Surely, said Schrödinger, the *singularity* of consciousness is more intuitively convincing than the western idea of a plurality of consciousnesses, which leads inevitably to the invention of souls — as many as there are bodies — and to unhelpful questions such as whether the soul survives death and whether animals (and bacteria) have souls? Towards the end of his life Schrödinger stated: "The world is a construct of our sensations, perceptions and memories. It is convenient to regard it as existing objectively on its own. But it certainly does not become manifest by its mere existence."

His friend and scientific colleague Albert Einstein could never bring himself to agree (and thus could never accept that quantum theory was the fundamental description of nature). Nature, for Einstein, had to be independent of human consciousness. In 1930, arguing with the Indian poet Rabindranath Tagore, Einstein stated: "Man defends himself from being regarded as an impotent object in the course of the universe. But should the lawfulness of events, such as unveils itself more or less clearly in inorganic nature, cease to function in front of the activities in our brain?

"Leaving aside the inconsistency of such a view, the influence of alcohol and other sharply controllable factors on our thoughts, feelings and activities should show very distinctly that determinism does not stop before the majesty of our human will.

"Maybe, we and human society require the illusion of freedom in our human activities! … I believe that whatever we do or live for has its causality; it is good, however, that we cannot look through it."

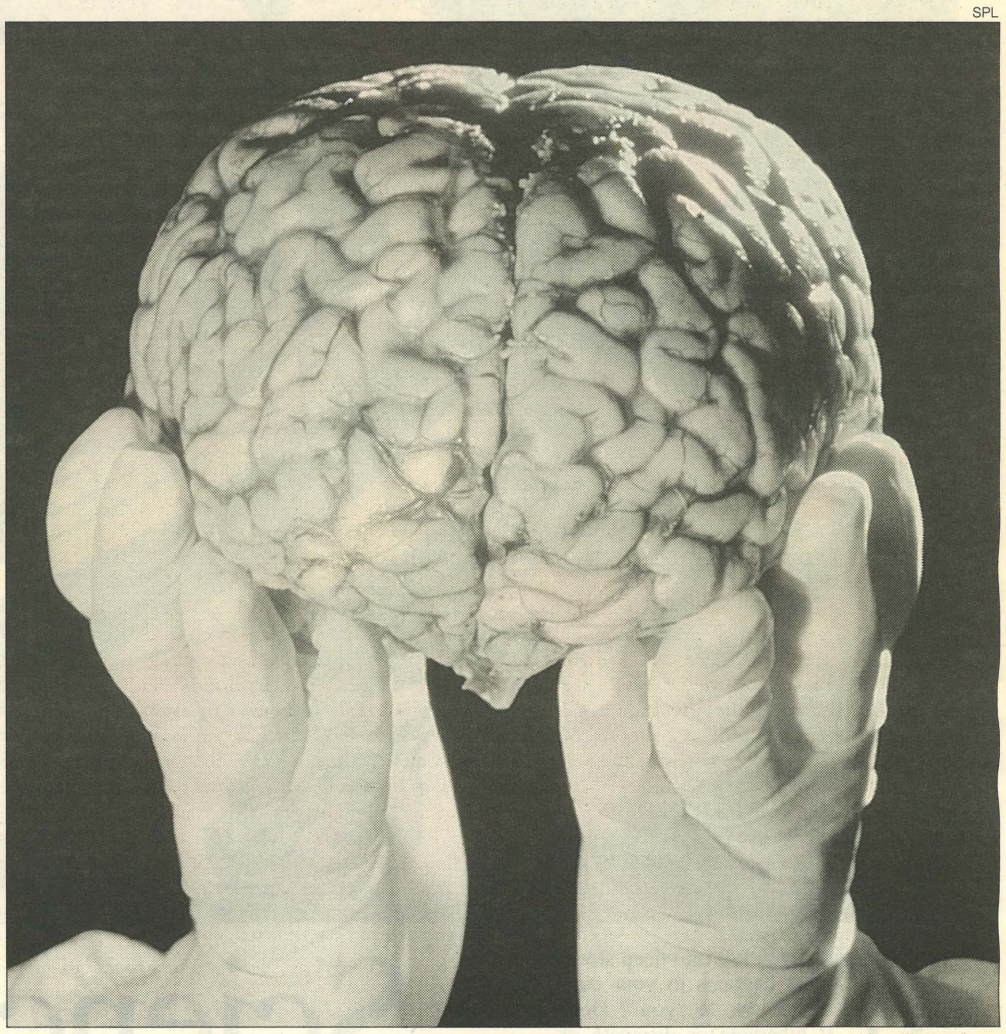
The nature of causality, determinism and free will continue to underlie the burgeoning scientific debate about the nature of consciousness. As Alwyn Scott wrote in 1995 in his *Stairway to the Mind*, any physicist who chose to tell a major scientific meeting that he believed in an omniscient God would most likely be written off as a "misguided fundamentalist" — if instead he were to profess belief in a Theory of Everything that determines every fact of the future from the facts of the past, many would welcome him. But what is new and exciting in the 1990s, is that through technological advances in many fields — bringing a vastly increased sensitivity and diversity of technique to the study of the brain — science is at last becoming capable of investigating old questions about mind and brain empirically. As a result, physicists, mathematicians, computer scientists, chemists, biologists, geneticists, psychologists, psychiatrists, philosophers, linguists, anthropologists, theologians and others, even mystics, are now listening to each other with renewed interest. As the articles and reviews in this special issue demonstrate, science cannot yet encompass the mystery of our introspections, but it is beginning to move in that direction.

Hume's internal bundle grows

Thinking about consciousness is no longer the

Below, **Kam Patel** outlines the main issues, while experts

SPL



When I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception, and never can observe anything but the perception.

David Hume
A Treatise of Human Nature (1740)

Most of us may feel we can identify with the 18th-century philosopher David Hume's inner world of thoughts and feelings. It is a private world which we would recognise as existing in some way within ourselves. But explaining the nature of what Hume thought of as his "bundle of internal perceptions" — and what 20th-century thinkers prefer to call "consciousness" — is a problem philosophers have wrestled with for hundreds of years.

In the past ten years the question has moved beyond philosophy; igniting a surge of interest among scientists working in disciplines as diverse as neuroscience, cognitive science, psychology and artificial intelligence.

Although a bewildering number of approaches are being deployed, it is possible to trace a number of key assumptions underpinning much of this interdisciplinary work.

Most scientists and philosophers assume that consciousness emerges from the operation of nerve cells in the brain (neurons). A notable opponent of this (seemingly obvious) view is the mathematician Roger Penrose, who controversially argues that consciousness is generated by quantum activity that occurs in tiny parts of the brain's nerve cells called microtubules. Penrose also maintains that computers can never emulate human understanding and therefore never attain consciousness, a claim that has attracted fierce criticism from the artificial intelligence community.

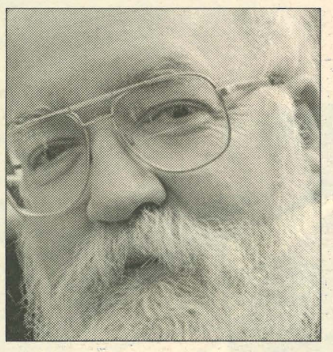
More generally, arguments over approaches adopted by most researchers and thinkers in the field become focused when two fundamental issues are considered. The first of these revolves around the relationship between conscious experience and the person, organism or artificial system owning that experience. The second concerns the ability of a person, organism or system to describe and understand the conscious experience of another.

Some thinkers, such as John Searle, base their theories on the essential privacy of the conscious experience and take what can be called a "first-person perspective" on the problem. Searle argues that consciousness should be seen as a "high level feature" of the workings of the neuronal system. But he stresses the privacy of that feature: "There is a sense in which each person's consciousness is private to that person, a sense in which he or she is related to his pains, tickles, itches, thoughts and feelings in a way that is quite unlike the ways others are related to them". This means, he says, that the essence of consciousness — a subjective, qualitative phenomena — cannot be described purely by examining the subject's behaviour or through efforts to build computational models of consciousness. Attempts to study the conscious experience as if it were a third-person phenomenon are, according to Searle, doomed to failure. Meanwhile some others committed to the first-person point of view argue for "qualia", the absolute, inexplicable, unassailability of first-person sensations such the redness of red or the *thrill* of seeing a beautiful sunset.

Putting forward a counter-argument to those in the first-person camp are thinkers who argue that while consciousness may appear to be a purely subjective experience, this should not prevent the development of a description of the conscious experience from a third-person perspective. The philosopher Daniel Dennett believes that the power of an objective, third-person driven analysis of consciousness is "woefully underestimated" by the first-person camp. He believes that it is quite possible objectively to capture everything about a person's conscious experience by scrupulous, patient and subtle experimental analysis. "You let the subjects tell you what it is like to be them. You

---

Daniel Dennett



Old habits die hard, especially habits of thinking, and our "intuitive" ways of thinking about consciousness are infected with leftover Cartesian images, an underestimated legacy of the dualist past. Of course the brain is the seat of consciousness, and all the phenomena that compose our pre-theoretical catalogue of conscious phenomena are ultimately explicable in terms of the activities in our brains and bodies, but the paths of explanation (or "reduction") are not as direct as many materialists have supposed.

My work has two roughly equal components. On the positive side, I have tried to show how to construct a truly non-Cartesian materialist theory of consciousness, in which conscious events are not those that occur in any privileged medium in the brain, but those which triumph in competition with other events, and hence have more persistent influence over subsequent events. This model explains many phenomena that otherwise are baffling, and has even predicted a few strikingly counter-intuitive phenomena that have subsequently been observed. On the negative side, I have tried to show theorists in several disciplines how their presumed-to-be-innocent formulations typically harbour Cartesian presuppositions that still need to be discarded and replaced. For instance, the distinction drawn by David Chalmers between the Easy Problems of consciousness (questions about the mechanics of nerves and brain cells) and the Hard Problem (the problem of phenomenal consciousness or "qualia") is not, I argue, the fruitful insight many have taken it to be, but a symptom of the failure to appreciate that all the work of consciousness, including the "phenomenal" work of appreciation and emotional reaction, must be fragmented and distributed around in the asynchronous activities of many networks extending throughout the body. There are no qualia left to be explained, once these tasks are accounted for, so the Hard Problem is an artefact of false accounting; once all the Easy Problems are solved, consciousness is explained.

Daniel Dennett is director of the Centre for Cognitive Studies, Tufts University.

Igor Aleksander

I am working on a programme involving neural networks. This research may enable an artificial system to have a point of view of its own, one that includes enough knowledge and experience to let it consider itself to be conscious.

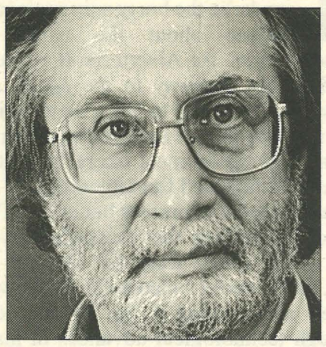
This is not through some precise definition of consciousness, but through the same notions we all have about our own consciousness. We know we are conscious without being able to say exactly what consciousness means. I can say roughly that it means first of all that I am sufficiently awake to notice my surroundings. Then, even if I close my eyes, and I am in a quiet room, I still know that I am conscious because I have a sensation that I call "thought". This consists of a variety of internal sensations that are a bit like that which I sense with my eyes open. I can even use natural language to describe my thoughts to someone else. I call this a "folk" description of consciousness.

The machine I am using, called Magnus, is a neural net. It can be configured by an experimenter to test hypotheses about how representations of sensation may be created that resemble the perceptual sensations themselves. They are also capable of representing that which they themselves might do with artificial actuators (hands, fingers, voice chords) — that is, an awareness of self. Also central to our Magnus project is the absorption of natural language, as used by human beings. This then develops, with Magnus learning about named objects and more abstract concepts. Such activity includes the build up of emotions from instincts and representations of some philosophers' pet notion, "qualia".

The key to all this is a discovery that something we call "iconic learning" can take place in a neural net. Neurons have patterns of behaviour when they are exposed to sensory input. Iconic learning is a phenomenon of these patterns being sustained when the perceptual input is no longer there. Such patterns are called "states" and the world of the organism is represented by a rich structure of such states, which is the seat of the organism's mechanism for consciousness. So, if such things go on in Magnus, why not in living beings?

Igor Aleksander is professor of neural systems engineering, Imperial College, London.

David Rosenthal



The problem of consciousness is to say what it is for some of our thoughts, feelings, and sensations to be conscious, given that others are not. This is different from saying what it is for a person to be conscious or not conscious. Even when people are conscious, many of their thoughts and sensations typically are not. And there is nothing problematic about a person being conscious — it is just the person being awake and responsive to sensory input.

When a thought or feeling is conscious, one is always conscious of that thought or feeling. Being conscious of things means either sensing them or having thoughts about them. And the thought or sensation that makes us conscious of something need not itself be conscious; even sensing subliminally makes us conscious of things in a certain way.

We do not *sense* our conscious thoughts and sensations — there is no sense organ for doing so. The only alternative is that we are conscious of our conscious thoughts, feelings, and sensations by having *thoughts* about them. These "higher order thoughts" are themselves seldom conscious; so we are unaware of them.

This model explains many striking phenomena. For example, creatures that can talk about thoughts and sensations at all can readily describe their own conscious thoughts and sensations, but not those that are not conscious. The model predicts this. When one says that one has a thought or sensation, one expresses a *thought about* that thought or sensation; indeed, one expresses the very higher order thought in virtue of which one is conscious of that thought or sensation. Similarly, we cannot report thoughts and sensations that are not conscious because then there is no higher order thought for us to express.

The absence of higher order thoughts prevents us from responding to our experiences verbally, but does not preclude non-verbal responses. This helps explain what happens in disorders such as blindsight, where brain lesions prevent subjects from consciously experiencing visual input in certain areas of their visual field. These patients cannot verbally report about that input in the automatic way that is characteristic with conscious experiences, but they can respond non-verbally to such input. These help confirm the higher-order-thought model, and make it promising for future research.

David Rosenthal is professor of philosophy, City University of New York.

a few extra knots

preserve of philosophers; it is part of science too. from a range of disciplines give their individual views

consider everything they say and what they say is part of your data. But you don't necessarily give it credence because the subject may not be right about their own subjective world," he explains. Dennett says that the third-person approach is leading to the discovery and prediction of new kinds of experience in people. "The fact that we can manipulate the very subjective experiences of people shows that we can understand it from a third-person point of view. So the claim that there is this ineffable residue that cannot be got at is shrinking fast."

For Dennett, scientific work on consciousness "is where the action really is". He says there has been a huge expansion in theoretical studies of consciousness because of advances in technology. This has made it possible to frame hypotheses that investigators simply did not have a language for a few years ago. For Dennett, conceptual advances in consciousness research go hand in hand with the availability of new technology.

At the leading edge of scientific research on consciousness is Francis Crick, father of molecular biology. Crick, based at the Salk Institute in San Diego, and his collaborator Christof Koch, have focused their efforts on the visual system. By the end of the century they hope they will be in a position to offer a neural explanation or "correlate" for visual awareness. This, they hope, will allow them to develop a coherent theory about consciousness which tackles the complexities of the subjective experience.

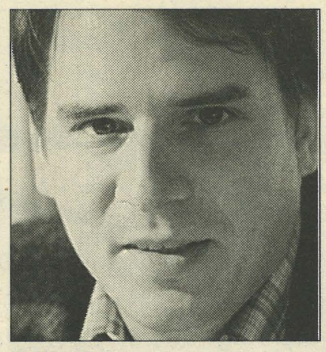
Crick says that the first-person experience should never be taken at face value in investigations of its nature. Visual psychology has shown that people deceive themselves enormously as to what they think is going on in their brains. For instance, people think they can see equally clearly in all directions whereas it is easy to show that they see most clearly in the centre of their gaze. The reason for thinking that we have all-round clear vision is because the eyes are busy moving about all the time, enabling the brain to fill in missing information in our field of view.

Another example of cerebral disinformation is human motivation: people say one thing when they are obviously motivated by something else. Crick says: "You cannot just take people's verbal reports about what they say on their own terms. You have to test them in a lot of circumstances. The subjective experience is what you might call real but it is not reliable in the sense that it is a true account of what is going on in your brain."

Crick and Koch believe that the sense of privacy in the conscious experience arises from the way the brain works. Crick explains that different areas of the brain can be thought of as being arranged in a pseudo-hierarchy with each "level" coding input-output information in a particular way. One level will perhaps code for movement, another for colour and so on. At each level the information is recoded so that what neurons are responding to is very different at each stage. And when it comes to the brain initiating speech or other kinds of motor output, the information has to be recoded again. Crick says: "This means that you cannot actually find out from the output what is going on inside the brain because the information is being recoded at all these stages. That is the explanation for why it is private. You can give an account of it but you cannot actually say what it is like. You cannot actually say what it looks like, or feels like to you in any way that explains it to other people. And that is what you would expect from our particular way of looking at the brain."

Considering the state of research on consciousness in general, Crick says it is not clear how rapid progress is going to be. He points out that even the most common assumption — that neuronal activity gives rise to conscious events — has not yet been proven, though he believes that it will be. He likens the lack of this evidence to the position people were in when they worried whether a vital force was needed to explain living things: "We now think that is unlikely because we understand the system so well. But in earlier times it was a reasonable assumption that you didn't have to have a vital force but you could not establish it. With consciousness we are in a state of ignorance. That is the main point. Most of the things we would like to know we do not know yet". While there has been an explosion in the number of people working on the problem, Crick says that this does not mean that the field is not in a "very confused state", with scientists disagreeing among themselves as much as the philosophers do.

He would like to see much more experimental work, particularly in neuroscience, being done and says that what is really lacking in the field is ideas. And it is experimental facts that he hopes will provide investigators with new ways of looking at the problem: "A lot of people would be loath to agree with that — they believe they can do it all in their heads," he says laughing.



Christof Koch

What is the relationship between our internal world of perceptions, thoughts and memories and the activity of nerve cells that must somehow be responsible for all this rich, subjective experience?

Francis Crick and I are pursuing our investigations on the basis of the tentative hypothesis that there are specific neurons in the brain whose activity mediates awareness that might include, of course, an event occurring in my own body. These neurons must expresses the fact that I am looking at the face of my daughter and that I "see" her blonde hair and "hear" her voice and that I can report on this awareness.

If — by some yet to be invented technical means — one could directly stimulate an appropriate set of such neurons in an awake human, the subject should have the experience associated with the features encoded by these nerve cells. It is quite likely that such neurons are distributed in a specific layer throughout parts of the cerebral cortex, that they have a specific shape and specific cellular properties and that they make specific connections with other neurons, most likely in the planning stages of cortex, that is in the frontal lobes. On-going experiments suggest that such neurons are absent from the earliest part of visual cortex.

But finding such neurons is only the beginning. Much will be learnt from studying areas of the brain that these neurons project to. And what about the crucial relationship between awareness and short-term memory? Is it not likely that these neurons will express the substrate of this type of memory? And what about the many types of mental diseases that affect awareness: can they be related to the specific loss of the awareness neuron? Finding the neurons that express the neural correlate of awareness is just another step, albeit a critical one, in understanding consciousness.

We now have the experimental tools in hand to attack these issues on an experimental basis. It is an open question as to whether all the puzzling aspects of consciousness will be explained once science has run its course or whether it will always defy any testable explanations.

Christof Koch is professor of computation and neural systems, California Institute of Technology.

KEY RESEARCH AREAS

1 Consciousness is in the brain. Approach pursued by hands-on neuroscientists, such as Susan Greenfield, who seek to understand consciousness by studying the actual material of which our brains are made. It is also the assumption of philosophers of mind, such as Patricia Churchland and John Searle, although their chief interests are philosophical rather than experimental.

2 Consciousness is in the mind. Here we have the experimental psychologists making a comeback. They have made some spectacular findings, such as Lawrence Weiskrantz's discovery of blindsight. He found that patients with a certain kind of brain damage cannot "see" certain things, but when asked to "guess" what is there, they get it right with inexplicable consistency. This seems to be a case of "awareness" without "consciousness".

3 Computers will explain consciousness. Computers have been one of the main spurs to a renewed interest in consciousness studies. Dan Dennett claims that a computerised robot of sufficient complexity would be both intelligent and conscious. He is even building one to prove it. Roger Penrose disagrees. He accepts that you might be able to build a machine that is conscious, but it would have to be more than a mere computer.

4 The allure of quantum mechanics. One of today's unsolved mysteries is the subatomic world of quantum mechanics. It seems reasonable to look here for a solution to consciousness. So it seems to researchers such as Stuart Hameroff. Others are sceptical, arguing you do not explain a mystery by an enigma.

5 Out of this world. Two groups are seeking the solution to the question of consciousness outside the material realm. One looks to the realm of meditation and mystical experience, the other sees in the mind-over-matter accounts of parapsy-chology a clue to the mind-over matter experience of, say my consciously raising my arm. Susan Blackmore opposes this approach. She can see no connection at all between the paranormal and consciousness.

6 The commonsense approach. The experts deride it as "folk psychology", but the common human experience is of being conscious, feeling in control of our lives and taking responsibility for them.

Compiled by Anthony Freeman

THES TUCSON II QUESTIONNAIRE

Join in the debate on consciousness without travelling to Tucson. *The THES* is conducting a survey to find out what people really think about consciousness: the evidence, the theories, and the prospects of a major breakthrough in our understanding. When we publish the results, they should make fascinating reading. Internet users can complete our questionnaire electronically at http://thesis.newsint.co.uk

1 Warm up with a thought experiment. How would the conscious experience of the following compare with your own?

a snail	a	b	c	d	e	f
a bat	a	b	c	d	e	f
a chimpanzee	a	b	c	d	e	f
an extremely intelligent computer	a	b	c	d	e	f
yourself with part of your brain replaced by silicon chips	a	b	c	d	e	f
your "brain program" copied to silicon	a	b	c	d	e	f
another human being	a	b	c	d	e	f
a zombie	a	b	c	d	e	f

(a) similar; (b) different in intensity; (c) different in quality; (d) different in intensity and quality; (e) absent; (f) the example is impossible.

2 Rate the following as sources of evidence about consciousness:

physics	a	b	c	d
neuroscience	a	b	c	d
philosophy	a	b	c	d
art	a	b	c	d
introspection	a	b	c	d
experimental psychology	a	b	c	d
other species	a	b	c	d
other cultures	a	b	c	d
computers	a	b	c	d
altered states	a	b	c	d
paranormal events	a	b	c	d

(a) likely to mislead; (b) irrelevant; (c) useful; (d) essential.

3 Which kind of theory of consciousness would you find most plausible?___

(a) it is a computational process; the "hardware" does not matter;
(b) consciousness can be traced to fundamental physics;
(c) consciousness can be traced to something fundamental, new and possibly informational;
(d) it is something to do with neurons and networks;
(e) it is something to do with the whole body and the world it inhabits;
(f) each of these levels adds something unique to make consciousness possible;
(g) there is a realm of mind and a realm of matter; consciousness is a mind thing.

4 What are the prospects for a solution of the "hard problem" (how physical processes in the brain give rise to subjective experience)?___

(a) it cannot be solved in principle;
(b) it cannot be solved in practice;
(c) it might be solved some day;
(d) it will definitely be solved;
(e) it is nonexistent or already solved.

5 Name

.....

6 Address

.....

.....

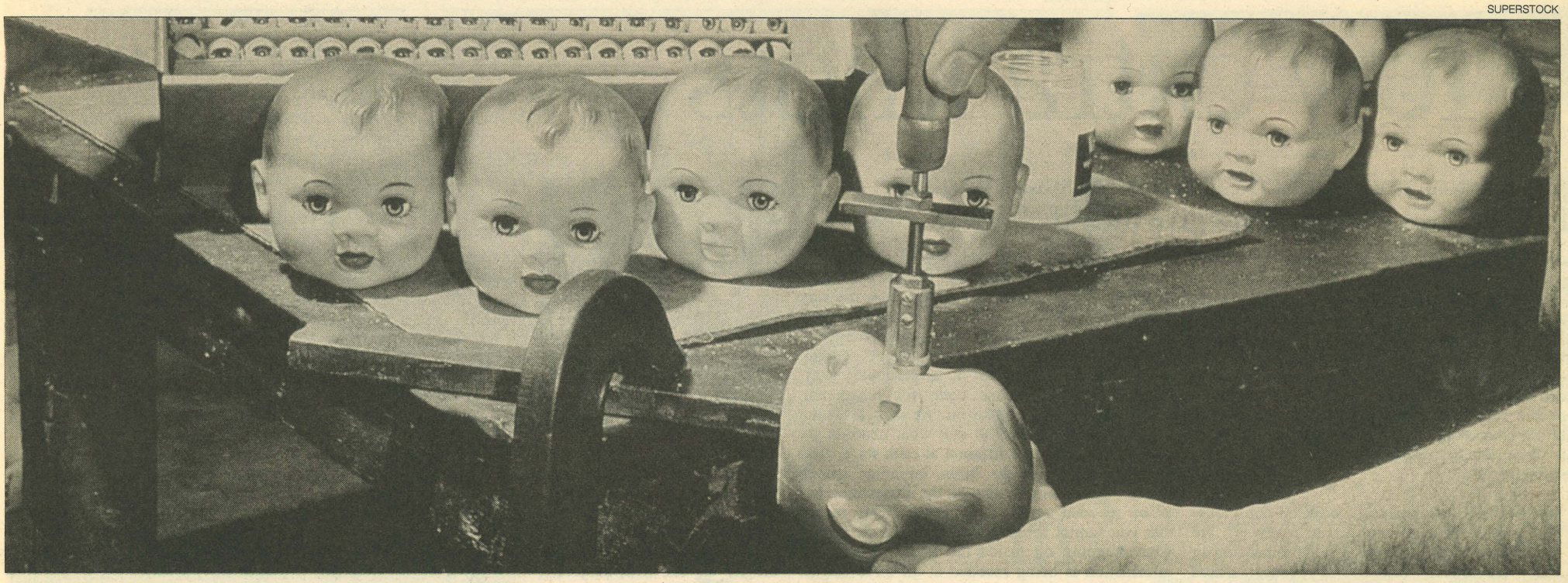
.....

.....

.....

Send your completed questionnaire to Lynne Williams, *The Times Higher Education Supplement*, Admiral House, 66-68 East Smithfield, London E1 9XY, United Kingdom. Fax +44 171 782 3300.

SUPERSTOCK



Can machines ever be conscious?

Igor Aleksander (left) believes they can, **Jaron Lanier** disagrees

Can machines be conscious? "I should co-co! You should see my new video recorder. Set it up to record *Neighbours* and it gives you the early evening news instead. It's got a will of its own."

People are only too ready to assign consciousness to inanimate objects; ancient Egyptians ascribed consciousness to the statue of Memnon because it emitted sounds as air expanded within it, warmed by the rays of the sun. However, were someone at a party to reveal that he was made of silicon chips, a perfectly understandable reaction might be "Good heavens! I've been talking to an unconscious zombie all evening". Curiously, people ascribe consciousness to the daftest selection of objects and then argue that sophisticated information processing machines are definitely not conscious. The judgement as to whether a manufactured object is conscious need not be a property of the machine, it is in the gift of the beholder. So, were it possible to make artefacts which some people agree are conscious, it would be impossible to do this in a universally agreed way.

Philosophers call this the "third person" problem; it is impossible, some argue, to tell whether anything outside ourselves is conscious or not. It is impossible (argues the philosopher Thomas Nagel) to know what it is like to be a bat. You may work out every last detail of the working of a bat's brain but this will not tell you what it is like to be that bat. So the only organism of whose consciousness I can be absolutely sure is me. This is the "first person" hallmark of consciousness. So, philosophically minded people will perfectly properly argue that science and the engineering of machines, being about objects outside ourselves, cannot cope with anything but third-person problems and therefore cannot explain what consciousness is.

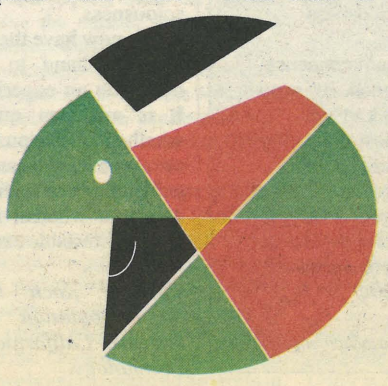
My argument is that the engineer can conjure up a sneaky finesse of this problem. She can ask the question, "What properties does an organism, artificial or real, need to have to decide that it itself is conscious?" If it can be argued that an object does not have the machinery to be able to decide it is conscious, then any argument that it might be conscious is not based on any principle and is just arbitrary attribution.

To work with neural machines (one of which — called Magnus — is used by me and my co-workers) is to ask precisely the above question. It is not about machines that everyone will agree are conscious. The consciousness of anything we produce could be refuted simply though "third person" doubts. But we are defining the nature of a minimum of machinery required for a machine to be able to operate in a "first-person" way. The critics of this approach, particularly those who are critical of artificial intelligence as a programmer's act of puppeteering, will say: "but any old machine can be given a voice synthesiser that regularly says things like 'I think, therefore I am.'"

Creation of animated computer puppets is precisely what we are *not* doing. Magnus is generating explanations of "first person" mechanisms and particularly how these may be derived from the operation of neural networks. Neural networks are not used because they mimic the brain, but because they have the qualities to learn and generalise in what seems to be a fundamental fashion. So no wonder that the brain has such things too. Typical are questions as to how an artificial neural net might store a vivid, recallable impression of sights, sounds, smells, feelings and tastes; how the same net, given ways of acting on its environment, builds a representation of what, in the world around it, it can achieve and modify; how it then goes on to absorb language with which it can express its own internal activity (thoughts?) to a human interlocutor. Theory has it that such an organism could build up emotions from instincts and even have a true "will of its own". Indeed, the philosophers' concept of "qualia" can be shown not to be outside the representational power of a neural net.

Some thinkers do not like what we do and argue that "mere computer simulation" cannot capture the "sentience" of a living being. But what we do is not "mere simulation", it is an inquiry into the nature of sentient organisms that aims at an explanation of the first person and merely uses a computer as a useful tool with which to develop and demonstrate ideas about mechanisms. Others argue that no matter what mechanism we may discover, this could well be necessary but not sufficient to explain consciousness. What is missing is something that is not available to science. There is a gap, they argue, between neural mechanisms and consciousness.

I disagree. Their viewpoint is no more than a belief. Phlogiston and spatial ether were such beliefs until explanations were developed which made them unnecessary, redundant. Some philosophers may accurately call me an eliminativist, but it is up to them to show that neural mechanisms are insufficient for any reason other than that they believe this to be the case.



Igor Aleksander is professor of neural systems engineering, Imperial College, London.

It is collective self-flattery for the computer science community to argue that computers can be conscious. I will argue that they cannot. Arguments about machine intelligence hinge on questions of epistemology, our ways of knowing what we know. The most basic argument of this kind is the Turing test. Alan Turing proposed that if a computer was programmed in such a way that it could fool a human observer into believing that it was conscious, then it would be sentimental foolishness to suggest that it was not conscious — like claiming the earth was at the centre of the universe; a desperate attempt to hold onto our uniqueness.

I claim that there are different ways of knowing things. Consciousness is the thing we share that we do not share objectively. We experience it subjectively, but that does not mean it does not exist.

How could we decide whether machines might also experience consciousness? In Turing's set-up, it is impossible to tell whether the computer has become more human-like, or whether the human has become more computer-like. All we are able to measure is their similarity. This ambiguity makes artificial intelligence an idea that is not only groundless, but damaging. If you observe humans using computer programs that are designated to be "smart", you will see them make themselves stupid in order to make the programs work.

What starts as an epistemological argument quickly turns into a practical design argument. In the Turing test, we cannot tell whether people are making themselves stupid in order to make computers seem smart. Therefore the idea of machine intelligence makes it harder to design good machines. When users treat a computer program as a dumb tool, they are more likely to criticise a program that is not easy to use. When users grant autonomy to a program, they are more likely to defer to it, and blame themselves. This interrupts the feedback that leads to improvements in design. The only measurable difference between a smart program and a dumb tool is in the psychology of the human user.

This argument suggests that it is better for us to believe that computers cannot be conscious. But what if they actually are? This is a different kind of question, a question of ontology. I argue that computers are not conscious because they cannot recognise each other. If we sent a computer in a spaceship to an alien planet and asked for a definitive analysis of whether there were computers present, the computer would not be able to answer. There are theoretical limits on one program's ability to fully analyse another that make this so. People can recognise and use computers, so people are not in the same ontological category as computers.

This is just another way of saying that without consciousness, the world as we know it through our science need not be made of gross objects at all, only fundamental particles. For instance, one has to be able to distinguish cars from air in order to measure "traffic". Our most accurately confirmed scientific hypotheses, those of fundamental physics, do not, however, acknowledge cars or other gross objects.

It is easy to claim that the state of a person's brain is what notices cars or computers, but that avoids the question of how the brain comes to matter as a unit in the first place. If consciousness is associated with a brain, why is it not also associated with a momentary correlation between a brain and the arrangement of noodles on a plate of pasta being eaten by the owner of the brain? Even brains exist only by virtue of conscious acknowledgment. The alternative idea would be that the right kind of complex process gives rise to consciousness. In that case there would be huge swarms of slightly different consciousness around each person, corresponding to every combination of their brain, or sections of it, with other objects in the universe.

A world without consciousness would be a world of elementary particles at play. They would be the same particles in the same positions and relationships as in our world, but no one would notice them as members of objects like brains. I am not claiming there is something outside my brain that contains the content of my experience. I can accept that the content of my subjective experience might be held in my neurons, and still claim that it is experience itself that makes those neurons exist as effective units.

The first argument presented above, about the Turing test, turns out to have practical relevance because it influences our ability to design better user interfaces. And I think the second ontological argument does too — computers have come to play such a central role in our culture that our way of thinking about them affects our ways of thinking about each other. The tendency to think of computation as the most fundamental metaphor for experience and action leads inevitably to sloppy computer metaphors in politics, economics, psychology, and many other areas. I hope that if we acknowledge just how strange and wonderful it is that we are conscious, that wonder will translate into less bland and nerdy metaphors to guide us in those areas.

Jaron Lanier is visiting scholar in the computer science department, Columbia University, and chief scientist at Talisman Dynamics.

Do you believe in psychic phenomena? Are they likely to be able to explain consciousness?

Jessica Utts and **Brian D. Josephson** (left) are open to the idea, while **Susan Blackmore** is sceptical

Those who recognise that significant discoveries in science are often prompted by observations that do not fit expectations will find a stimulating challenge in accumulating evidence that it is possible to elicit psychic functioning in experiments with ordinary volunteers acting as subjects. Even more convincing results occur with specially selected subjects.

In one type of experiment, a "target" photograph or video segment is chosen randomly from a set of four possibilities. A "sender" attempts to transmit it mentally and a "receiver" is then asked to provide an account, either verbally or in writing, of what she imagines it might be. She is then shown the four possibilities, and selects the one she thinks best matches her perception. By chance alone, a correct match is expected on average one time in four, whereas the experiments typically show the considerably higher success rate of around one in three.

The recent declassification of the US government's psychical research programme (experiments on "remote viewing", similar to the type just described except that they used independent judges to assess the matches rather than having the subjects judge themselves) has permitted a comparison to be made of the results of this programme with those described in the open literature. Despite the different judging procedure, similar success rates were found. In addition, many of the governmental experiments used gifted subjects. The success rate was then even higher, typically over 40 per cent. The few experiments in the open literature that used gifted subjects found similar success rates.

In the past, critics have attempted to discredit positive results in psychical research on grounds of lack of repeatability. But, as any statistician knows, even where an influence exists, an isolated experiment with an insufficient number of trials may not demonstrate a statistically significant effect. Accordingly, without a more sophisticated analysis, "failure to reproduce an effect" does not demonstrate its absence. Suppose psychic abilities, in line with the results already achieved, increase the chances of a successful match between real and imagined images from one quarter to one third. Then (according to statistical theories), an experiment with 30 trials, which has been typical of these experiments, would have less than a 17 per cent chance of achieving a result of statistical significance. More recent larger experiments still utilise only about 100 trials, and have only about a 57 per cent chance of achieving statistical significance.

Detailed analysis of the complete collection of experiments on this type of phenomenon shows that what holds, despite changes in equipment, experimenter, subjects, judges, targets and laboratories, is far greater consistency with the one-in-three success rate already mentioned than with the one-in-four chance expectation rate. Such consistency is the hallmark of a genuine effect, and this, together with the very low probability of the overall success rate observed occurring by chance, argues strongly for the phenomena being real and not artifactual.

Re-examination of other types of psychical investigations reveals that they too achieved replicable effects, which went largely unappreciated because of a poor understanding of statistics. For instance, an analysis of experiments in precognitive card guessing and related "forced-choice" experiments, published by Honorton and Ferrari in the *Journal of Parapsychology*, found that gifted subjects were able to achieve consistently about a 27 per cent success rate when 25 per cent was expected by chance. Similar US government experiments have achieved the same 27 per cent success rate over thousands of trials. If chance alone were the explanation for these results, it would be truly remarkable to achieve a 27 per cent success rate over thousands of trials, and it would be even more remarkable to see identical results in the government work.

Strong statistical results are of course meaningless if experiments are not properly conducted. Debunkers of parapsychology are fond of showcasing the very few experiments that have been found to have serious problems. But that ignores the fact that the vast majority of experiments were done using excellent protocols. For the past decade the US government experiments were overseen by a high-level scientific committee, consisting of respected academics from a variety of disciplines, all of whom were required to approve the protocols in advance. There have been no explanations forthcoming that allow an honest observer to dismiss the growing collection of consistent results.

What are the implications for science of the fact that psychic functioning appears to be a real effect? These phenomena seem mysterious, but no more mysterious perhaps than strange phenomena of the past which science has now happily incorporated within its scope. What ideas might be relevant in the context of extending science to take these phenomena into account? Two such concepts are those of the observer and nonlocality. The observer forces his way into modern science because the equations of science, if taken literally, imply a universe that is constantly splitting into separate branches, only one of which corresponds to our perceived reality. A process of "decoherence" has been invoked to stop two branches interfering with each other, but this still does not answer the question why our experience is of one particular branch, and not another. Perhaps, despite the unpopularity of the idea, the experiencers of the idea are also the selectors.

This idea perhaps makes sense in the light of theories that presuppose that quantum theory is not the ultimate theory of nature, but involves the manifestations of a deeper "subquantum domain". In just the same way that a surf rider can make use of random waves to travel effortlessly along, a psychic may be able to direct random energy at the subquantum level for her own purposes. Some accounts of the subquantum level involve action at a distance, which fits in well with some purported psychic abilities.

These proposals are extremely speculative. What needs to be done, in any event, is to integrate mental phenomena more thoroughly into the framework of science. The research of Lawrence LeShan, where interviews with psychics disclosed that they were aware of a "hierarchy of meaningful interconnections", perhaps provides a hint of what might be involved.

Science has a poor handle on ideas such as meaningful interconnections since they are alien to its usual ways of thinking. Perhaps it will need to overcome its abhorrence of such concepts in order to arrive at the truth.

Jessica Utts is professor of statistics, University of California, Davis, and was one of two experts commissioned by the CIA to review the US government's psychic research. Brian D. Josephson, Nobel laureate, is professor of physics, University of Cambridge.

DAVID BROMLEY

If you took a time machine and travelled back anywhere in parapsychology's 60-year history you might hear parapsychologists say something like this: "The old experiments had problems — but we've really found the repeatable experiment this time." I first heard this in the early 1970s. After a dramatic out-of-body experience I had found myself in states of consciousness that were completely ignored by ordinary psychology. It seemed logical then, though it certainly does not now, to turn to parapsychology. When I decided to become a parapsychologist I had no idea it would mean 20 years of failing to find the paranormal.

At that time card-guessing experiments were still the norm. Samuel Soal's famous telepathy experiments at Queen Mary College, London, providing odds against chance of millions to one, had not yet been exposed as a fraud. Results with children in classrooms seemed promising, as did experiments with telepathy during dreams.

Until recently the latest "real thing" has been the *ganzfeld*. Subjects in this experiment lie comfortably, listening to white noise or sea-shore sounds through headphones, and wear half ping-pong balls over their eyes seeing nothing but a uniform white or pink field (the *ganzfeld*). Meanwhile, a sender in a distant room views a picture or video clip. After half an hour or so the subject is shown four such pictures or videos and is asked to choose which was the target. Several researchers have claimed positive results, and meta-analyses have combined the results of many experiments to show that the results are consistent, do not depend on any one experimenter, and reveal regular features of extrasensory perception.

The *ganzfeld* reached scientific respectability in 1994 when Cornell psychologist Daryl Bem and parapsychologist Charles Honorton published a report in a prestigious journal, *Psychological Bulletin*. They reported impressive new results with a fully automated *ganzfeld* procedure, claiming to have demonstrated a repeatable experiment. So had they?

My own conclusion is biased by my personal experience. I tried my first *ganzfeld* experiment in 1978, when the procedure was new. Failing to get results myself I went to visit the laboratory in Cambridge where some of the best results were being obtained. What I found had a profound effect on my confidence in the whole field. The experiments, which looked so beautifully designed in print, were easily open to fraud or error. Eventually the experimenters and I all published our different views of the affair, and the main experimenter left the field. I turned to other experiments.

This depressing incident is only still relevant because the Cambridge data is all there in the Bem and Honorton review. Indeed, out of 28 studies included, nine came from the Cambridge lab, more than from any other laboratory. Yet not a word of doubt is expressed, no references are given, and a reader could not guess there was such controversy.

Of course the new auto-*ganzfeld* results are even better. Why should I doubt them because of events in the past? The problem is that my personal experience conflicts with the succeses I read about in the literature and I cannot ignore either side. The only honest reaction is to say "I don't know".

Now that the CIA has released details of more than 20 years of research into remote viewing the spotlight has left the *ganzfeld*. "Oh yes, the old *ganzfeld* experiments had problems", we might soon hear "but we've really found the repeatable experiment this time". But what if they have? What if my doubt is misplaced and there really is extrasensory perception after all? What would this tell us about consciousness?

The popular view seems to be something like this — if ESP exists it proves that mental phenomena are nonlocal, or independent of space and time. If psychokinesis exists, it proves that mind can reach out beyond the brain to affect things at a distance. If you equate mind with consciousness — hey presto — ESP and PK prove the power of consciousness.

It is a desire for this "power of consciousness" that fuels much enthusiasm for the paranormal. Parapsychologists have often been accused of wanting to prove the existence of the soul, and denied it, so I will instead accuse them of wanting to prove the power of consciousness. Will they succeed?

First they need to make their case that any effects they find really involve consciousness. For example, recent PK experiments apparently show "the effect of consciousness" on random-number generators, computers or dice. Yet what they have really shown is correlations between instructions given to subjects and the physical system being tested. The really interesting questions concerning consciousness are about subjectivity. There are no controls in the PK experiments to show that subjective experience is involved in any way.

As our understanding of conscious experience progresses, this desire to find the "power of consciousness" sets parapsychology ever more against the rest of science (which of course is part of its appeal). The more we look into the workings of the brain the less it looks like a machine run by a conscious self. There is no place inside the brain where consciousness resides, where mental images are "viewed" or where instructions are "issued". There is just massive parallel throughput and no centre. There are even a few crucial experiments suggesting that conscious experience takes some time to build up and is much too slow to be responsible for making things happen. Indeed the brain seems to be a machine that runs itself very well and produces an illusion that there is someone in charge. This illusion is just what meditators and spiritual practitioners have been saying for millennia; that our ordinary view of ourselves, as conscious, active agents experiencing a real world, is wrong — an illusion. Now science seems to be coming to the same conclusion.

Parapsychology is going the other way. It is trying to prove that consciousness really does have power; that our minds really can reach out and "do" things, not only within our own bodies but beyond. Odd, then, that so many people think of parapsychology as more "spiritual" than conventional science. I think it could be the other way around.

I look forward to the kind of psychology that can bring together the spiritual insights with the scientific ones — that can reveal what kind of illusion we live in and how it comes about, and perhaps even help us to see our way out of the illusion. This would indeed be progress in understanding consciousness, and in being conscious in a different way. And as far as this hope is concerned parapsychology is going nowhere. This is why my answers to the two questions are "probably not", and definitely "no".

Susan Blackmore is senior lecturer in the department of health and community studies, University of the West of England.

Does consciousness emerge from quantum processes?

Stuart Hameroff and **Roger Penrose** (left) think it does, while **Patricia Churchland** is less convinced

How can we comprehend the nature of our conscious experience? This question provokes four types of explanation. "Reductive materialists" believe that conscious experience simply emerges from computer-like excitations among the brain's neurons. "Dualists" view consciousness as separate from the brain, but able to influence brain activities. "Idealists" argue that consciousness is primary and itself creates reality: consciousness is all there is. "Panpsychists" say that conscious experience is intrinsic to physical reality, that a "protoconsciousness" (a "funda-mentality") is present even in inanimate structures.

Consider this fourth view. Could the raw components of conscious experience actually be "built-in" to the universe? Philosopher Alfred North Whitehead proposed that at a deeper level than atoms or electrons are fundamental units, which Whitehead termed "occasions of experience". Some modern thinkers argue that what makes up the universe is fundamental information with experiential properties. Perhaps neurobiological systems somehow access and organise precursors of conscious experience that are embedded in the physics of reality.

Present-day understanding of physical reality rests upon "space-time geometry", as described by both Einstein's general relativity and quantum theories. General relativity shows that our perceived reality of three spatial dimensions "moving through time" is, more appropriately, a four-dimensional space-time continuum. The presence of a physical object induces a curvature of this underlying space-time. Whereas large objects (eg planets, stars) produce measurable space-time curvatures, those produced by small ones (eg atoms, proteins) are tiny.

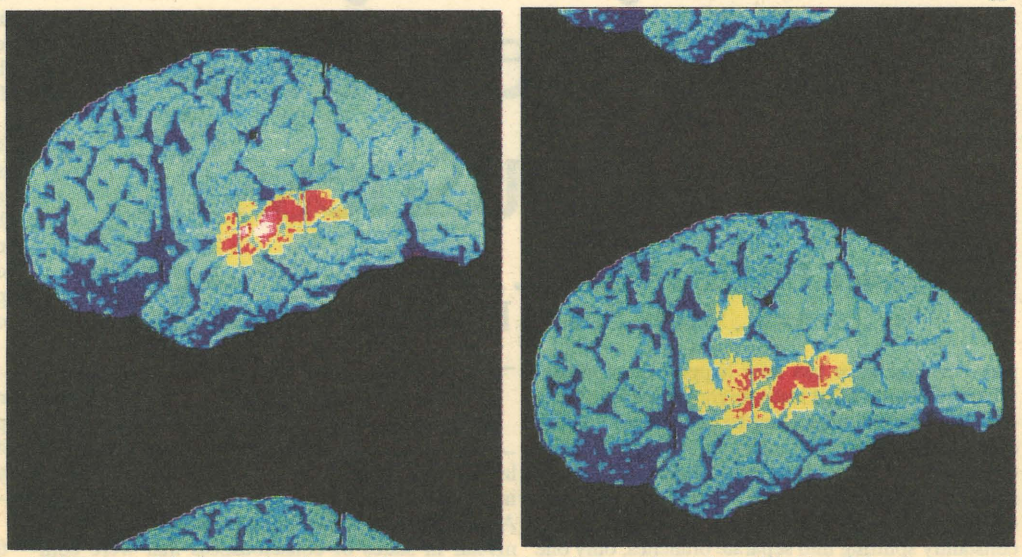
More relevant at this small scale is quantum theory, which has shown that individual particles (or even large collections of them) can coexist in a "superposition" of separate locations at once. As strange as this seems, simultaneous existence of a quantum object in two locations at once has been experimentally verified repeatedly. The major unresolved issue is that quantum superpositions seem to perish when the systems get "too large", and mysteriously "collapse" into definite locations. Because very small superposed systems do not spontaneously collapse in this way, conventional theory holds that quantum systems remain in superposition until consciously observed. Accordingly, the mythical Schrödinger's cat would remain both dead and alive in its closed box.

However, many physicists now believe that at some point between very small quantum-scale systems and large cat-size systems, an "objective" factor disturbs the superposition and causes an actual physical collapse (or reduction) to definite states and locations. This putative process is called "objective reduction" (OR). Moreover, some scientists believe that the measure of an object's "largeness" which elicits such "OR" is the degree of overall curvature it induces in space-time.

What about the space-time curvature induced by an object in superposition? An object in quantum superposition existing simultaneously in separated locations will evolve separate and distinct space-time curvatures. This leads to a "blistering" of space-time. If such blisters were to continue to enlarge indefinitely, the entire space-time geometry would separate (leading to "multiple universes"). According to the principles of OR, however, collapse abruptly occurs when the space-time separation reaches a critical degree. The instantaneous OR event selects a particular mass distribution and a particular configuration of space-time geometry. If "funda-mentality" is indeed embedded in the universe, here is a natural place for it: a self-organising OR process could select the individual space-time geometry of experience.

Could self-organising OR events of this nature be occurring in our brains? Why is neurobiology better than rocks or tables at selecting fundamental experience? A requirement for brain OR events would seem to be structures in nerve cells which could support quantum superposition of sufficient "largeness" to elicit OR at appropriate time-scales. Such structures should be highly prevalent, functionally important, coupled to quantum-level events and have crystal-like order.

SLP



A PET scan of the brain showing which areas are active when words are heard (left) and then interpreted

Most important, they should be capable of information processing and have the ability to be isolated from external interaction. Membranes, membrane proteins, synapses, cell water, DNA, clathrins, myelin, centrioles and other neurobiological structures have been suggested, but in our view "cytoskeletal microtubules" are best suited.

Microtubules are crystal-like protein cylinders that perform varieties of cellular chores within nerve cells including forming and regulating synaptic connections. The individual protein subunits ("tubulins") which make up the microtubules can switch between different configurations, known as "conformations", governed by quantum-level events. Conceded to be the cell's structural support, microtubules, according to accumulating evidence can also process information by means of cooperative interaction of their tubulin subunits.

In a series of recent articles, we have proposed that microtubules process information while in quantum-coherent superposition. In this form of "quantum computing", multiple computations may be performed simultaneously, in parallel. In our view, the quantum computing phase in microtubules corresponds to preconscious processing lasting up to one second (and involves microtubules arrayed in thousands of neurons). The climactic and instantaneous ORs are conscious events. Sequences of such "occasions of experience" create a flow of time, giving rise to a stream of conscious thought.

Although dependent upon some unproven assumptions, our model has significant advantages. It is specific, and attempts to deal directly with the nature of experience. Moreover, like consciousness, quantum coherence has an essential global unity.

Consciousness undoubtedly has an important place in the universe. We take a new scientific approach towards understanding how conscious experience might deeply integrate with the workings of the physical universe.

Stuart Hameroff, MD, is a practising clinical anesthetist and professor at the University of Arizona. Sir Roger Penrose is Rouse Ball professor of mathematics, University of Oxford.

When I am asleep, I am unaware of the smell of cinnamon in the air. Distracted by a video game, I am unaware of the movement of my tongue. If a stroke renders me paralysed, I may be unaware of my paralysis. Each of these cases presents an opportunity to the neuroscientist trying to understand the nature of conscious experience. So what are the differences in the brain when I am aware of a stimulus, and when I am unaware of it?

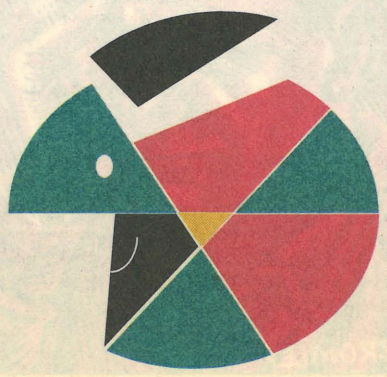
Discovering the relevant differences marks an important starting point in discovering the mechanisms of conscious experience. To advance beyond the starting point will require a detailed understanding of the brain's anatomy and physiology.

At this early stage in the project, what is meant by "consciousness" is best specified by example, using well-attested instances. More contentious examples, such as whether frogs are visually aware, can be sorted out once cognitive neuroscience is a little more solid. In other words, we start with common sense, and see where the science leads.

Is the above approach hogwash? Possibly. Dualists, who believe that there is a nonphysical soul in addition to the physical brain, will certainly say so. Colin McGinn and Jerry Fodor, for example, have declared that the brain *is* more complicated than it is smart and that consciousness will forever be a mystery to us. Given how much remains to be discovered about the brain, not to mention the unpredictability of technological innovation and the doors thereby opened, it is surprising that they rely on poverty of the imagination.

Now suppose we *do* find some phenomenon really mysterious. This is a psychological fact about *us* — not a metaphysical fact about the nature of the world. It is a fact about what we do and do not know, about where science has and has not reached.

For Roger Penrose, the key to consciousness lies in quantum events in tiny protein structures — microtubules — within neurons. Why *there,* and why quantum mechanical properties? Because the nature of mathematical understanding, Penrose believes, transcends the kind of computation that could conceivably by done by neurons and networks. As a demonstration of neuronal inadequacy, Penrose cites the Gödel incompleteness result, which concerns limitations of provability in axioms systems for arithmetic. What is needed, according to Penrose, are unique operations at the quantum level. Quantum gravity, were it to exist, would do the trick. Granting that no theory of quantum gravity exists, Penrose and his colleague Stuart Hameroff argue that microtubules are about the right size to support the envisioned quantum events, and have the right sort of sensitivity to anesthetics to suggest they do sustain consciousness.



The details of the Penrose-Hameroff theory are highly technical. Before investing time in mastering the details, most people want a measure of the theory's "figures of merit", as an engineer might put it. Specifically, is there any hard evidence in support of the theory, is the theory testable, and if true, would the theory give a clear and cogent explanation of what it is supposed to explain?

The figures of merit are not encouraging. First, mathematicians generally disagree with Penrose on what the Gödel result implies for brain function. Additionally, the link between conscious experiences such as smelling cinnamon and the Gödel result is obscure at best.

Now, is there any significant evidential link between microtubules and awareness? Hameroff believes microtubules are affected by hydrophobic anesthetics in such a way as to cause loss of consciousness. But there is no evidence that loss of consciousness under anesthesia depends upon the envisaged changes in microtubules, and only indirect evidence that anesthetics do in fact (as opposed to "could conceivably") have any effect on microtubules. On the other hand, evidence points to proteins in the neuron membrane as the principal locus of action of hydrophobic anesthetics.

Is there any hard evidence that quantum coherence happens in microtubules? *Only that it might.* Surely the presence of cytoplasmic ions in the microtubule pore would disrupt these effects? *They might not.* Surely the effects of quantum coherence would be swamped by the millivolt signalling activity in the neuronal membrane? *They might not be.* Can the existence of quantum coherence in microtubules be tested experimentally? *For technical reasons, experiments on microtubules are performed in a dish, rather than in the animal.* If tests under these conditions failed to show quantum coherence, would that be significant? *No, because microtubules might behave differently in the animal, where we cannot test for these effects.* Does any of this, supposing it to be true, help us explain such things as recall of past events, filling in of the blindspot, hallucinations and attentional effects on sensory awareness? *Somehow, it might.*

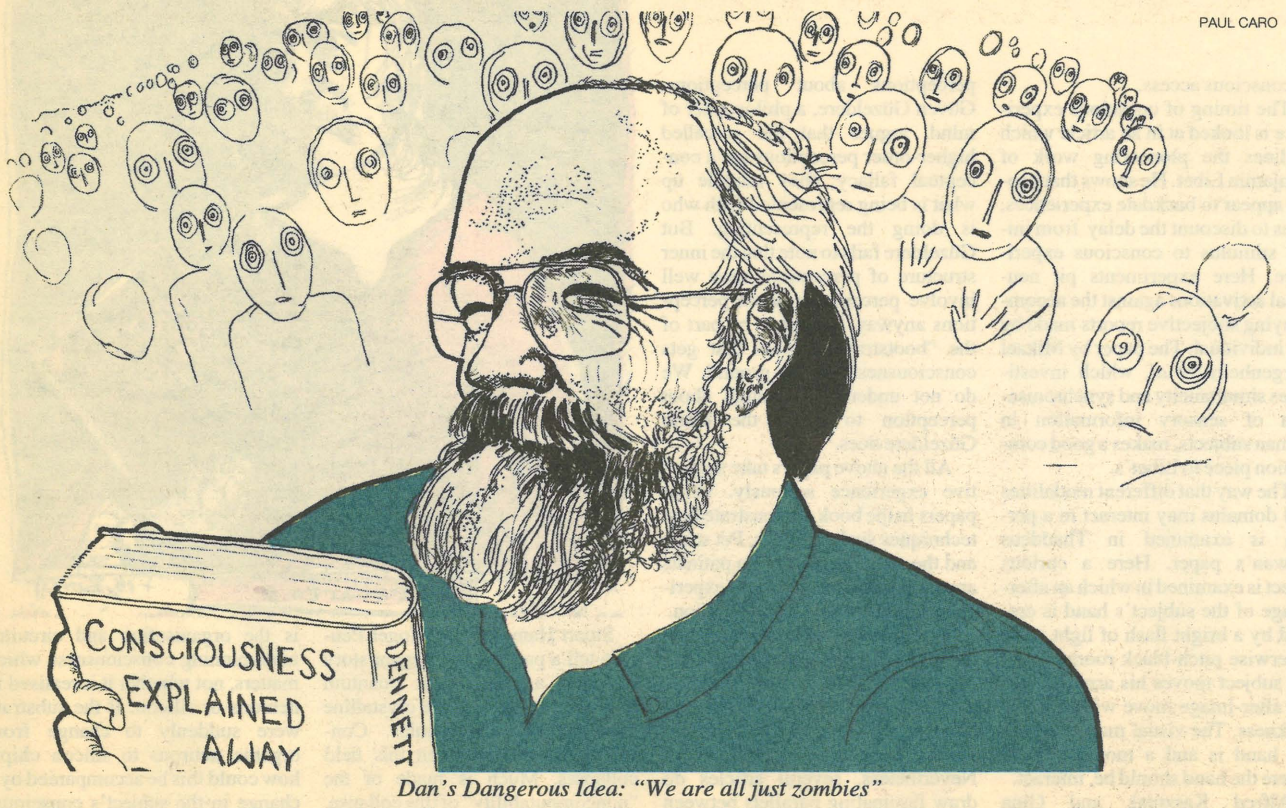
The want of directly relevant data is frustrating enough, but the explanatory vacuum is catastrophic. Pixie dust in the synapses is about as explanatorily powerful as quantum coherence in the microtubules. The theory needs work.

What I share with Penrose — and essentially all neuroscientists — is perplexity. So far, cognitive neuroscience really does not have its hands on a conceptual framework for explaining how experience is based in brain operations. Significantly, awareness is by no means the lone enigma. For virtually *no* higher function do we have a theoretical framework adequate to yield genuine, full-blooded explanations — of how, say, one recalls the punch line of a joke Uncle Bart told last month, or how a skill like typing becomes automatised or even how eye movements are controlled in reading. For none of these cases do we have a theory that genuinely *explains* how the effect is achieved.

To be sure, most new ideas are bound to go the way of the three-legged trout. But the climate should not be so harsh as instantly to snuff out any contender that looks outlandish. For this reason alone, I applaud the boldness of Penrose and Hameroff. Having looked closely at the details of their proposal, however, I am inclined to pin my explanatory hopes on cognitive neuroscience.

Patricia Churchland is director of the Experimental Philosophy Lab, University of California, San Diego.

Daniel Dennett's stimulating *Consciousness Explained* failed to live up to its title. Now another philosopher, David Chalmers, offers a new theory. **Colin McGinn** assesses it philosophically and **Roger Penrose** scientifically



PAUL CARO

Dan's Dangerous Idea: "We are all just zombies"

Wise incomprehension

BY COLIN MCGINN

THE CONSCIOUS MIND: IN SEARCH OF A FUNDAMENTAL THEORY
BY DAVID CHALMERS
Oxford University Press
392pp, $29.95
ISBN 0 19 510553 2

It is very hard to devise a theory of consciousness that is not open to decisive objection. This is not because consciousness is so amorphously ill-defined that anything goes and we find it impossible to choose among a plethora of options. Rather, no matter what theory we come up with, it always seems to run into some shattering difficulty. The problem of consciousness is like a chess game in which a series of forced moves always ends in checkmate, more or less humiliating. Sometimes it seems that the best we can hope for is some teetering *ad hoc* contrivance that just manages to evade outright refutation — for the moment at least. Philosophy is like that, we know; but with consciousness the constraints are especially tight.

David Chalmers's book is an attempt to develop a theory that escapes knockdown refutation, while tolerating some counterintuitive and uncomfortable features. The book is very well argued, thorough, sophisticated, honest, stimulating — and almost plausible. It is certainly one of the best discussions of consciousness in existence, both as an advanced text and as an introduction to the issues. One feels that Chalmers has done about as good a job as could be done on this most intractable of problems. That said, I do not think the position he defends ultimately works, and for reasons that are not surprising. Still, there is much to be gained by following his argument: checkmate, yet again, but an impressive game nonetheless.

The book has two central theses, one negative, the other positive. The negative thesis is that materialism is false, because the mental is not logically supervenient on the physical. The mental is not explained and necessitated by the physical in the way that the observable macroproperties of water are explained and necessitated by the molecular structure of water. Since facts about consciousness are not entailed by physical facts, the former are something over and above the latter. This is argued to follow from the conceivability of zombies — entities physically just like us but without any consciousness: since these are logically possible, the physical facts alone cannot conceptually guarantee the presence of a conscious life. We cannot then come to know anything about conscious experience itself just from knowing all the physical truths of the universe; nor, *a fortiori*, is it possible to analyse experience in physical or functional terms. Experience is irreducible. It follows that dualism of some form must be true.

The positive thesis is that this dualism consists in fundamental laws that connect physical and mental properties by mere natural (not logical) necessity. We cannot reductively explain experience in physical or functional terms, but we can suppose there to be a contingent empirical law-like connection between them. This is nomological dualism instead of the rejected reductive monism. The physical does indeed "give rise" to the phenomenal, but it does so only with the force of natural necessity. Experience is thus a basic feature of the universe, like space and time, tacked on (as it were) to the swarms of particles that constitute matter.

In addition to these two main theses Chalmers speculates that the notion of *information* might provide some sort of link between the mental and the physical. Since the concept of information he employs is correlative with the notion of causation (the Shannon-Weaver concept of selection among possibilities), it turns out that experience is ubiquitous in the world — which leads Chalmers to endorse a version of panpsychism. Thermometers can now boast consciousness of some primitive form, a result Chalmers declares himself willing to live with. He also ingeniously defends a version of functionalism that makes experiences lawfully correlated with (but not reducible to) computational-functional properties. The argument here turns on the implausibility of dissociating qualia from the subject's first-person access to them, as would have to be so if experience could float free of a subject's cognitive processing.

There are two large problems with the theory as presented. The first, which Chalmers fully acknowledges, is that epiphenomenalism about experiences is entailed. Since my zombie and me share our physical and functional constitution, nothing in our behaviour differs, so that the doings of *both* of us can be explained without ascribing conscious states to *either* of us — yet I have them and he does not. In particular, we make the same judgements — including, for example, "I am conscious and currently having a red experience" — despite the vast difference in respect of conscious experience. But now it follows that my utterance of this is not explained by what makes the judgement true, since my zombie's utterance cannot be so explained — it being false in his case. My experience thus turns out to be epiphenomenal with respect to my self-ascriptions of experience. Chalmers himself spells out this consequence and tries his best to draw its sting; but he is clear that it would be better if it could be avoided, and he does not succeed in removing the attendant air of paradox. What needs to be noted is that it is the denial of logical supervenience that leads directly to epiphenomenalism; so we need to be very sure that this denial is compulsory.

The second problem, which he nowhere confronts, is that just as the alleged conceptual contingency of the link between the physical and the mental leads to the logical possibility of zombies, so also does it lead to the logical possiblity of disembodied consciousness. For if the link is merely that of natural necessity, then there are possible worlds in which the laws are abrogated — which means that the correlated properties could be instantiated independently of each other. There are pure spirit worlds as well as zombie worlds! I do not know whether this consequence would alarm Chalmers, but I suspect it would — and rightly so. How would such disembodied experiences be connected to the rest of nature? What might their causal powers depend on? How could they have any dynamic role in anyone's psychology? Where would they *come* from? The trouble is that once the psychophysical link is loosened to mere natural necessity the ontology of mind comes out looking pretty radically Cartesian.

Both problems have a common source: the denial of logical supervenience. It is therefore extremely important that this denial be shown to be undeniable. Chalmers is aware of this and argues that putative notions of *a posteriori* supervenience, in which there is no *conceptual* entailment from one level to the other, will not provide a viable alternative. Only logical supervenience can block the conceivability argument to the possibility of zombies. I find him quite convincing on this, but he underestimates how pressing it is to find some way to defend strong metaphysical supervenience, in view of the problems that arise from denying it. The crucial question here is whether all forms of logical supervenience must be epistemically transparent to us. Must our present concepts allow us to appreciate the nature of the supervenience relations that constitute the psychophysical link? Might we not instead be confronted by a case of *opaque* logical supervenience? If that were so, then there would exist concepts of both the physical and the experiential, and of whatever relations might connect them, such that there is an *a priori* explanatory

continued on page ix

The problem with zombies

BY ROGER PENROSE

David Chalmers presents his views on consciousness and its relation to the physical world very much from a philosopher's perspective. Since I find many of the issues that tend to occupy philosophers' attentions hard to relate to, I am approaching his work more as a contribution to our scientific understanding of the elusive nature of consciousness, than as a philosophical discourse.

Yet, scientists have much to gain from the philosopher's input. There are many confusions to be clarified at the very foundations of our physical pictures. The role of consciousness, in relation to the physical world, provides perhaps the deepest of these potential confusions. Chalmers is a philosopher of distinction who has thought long and hard on these matters and is well-versed in most of the fundamental issues underlying present-day physical theories. I therefore started with high hopes that his own insights could shed important clarifying light on these central issues. In this, I felt somewhat disappointed. However, I believe that there are valuable arguments given here which may contribute importantly to our final understanding of the puzzle, although perhaps not in the way that Chalmers intended.

Much of the book is in a philosopher's style that I find uncongenial. Unfamiliar words are introduced in order to make what seem to be hair-splitting distinctions. For example, numerous subtly different notions of consciousness are discussed at length. Can one be "aware" of a sensation without "consciously" noticing it? This relates to the philosopher's conundrum of a continuous noise that suddenly stops: one may become conscious of the noise only after it stops. (But is this not merely a reflection of the time-delays that can be involved in conscious awareness that we know are present from the important experiments of Benjamin Libet and his colleagues?) Definitions are often confusing to an outsider such as myself because they are frequently given only by example. I never grasped the sense of "naturally possible" — as applied to a randomly acting monkey typing *Hamlet*, but not to a persisting cubic mile of uranium 235. (Both are merely matters of probability.) There is also much analysis of the statement "water is H_2O". (Believing that *steam*, not water, is H_2O, I found this additionally confusing.)

"Zombies" — who act exactly like conscious human beings while being entirely unconscious — occur frequently. It is argued that such beings are conceivable, and the conclusion is drawn that consciousness cannot be a physical phenomenon. Why not more reasonably conclude that an unconscious physical being must behave differently from a conscious one, consciousness being a physical phenomenon? (Just because Chalmers can imagine a zombie does not make it possible; I can "imagine" a counter-example to Fermat's last theorem.)

Much faith is placed in the type of argument which depends upon the assumed theoretical possibility of successively replacing every individual neuron in a conscious person's brain by "a silicon chip that performs precisely the same local function as the neuron". Using ingenious new arguments (concerning "fading" and "dancing" qualia), Chalmers persuasively deduces that a robot whose silicon chips are wired in precisely the same way as are the neurons in his own brain would have just the same conscious experiences as himself.

Chalmers is much less convincing in deducing that John Searle's Chinese room could actually experience a "redness" sensation.

continued on page ix

Pilgrims on an unknown road

FELIX BENNETT

"Curiously enough . . . it is sometimes those basic problems that look impossibly difficult to solve which yield the most easily. This is because there may be so few even remotely possible solutions that eventually one is led inexorably to the correct answer." These words of Francis Crick (in *What Mad Pursuit*) could turn out to apply to the problem of consciousness.

This book, the selected proceedings of a conference held at the University of Arizona in 1994, is a landmark in the study of consciousness. It contains 56 articles from disciplines as diverse as philosophy, cognitive science, neuroscience, neural networks, quantum theory, hierarchical organisation and phenomenology. Many are specialised, but there are also some for the general reader.

Three criteria by which to gauge the articles might be as follows. First, do they take the first-person perspective of subjective experience seriously? Second, is a satisfactory explanatory framework supplied? Third, is an approach evident that brings together both philosophy and empirical science?

It is often thought that the first-person perspective can only be investigated in human subjects not animals. Thus blindsight is a phenomenon which has previously been studied in people with a damaged striate-cortex. They claim to be blind in the visual field dealt with by this damaged area but when shown something they cannot "see" and asked to guess what it is, they choose correctly with a high statistical significance. Here, in a fascinating paper by Petra Stoerig and Alan Cowey, monkeys with apparent blindsight are investigated. Through a series of experiments, Stoerig and Cowey coax the monkeys into effectively telling them whether they can guess the position of something in their blindsight or damaged receptive field, and reporting whether they can see it.

Britt Anderson and Thomas Head produce evidence of a kind of blindsight in language. They show that patients who are severely "aphasic" and overtly lack language comprehension, respond covertly to funny limericks, as demonstrated by changes in the electrical potential of their skin. It is amazing that the full comprehension of a complex spoken utterance can be demonstrated in the absence of overt comprehension, and possibly of conscious access.

The timing of conscious experience is looked at in an article which outlines the pioneering work of Benjamin Libet. He shows that people appear to backdate experiences, so as to discount the delay from initial stimulus to conscious experience. Here experiments pit neuronal activations against the accompanying subjective reports made by the individual. The paper by Mikael Bergenheim *et al*, which investigates simultaneity and synchronisation of sensory information in human subjects, makes a good companion piece to Libet's.

The way that different modalities and domains may interact in a person is examined in Thaddeus Cowan's paper. Here a curious effect is examined in which an after-image of the subject's hand is created by a bright flash of light in an otherwise pitch-black room. When the subject moves his arm, he sees the after-image move with it in the darkness. The visual map of where the hand is and a motor map of where the hand should be, interact.

Alfred Kaszniak and Gina DiTraglia Cristenson review research on patients who have no idea of their acquired disabilities, such as paralysis and memory loss. This lack of self-awareness seems to be correlated with specific types of brain damage peculiar to Alzheimer's patients (localised in the frontal cortex).

This brings us to the question of perceptions about perceptions. Güven Güzeldere, a philosopher of mind, argues that the so-called higher-order perceptions are a conceptual fallacy; they muddle up what is being represented with who is doing the representing. But Güzeldere fails to note that the inner structure of perceptions may well involve perceptions about perceptions anyway. This may be part of the "bootstrap" process that gets consciousness off the ground. We do not understand enough about perception to make the claim Güzeldere does.

All the above papers take subjective experience seriously. Other papers in the book demonstrate how techniques such as EEGs, Pet scans and the study of split-brain patients are revealing how conscious experience correlates with brain function.

A number of writers claim that quantum physics can explain consciousness. This may be a reductionist fallacy. How could the collapse of the wave function capture what it *feels* like to smell a rose? Nevertheless, several articles do draw fascinating parallels between quantum physics and the mind. These include the smeared-out nonlocality of both mental and physical events, their holism (where parts interact to form a global unity) and the actuality of the present. But these parallels may just be parallels. Quantum physics may not have anything directly to do with consciousness.

BY PAUL G. CARO

TOWARD A SCIENCE OF CONSCIOUSNESS: THE FIRST TUCSON DISCUSSIONS AND DEBATES
EDITED BY STUART R. HAMEROFF, ALFRED W. KASZNIAK AND ALWYN C. SCOTT
MIT Press, 786 pp, £47.50
ISBN 0 262 08249 7

Stuart Hameroff and Roger Penrose tell a particularly elegant story in which a macroscopic quantum field is supported by the crystalline structure of microtubules. Consciousness occurs when this field collapses. Much is made of the "noncomputability" of this collapse, said to be induced by quantum gravity. But if one is not hung up on noncomputability, there seems little reason to buy into this story. Is Penrose tilting at shadows?

An important neuroscience article by Christof Koch outlines his work with Crick on the neural correlates of consciousness. Koch suggests that area V1 of the visual cortex is not conscious, because it does not project directly to the frontal cortex, which is involved in planning and voluntary motor outputs. But it might be a mistake to discount the mail boy just because he has never met the chief executive. Maybe the mail boy has a soul as well: area V1 does not just pass on information, it also gets feedback from above.

What is known as the "binding problem" refers to the question of how the brain joins outputs from separate, functionally distinct regions together into the perception of a unified object, without having an assembly area. Synchronous 40 Hz firing between spatially separate groups of neurons might tie their informational content together by their coherent rhythmical activity. Valerie Hardcastle is sceptical on the basis of the evidence. However, this is a hot topic and new evidence is accumulating rapidly.

Neuronal rhythms might also be involved in attention. Tokiko Yamanoue presents a simple neural network model with oscillatory activity. Without any additional mechanism, it seems to mimic a number of well-known properties associated with human attention amazingly.

Alwyn Scott attempts to draw a global picture of the hierarchical structure of the universe, of which consciousness is one level among many. This is a delightful read. However, we need a tighter conceptual framework that will capture the logic of how consciousness is generated, why phenomenal experience is self-evident, what role self-reference plays, and why the mind both floats on the workings of the nervous system, yet is sealed off from it. I think mathematical logic may be a clue here.

Although we have to take the first-person perspective of subjective experience seriously, explanatory conceptual frameworks are seldom developed from this point of view. David Chalmers argues that it is the organisation and circuitry underpinning consciousness which matters, not whether it is realised in neurons or silicon. If the substrate were suddenly to change from organic neurons to silicon chips, how could this be accompanied by a change in the subject's consciousness, from, say, seeing red to seeing blue? Since there is nothing in the subject's behaviour to indicate a difference, and, more profoundly, there is nothing internally to allow the subject to *know* there is a difference, Chalmers concludes that the substrate cannot matter.

Another way of arguing this, I suggest, is to start by realising that explanations are purely relational structures or maps. If someone argues that the material substrate of the circuitry matters, in explaining why he would have to unpack the material and the workings of the circuits, ie regard them as a relational structure. But we could then substitute an alternative substrate to instantiate this structure, and so leave its conscious correlate unchanged.

Explanations as relational structures may apply, I suggest, at the phenomenal as well as the physical level. Thus we could ask what is required in order to distinguish between two percepts, such as seeing red and seeing green. In order to answer this question, we might consider a scenario where someone loses this discrimination, and can see only "gred" when either red or green is presented to the senses. How might he now reinstate his faculty of discrimination? He would need two independent markers that allowed the subject to tag "gred" in two different ways, one pair of which (say tag 1 plus gred) would then *be* the percept "red" and the other (say tag 2 plus gred) would *be* the percept "green". These tags must themselves be independent conscious percepts (if they are to allow the subject to *know* there is a difference). Seen in this way, the essential components of sensations are other related sensations, which are assembled to form more complex recursive hierarchies.

As Crick surmises, there may be so few even remotely possible solutions to the problem of consciousness that we are led inexorably to the correct answer. But this will require scientists to understand the philosophical constraints on a theory of consciousness, and philosophers to have the courage to create theories with testable predictions.

Paul Caro is honorary research associate, department of mathematics, University College, London.

Wise incomprehension

continued from page vii

connection between those concepts — even though they are not concepts we do or even could grasp. The conceptual dependencies would go outside of the circle of concepts we bring to bear in thinking about mind and body. Indeed, these concepts cannot be within our grasp or else it *would* be plainly inconceivable to us that zombies are logically possible. In other words, zombies *seem* possible to us only and precisely because we do not grasp the concepts that render them *im*possible. There is logical supervenience after all, but it is hidden to our epistemic faculties.

This is surely a coherent position, and it provides an alternative to the other relations Chalmers mentions. In fact, he does briefly discuss something like this at one point, correctly attributing it to me. But he does not see how serious are the consequences of rejecting it, since it seems to be the only viable way to avoid the twin problems of epiphenomenalism and disembodiment, while accepting that we cannot reduce experience to physical properties.

It is not dogmatic materialism that prompts insistence on strong supervenience but the need to escape the two problems cited. Indeed, the thesis of opaque logical supervenience is not materialist at all, if *that* means that the terms of current or foreseeable physics are adequate to explain consciousness. The view is actually quite compatible with theories that regard the physical as itself just the appearance of some deeper currently unconceived reality — or with idealism for that matter. Of course, the view assumes that we do not *know* the concepts that are necessary for a satisfying explanation of conciousness; what it does is use this fact to explain why it is that we can be misled into denying logical supervenience, with all the problems that stem from this.

It helps here not to be too wedded to the old framework of "materialism" versus "dualism". Both notions assume that materialism is a useful well-defined doctrine, but it is not, since the notion of the "material" is entirely theory-relative. We do not want to limit our theoretical concepts to those of current physics, but if we make the notion more inclusive it comes to include anything that might be relevant in explaining what happens in the world. There are really a *lot* of properties that might be identified and used in explanations of consciousness. Perhaps because he sticks to the old materialism-dualism dichotomy, Chalmers finds it hard to imagine how there could be concepts that transcend those now used in physics or commonsense psychology, and hence finds the idea of opaque logical necessitation difficult to accept. The first order of business here is not to declare materialism false, but to question its very significance.

The speculations on information and panpsychism are admitted to be a bit on the wild side, but the problems go beyond mere incredibility. Not only do we see no evidence in nature of the experiential properties allegedly associated with every causal process; it is also not the case that physics finds any need to postulate such properties in explaining the behaviour of matter. If all matter has experiential properties, should not this be relevant to the correct science of matter? Yet there seems no gap in the physics of the inanimate that calls for the ascription of mental properties to things. These alleged properties make no difference to the way a rock falls or water flows or any other purely physical interaction. The only motivation for invoking them is in order to provide an explanatory account of consciousness; they are idle otherwise. Subtract them from that thermometer and you will not observe any change in its behaviour.

Chalmers's defence of a weak form of functionalism uses some intriguing thought experiments, but the conclusion that there is a lawlike relation between functional properties and consciousness is too weak to be of much interest. We might equally claim that there is also a law-like relation between experiences and underlying neural states: if you keep the latter constant you will always get, as a matter of law, the same experiences. No *asymmetry* is established between the functional and the neural if law-like dependence is all that is asserted; so it is wrong to suppose that any interesting form of functionalism has been established. All we have is a three-way law-like relation between the mental, the neural and the functional.

The only way to avoid being checkmated by consciousness is to assume you do not understand it. Chalmers has done his level best to understand consciousness, but the result, despite its many merits, shows the wisdom of incomprehension.

Colin McGinn is professor of philosophy, Rutgers University.

RONALD GRANT

Zombies — physically like us, but the physical facts alone cannot guarantee the presence of a conscious life

Problems with zombies

continued from page vii

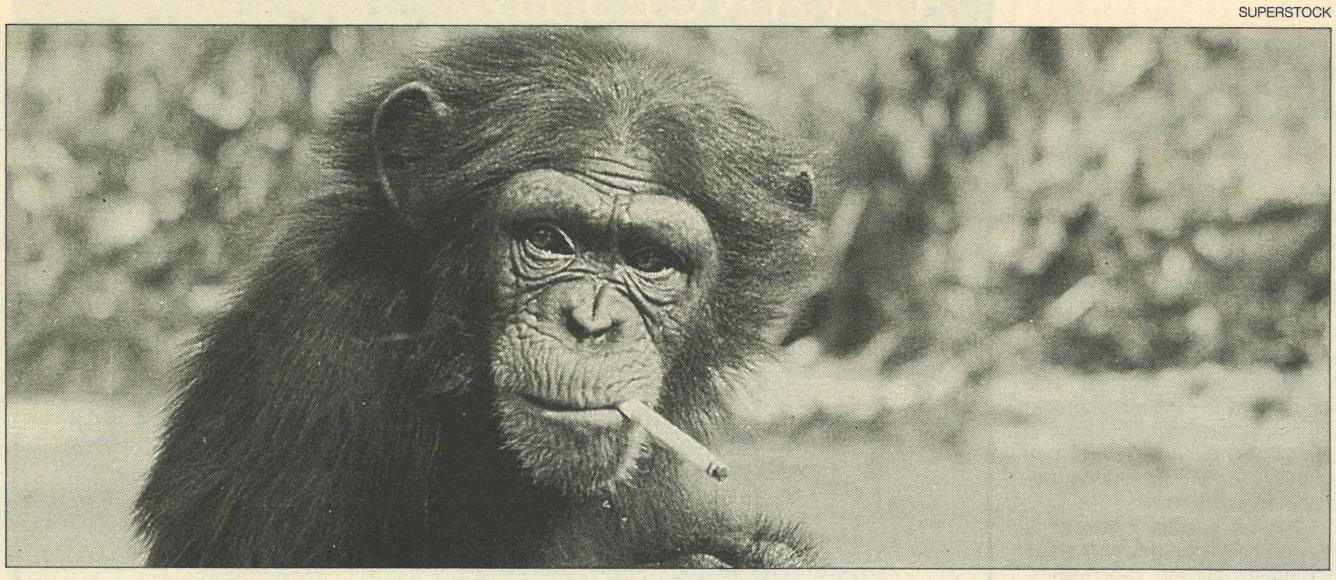
Accordingly, while I side with Chalmers on his first argument, I must support Searle on the second — which seems to lead us to a contradiction: both arguments require the *assumption* that such function-preserving neuron replacements are possible; hence they are *not* possible! Previously, I had depended upon arguments from Gödel's theorem to arrive at such a conclusion — but we now see that it also follows from this completely different line of reasoning. (Chalmers barely refers to the Göedelian case, dismissing it in half a page with an incorrect argument. Also, his interpretation of a finite-state Turing machine as an infinite "combinatorial-state automaton" is inappropriate, for reasons I cannot go into here.)

Perhaps a fundamentally non-computable physical input occurs at the neuron's cytoskeletal level (as Hameroff and I have proposed) — depending upon a presently unknown physics at the quantum/classical borderline. Chalmers, however, is dismissive of the possibility that our present-day quantum theory needs fundamental change. Instead, he is driven to the "many-worlds" (or "many-minds") Everett interpretation which he admits is "almost impossible to believe". It is here that his arguments become least credible (and if he really believes them, he should go over the entire reasoning of his book all over again, in the light of his changed perspective).

Perhaps it demands an unreasonable boldness to accept that our present-day quantum physics requires revolutionary change — as with Einstein's overturning of Newton's superbly accurate gravitational theory. Yet I believe that such a change is necessary, and the strong artificial intelligence/many-worlds deductions that Chalmers feels driven to, are unwarranted. But he is right to stress the inadequacy of "conventional" scientific approaches to consciousness, and his logic is normally impeccable. I believe that there is much of lasting value in his book, despite my profound disagreement with his final conclusions.

Sir Roger Penrose is Rouse Ball professor of mathematics, University of Oxford.

SUPERSTOCK



Do chimpanzees have a theory of mind?

When are the neighbours in?

BY DAVID ELWELL

THEORIES OF THEORIES OF MIND
EDITED BY PETER CARRUTHERS AND PETER K. SMITH
Cambridge University Press, 390pp, £14.95
ISBN 0 521 55916 2

THE POSTMODERN BRAIN
BY GORDON GLOBUS
John Benjamins, 188pp, $29.95
ISBN 90 272 5121 5

These days there is an interdisciplinary spirit abroad in the philosophy of mind. Arguably much of the most interesting work being done in this field draws extensively, and in an informed way, upon what is going on in other disciplines. Both of these books bear witness in different ways to that spirit of collaborative engagement.

Theories of Theories of Mind collects together papers by philosophers, psychologists and primatologists based upon their contributions to a series of interdisciplinary workshops sponsored by the Hang Seng Centre for Cognitive Studies at Sheffield. Here "theory of mind" is a term of art used to designate the ability, possessed by most human beings and, possibly, some other species, to make sense of the minds of others.

As a domain of research this is a particularly good example of the fertility of interdisciplinary collaboration: the question "do chimpanzees have a theory of mind?" posed by primatologists, led, via suggestions from the American philosophers Daniel Dennett and Gilbert Harman, to the development of the "false-belief task" used by developmental psychologists to investigate the acquisition of a theory of mind by children. The idea behind the task is that the correct attribution of a false belief to another person is evidence of the ability to think of others as having beliefs about the world which are distinct from one's own. Normally children become able to succeed at the task between the ages of four and five years old. In autism, however, success with the task is significantly delayed, if it is attained at all. This has led to the proposal that the primary deficit underlying autism is the absence of a theory of mind in the autistic subject.

These elements are all represented in the collection. The longest section, however, concerns the issue of what underpins the theory of mind abilities of mature human adults — how do we explain and predict the actions of ourselves and others? Here the field is occupied by two broad alternatives. "Theory"-theory attributes "mind reading" abilities to the possession of a theory of the psychology of others that sustains explanation and prediction of their behaviour. Simulation theory, on the other hand, denies that these abilities are due to the grasp of a theory, holding instead that they arise from our capacity to project ourselves in imagination into other people's perspectives, simulating their mental activities with our own.

There are papers by champions of each of these positions. Jane Heal, for example, argues against a thorough-going "theory"-theory view on the grounds that it would have to include a theory of relevance allowing us to predict which of the myriad beliefs which comprise a person's world-view it is appropriate to consider in any particular circumstance — such a theory would be massively complex and, insofar as we have no idea what form it might take, deeply tacit.

Nichols *et al* argue against simulation theory that it cannot adequately account for our failures to predict behaviour in some situations. On the simulation view such a failure must be due either to the subject of prediction being different to the predictor, or because the wrong beliefs and desires are fed into the simulation, and it is claimed that there are cases of mistaken prediction that fall into neither category. Only if prediction is based on a theory, it is argued, subject as it is to error in its theorems, can we explain these mistakes.

These papers, while they do not resolve the debate, suggest that a consensus may be emerging involving elements of both broad views. The collection is valuable both as an introduction to an exciting area of research and as a snapshot of the current state of the debate.

The Postmodern Brain seeks to forge different kinds of interdisciplinary connections, leaping the apparent chasm between contemporary philosophical reflection upon cognitive science and the postmodern philosophy of Heidegger and Derrida. Gordon Globus holds that we are bound to make no progress in understanding how the brain is the physical substratum of the mind so long as we see the brain as a computer engaged in the manipulation of representations. He sees in biologically realistic neural networks the resources for a noncomputational conception of brain functioning, the characteristics of which are the basis of the connections he strives to establish between this radical connectionist view and some central concepts in postmodern continental philosophy. A deconstructive approach is applied to types of narrative on aspects of the mental, including the computational theory of the mind and classical dynamic psychotherapy, which essentially embody the "metaphysics" that postmodern philosophy rejects. And in parallel, the postmodern, connectionist picture of brain functioning is applied to the explanation of mental illness and dreams.

In the field of cognitive science there is as yet no consensus on how connectionist models relate to computational theories of cognition. Are neural networks simply implementations of computational processes by a brain-like structure, or, more radically, do connectionist models rival and replace classical computational models? Globus sketches some arguments from the literature for taking the latter position.

The central contention of this book is that there is some interesting connection, or "resonance", between postmodernism and the radical connectionist view of brain functioning. Echoes there may indeed be though I found them difficult to assess, but they seem insufficient to sustain claims about the postmodern brain. The attempted assimilation seems to me to fail to sufficiently address the challenge that from a postmodern perspective the radical connectionist model of brain functioning is itself engrossed in "metaphysics", and as such is grist to the deconstructionist mill. Indeed a related methodological tension pervades the whole book, with the writer moving untroubled between "constructive", linear accounts setting out the connectionist model and making the connections with postmodernist concepts, and the deconstructive treatments elsewhere. On the whole I found this book rather thinly argued and, in places, decidedly opaque.

David Elwell is a psychiatrist and graduate student in philosophy, University of Oxford.

In brief

QUESTIONING CONSCIOUSNESS: THE INTERPLAY OF IMAGERY, COGNITION AND EMOTION IN THE HUMAN BRAIN
BY RALPH D. ELLIS
John Benjamins, 260 pp, $34.95
ISBN 90 272 5122 3

This book's title is a three-way pun. The author not only asks many questions about consciousness, he also questions many received views on it, and goes on to claim that a "questioning process is at the heart of consciousness". It all makes for an exhilarating and informative read.

He takes imagination as the basic building block of consciousness. The initial brain processes are the same for imagining, remembering and actually seeing: in each case a mental image is formed and "looked for". Only if these "outflowing" signals from the prefrontal cortex are matched by "incoming" signals from the retina, is the object "seen" rather than just imagined or remembered. But — and this is crucial — it is not sufficient merely for light rays from the object to fall on the retina and be passively received: one must already be "looking for" the thing in order to see it. This is why Ellis claims that questioning is at the heart of consciousness. The incoming sensory information has to be "questioned" by the brain, and only if it matches something "looked for", will it enter consciousness.

This brings us to the mind-body problem, to which there are three familiar solutions. One is dualism, where consciousness is a feature of a nonmaterial mind which interacts with and controls the physical brain. The alternatives are both materialistic: either the conscious mind is an epiphenomenon caused by brain function, or it is to be identified with brain function. Ellis rejects all of these and offers instead a further possibility, that consciousness is a process which takes the functioning of the brain as its substratum. An analogy would be a wave travelling across the sea: the wave and the water are not identical, nor is one the cause of the other, but they are inseparable (a key term). Similarly consciousness is neither identical with nor caused by brain functioning, but it is nonetheless inseparable from it. Where consciousness goes beyond the wave analogy — and this gives it its unique character — is in its ability to create, replace and reproduce elements of its substratum as required.

ANTHONY FREEMAN

LOCATING CONSCIOUSNESS
BY VALERIE GRAY HARDCASTLE
John Benjamins, 264pp, $34.95
ISBN 90 272 5124 X

The "problem of consciousness" is at last emerging from the shadows of philosophy into the full glare of scientific day. The signs are everywhere — why, even that scientific holy book, *Nature*, now regularly publishes articles using the hitherto taboo word. But philosophical issues still cling to this lusty new field of scientific inquiry, like almost outworn swaddling clothes. So, to orient us to this emerging field, we need a guide as well versed in philosophy as in the relevant sciences. Valerie Gray Hardcastle fits the bill admirably: I cannot tell whether she is a philosopher turned scientist or the reverse, she seems so equally at home with both. She steers an expert course through the contending views of the relations between brain and mind that come from philosophy, but refuses to be tempted by any of them away from the key new understanding: that the problem of consciousness is one to be solved by the normal methods of empirical research and theory construction that make up the natural sciences, not by the purely conceptual analysis of philosophy. She sketches out the major areas of relevant knowledge gleaned from difficult and ingenious experiments in both psychology and neurophysiology. But she is not content merely to rehearse the views and findings of others; on the contrary she comes up with some novel scientific hypotheses of her own, which are likely, I believe, to be taken seriously.

JEFFREY GRAY

QUANTUM BRAIN DYNAMICS AND CONSCIOUSNESS: AN INTRODUCTION
BY MARI JIBU AND KUNIO YASUE
John Benjamins, 242pp, $29.95
ISBN 90 272 5123 1

Mari Jibu and Kunio Yasue present "a theoretical framework, called quantum brain dynamics, to investigate consciousness scientifically in light of the first principles of physics, that is, the most fundamental laws of quantum field theory. This framework is based on the original physical theory of memory and brain functioning found in the quantum field theory developed by Ricciardi and Umezawa in the 1960s".

Here is a telling paragraph: "What is consciousness? What is mind? Of course, we are not going to delve deeply into philosophical and epistemological considerations on consciousness. Rather, we are going to reveal what kind of physical phenomena might be identified with the fundamental process of consciousness from the purely physical point of view of quantum brain dynamics. The QBD system is nothing but water extending across and penetrating the whole assembly of brain cells and interacting with the electromagnetic field inside the cranium".

If we work up a competent knowledge of the physics and electrical engineering of our TV sets, that knowledge will not help us to decide how far to trust "the person talking on the screen". Jibu and Yasue's book is a competent history of theoretical physics. But what does that have to do with consciousness?

ALAN RIDDIFORD

Revd Anthony Freeman is managing editor of the Journal of Consciousness Studies*; Jeffrey Gray is professor of psychiatry, Institute of Psychiatry, London; and Alan Riddiford is an engineering physicist at the Fermi National Accelerator Laboratory, USA.*

Spooky systems and magic bullets

BY SUSAN GREENFIELD

EVOLVING THE MIND: ON THE NATURE OF MATTER AND THE ORIGIN OF CONSCIOUSNESS
BY A. G. CAIRNS-SMITH
Cambridge University Press
329pp, £16.95
ISBN 0 521 40220 4

Of all the frustrating, murky biological phenomena it is consciousness, with its quintessential subjectivity, that exercises a surprisingly strong pull on a growing group of physical scientists and mathematicians. The latest moth to the flame is Alexander Cairns-Smith, an organic chemist. The particular path that the author follows, although it is not stated explicitly until two thirds of the way through the book, is one whereby we might obtain a scientific model of what consciousness "is like". Curiously, the author dismisses the more understandable aim of discovering what consciousness actually is without explaining clearly why, and without giving any estimate of just how much, or how little, he hopes his model will achieve.

It is a model that is nonmathematical, physical and "spooky". Indeed, the theme of spookiness is with the reader from the outset. Cairns-Smith's vision of the brain is one viewable as three different systems, accounts of which give the book its overall structure.

System one is the world of the chemist-physicist, where molecular and submolecular interactions occur without reference to any grander design. Cairns-Smith is clearly at home here: he writes with authority and the ease of metaphor and image that betokens the experienced teacher. On the other hand general readers will find certain passages very tough going and might wonder why they have to take on board a morass of technical material occupying a good third of the book, yet which has no obvious link to consciousness.

Nonetheless, even a rather cursory read would reveal the take-home message in the author's view of the fundamental components of the physical world. It is a world where atoms are not autonomous, where events are not local but global, and where energy and matter blur into each other. Nothing is quite what it seems.

Having explored the concept of matter, we are shown how matter underlies life. It is a shame that the "spookiness" or otherwise of life itself is not discussed. However valuable thorough discussion of enzymes, ATP and so forth might be, a more engaging slant for the general reader would have been to demolish once and for all the vitalist stance. One feels that the author missed an obvious opportunity here to show that life can be reduced to familiar, nonmagical elements — but that the emergent properties of those elements are, for the time being at least, exclusive to biological systems. Surely it is this counterintuitive aspect of seemingly magical emergent properties that agrees rather well with Cairns-Smith's basic disposition.

Another more helpful key word that is stressed, however, is "communication". This idea dominates once we enter system two. Here we deal with signalling within the physical brain, from one neuron to another. Cairns-Smith rather misleadingly refers to this system as "neural computation", presumably to draw a sharp distinction with the ethereal, holistic consciousness that will constitute level three. On the other hand, the term might mislead the casual reader into assuming the physical brain is like a computer, a stance which Cairns-Smith does not particularly try to sell, even though he persistently describes neuronal connections (as many have before him) as "wiring", an inflexible, inelegant and to my mind inaccurate metaphor.

Sometimes the images pay off, for example, the action of cell adhesion molecules is described as "neurons waving their sugary flags", while axonal transport is crisply portrayed as cellular products efficiently "going by rail". On the other hand there are times when the analogies do not speak immediately to the nonphysical scientist: the paradoxically passive process of generation of action potentials is compared to Westinghouse brakes, rather than, say, the more banal yet familiar example of a deflating balloon squirting through the air. However, the overall impression is of a writer making an enormous effort to convey obscure and difficult ideas in an accessible way. It would be very surprising if Cairns-Smith hit the bull's eye with every metaphor every time.

A more worrisome problem is that by attempting a rather detailed review of the brain from the neuronal level, "bottom up", the reader tends to slip uncomfortably between stools. Too much space is devoted to too much detail for the general reader to be able to see the wood for the trees. And there are features of the wood that are relevant but have been omitted. For instance, it would be helpful to know that the firing of action potentials is not the only way of observing discrete populations of neurons at work. The elegant work by Grinvald and Aertsen showing how neuronal assemblies of varying size can form and reform within fractions of a second, may well be of relevance to anyone interested in seemingly "spooky" brain functions. Similarly, discussions of clue-laden phenomena such as synesthesia, phantom limb pain or prosopagnosia would be more likely to rivet the reader's interest than even the most cursory mention of ligand-gated ion channels and G proteins.

Meanwhile, golden opportunities slide away. We are given a relatively detailed neuroanatomical description of the brain, but are not treated to any discussion of the rationale and problems of localisation of function. The chemical structures of key transmitters are shown in a figure, but not their distribution in the brain. System two could have served as an invaluable bridge between the matter of system one and the mind of system three. As it stands, there are no clear clues, and readers are left with the impression that they have leapfrogged over the physical brain in order to relate molecules, atoms and other particles directly to consciousness.

Cairns-Smith as much admits this bias in the final chapter, which is couched as an engaging dialogue between himself (presumably), "Advo", and "Crit", a hypothetical adversarial chum. Advo: "How is it [consciousness] affected by system two? How does it act on system two?" Crit: "That I can answer easily. Don't know." True, none of us knows, but the phenomena of consciousness (system three) that are clearly rooted in perturbations in system two, for example phenomena involving mood-modifying drugs or, more specifically, blindsight, get very little airing.

Even though Cairns-Smith would probably agree that system three is most likely to arise from system two in some way, there is a feeling of rupture, a noncontinuity between the boring old slavish neurons plodding away algorithmically, and the will-o'-the-wisp of consciousness as it flits about the brain, delocalised and defiantly special.

Consciousness is hastily defined as "feelings", but as this profile is filled out during the narrative, we start to encounter some contradictions. On the one hand Cairns-Smith refers to consciousness when one is not paying attention as "spread out", but a little later he speaks of the relative simplicity of an actual experience at any one instant. It is hard to know how the author actually views consciousness since he spends very little time exploring any of its characteristics.

Still, a less-than-clear picture of consciousness should not deter us from the author's central theme: a quantum version of events along the lines originally proposed, we read, by Danah Zohar and Ian Marshall. Only in this final section of the book does Cairns-Smith's picture finally become less opaque. Having laboured in the early chapters through the fuzzy interconnections of time and space, energy and matter, we can now apply a comparable approach to mind and brain, where wonderfully ambiguous, delocalised quantum events collapse into the humdrum Newtonian world of the physical brain. Like Roger Penrose, Cairns-Smith has a vision of quantum coherence across banks of neurons: in his case, however, the orchestration takes the form of a Bose condensate (a Marxist-like organisation of particles whereby the individual is subsumed into the general whole so as to pull together towards a collective end).

I have two immediate scientific objections to this idea being actually implemented in the brain. First, Cairns-Smith conjures up the idea of "specialist brain cell proteins" that play a crucial, and presumably committed role as "suitable oscillators" in the generation of consciousness, but which thereby come perilously close to resembling that anathema of nonvitalist scientists and philosophers alike: magic stuff. But perhaps I am reading too much into terminology. A second anxiety is far more serious, since it concerns whether quantum events could ever unfold in the first place. The temperature in our heads is so high as to make it extremely improbable that boson condensation could play a significant role in brain function. Granted, Cairns-Smith acknowledges this problem, but to dismiss the issue by saying that "it is by no means implausible" that natural selection has circumvented the problem in an as-yet-mysterious way, just will not do.

A more general complaint is that all the favourite topics relating to the quality of the first-person experience are not elucidated at all. Whether or not Newtonian grounded action potentials are replaced by quantum coherence, the philosopher is no nearer to enlightenment on how such events might be translated into, or even correspond with, a red quale. Neuroscientists too might be frustrated, since no attempt is made to interpret phenomenological events such as Alzheimer's disease, schizophrenia (although they are mentioned) or even dreams, in terms of the model. As it stands therefore, the concept of boson condensation being "like" consciousness, currently has only limited usefulness. On the other hand no book of this sort should be dismissed for not producing the magic bullet of the physical basis of the mind. Cairns-Smith has a story to tell and he does so, eloquently and well.

Susan Greenfield is lecturer in pharmacology, University of Oxford, and Gresham professor of physic.

KELLIE STRØM

Will this erring be read?

BY STUART SUTHERLAND

AWARENESS: WHAT IT IS, WHAT IT DOES
BY CHRIS NUNN
Routledge, 167pp, £37.50 and £11.99
ISBN 0 415 13226 6

In 1989 Roger Penrose put forward a revolutionary thesis. He argued that consciousness was associated not with macroscopic events in the nervous system as classically described, but with phenomena at the quantum level with all the uncertainty that that entails. In brief, his arguments were as follows. There are theorems in mathematics that cannot be proven by algorithms, but that mathematicians can see to be true. In so far as the nervous system operates algorithmically like computers (as usually accepted), it could not see the truth of such theorems. Some unspecified noncomputational mechanism must therefore be at work in the brain. Since individual particles could not affect the nervous system as we know it, coherent aggregates of particles (the Bose-Einstein condensate — particles that have all their attributes in common, as in a laser beam) must be formed somewhere in order to affect the conventional nervous system. Penrose believes such aggregates occur in the micro-tubules, which are part of the cell's cytoskeleton: he gives a series of cunning arguments to render this idea plausible. For example, unicellular organisms behave in a complex and purposive way: they have no nerve cells but they do have a cytoskeleton. Again, substances that render people unconscious, anaesthetics, are thought to affect the microtubules.

Penrose is meticulous in his use of evidence, but his highly speculative arguments are not completely persuasive. Although there are true propositions that are incapable of being proved within a given calculus, they might be proven within a higher level one (a metalanguage). Furthermore, one of the implications of Penrose's arguments appears to be that there are no true propositions that people cannot prove, but this is almost certainly untrue. For example, I cannot prove that I will perform a particular action (even if there is no change in the external world) since attempting to decide may change my state and I cannot take such changes into account without an infinite regress. Nor can I be certain that the decision process will halt before a solution is found, as opposed to being abandoned without a result being reached. Second, Penrose gives no examples from everyday life in which noncomputational methods must be used. Third, he is vague about how such methods operate in order to solve problems. He is reduced to saying that we do not at present understand the collapse of the wave function (the interconnection between the world of fundamental particles and the macroscopic world) and that once this is understood, which would require a revolution in physics, we shall understand noncomputational methods and also the nature of consciousness. This is a wish and a prayer, not a promise. Finally, if all consciousness depends on nonconscious quantum phenomena, the attempt of cognitive science to correlate consciousness with computational mechanisms would be a waste of effort, despite the successes it has already produced.

It was only a matter of time before someone would seize on Penrose's ideas and use them in ways he never intended. Ostensibly, Chris Nunn attempts to contrast Penrose's way of looking at consciousness and the brain with the classical view. In fact, the book defends Penrose's ideas by extending them to explore a number of bizarre phenomena — some real, most imaginary.

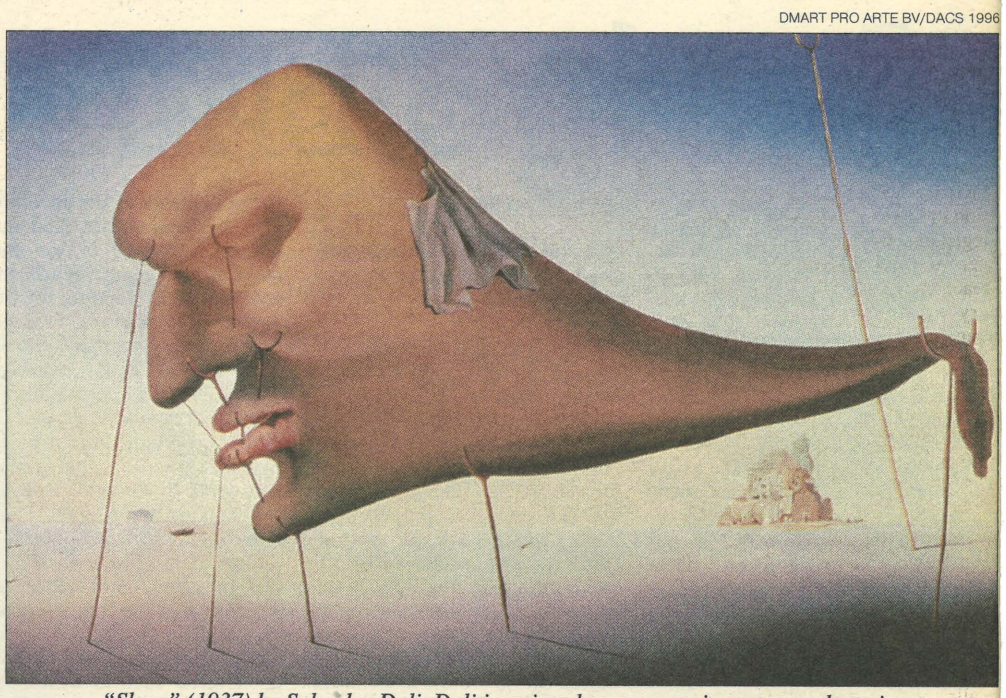
Nunn claims that telepathy can be explained by the Einstein-Podolsky-Rosen phenomenon — the fact that if two particles become "entangled" the attributes of one can be fixed by an observation on the other, however far away they are, even though before the observation is made, these attributes are not determined. Similarly, according to Nunn, psychokinesis is achieved by particles in the brain being somehow in tune with particles in the object affected by psychokinesis. No account of how quantum waves operate to achieve these effects is supplied. Nunn uses similar arguments to explain Jung's archetypes, and even the tendency for members of a group to think in the same way as one another. Talking to one another is apparently not sufficient. He also believes that the indeterminacy of the collapse of the wave function underlies freedom of the will: this opinion is inconsistent with the predictability of most human actions and if decisions are influenced by random events, the concepts of credit and blame would have to be abandoned.

Awareness contains many errors: for example, NMDA is not a neurotransmitter as alleged; we see the world in three dimensions not two; according to "identity theory" consciousness is not "the result" of conventional brain activity, it is such activity (or aspects of it) under a different description. Some of Nunn's arguments are, to put it mildly, odd. For example, he infers that consciousness must have evolutionary value because it uses up 30 per cent of the body's energy: hence, it must be useful or it would not have evolved in the face of its disadvantageous energy consumption. But surely it is the brain not consciousness that uses energy: there is no need to believe that awareness requires any extra energy whatever.

It is unclear whether Nunn understands what he writes, but it is certain that few if any of his readers will. *Awareness* may well make Penrose feel queasy.

Stuart Sutherland is emeritus professor of experimental psychology, University of Sussex.

DMART PRO ARTE BV/DACS 1996



"Sleep" (1937) by Salvador Dali. Dali imagines human consciousness on the point of falling asleep, as about to float away from reality's tenuous anchors. But consciousness is really far more mysterious

Minds, for the use of

BY ROBIN DUNBAR

COMPLEXITY AND THE FUNCTION OF MIND IN NATURE
BY PETER GODFREY-SMITH
Cambridge University Press, 311pp, £30.00
ISBN 0 521 45166 3

Philosophers' attempts to grapple with the mind have largely focused on answering questions about what minds are. Peter Godfrey-Smith's book is a welcome attempt to take a long philosophical look at the question of why minds exist at all. Evolutionary biologists (unlike almost all other disciplines in science) have long regarded "why?" questions of this kind as being of great importance.

"Why?" questions are important in that they broach the most fundamental questions of existence. With the possible exception of cosmologists, scientists have generally eschewed such issues, mostly in the mistaken belief that to ask them commits us to metaphysics or worse. But in biology, "why?" questions are at the very heart of the evolutionary process. To be sure, the answer must always be one of adaptive purposiveness and not the kind of teleological response so characteristic of the philosophical past. To distinguish them clearly from the latter kind of nonsense, the biological (or evolutionary) approach is sometimes said to be teleonomic — that is to say, it focuses on the mechanisms that drive the system in one direction rather than another. The thermostat is a teleonomic device in this sense. In the biological context, the mechanism is natural selection.

Godfrey-Smith's claim is that the mind evolved as a response to the need to deal with environmental complexity. "A central project in philosophy of mind over the last few decades," he observes, "has been the attempt to give a naturalistic analysis of intentionality, the attempt to say … what it is for an internal state of a system like a brain to represent or be about some other object or state of affairs". His aim, in contrast, is to show "that a basic mental tool kit has the function of enabling agents to deal with environmental complexity: that is, why the tool kit is there." His thesis, as he points out, does not depend on any particular theory as to how this cognitive machinery actually works; indeed, it is compatible with a number of theories of how thought manages to perform its everyday tasks.

Godfrey-Smith conceives his task in two parts. The first part of the book is a lengthy analysis of the positions of two once-influential philosophers, the Victorian Englishman Herbert Spencer and the latter-day American John Dewey. Godfrey-Smith intends these two to provide him with foils for exploring an externalist theory of the mind without need to commit himself to espousing either of their views in any detail.

He argues that, in their different ways, Spencer and Dewey provide complementary standpoints. Spencer's evolutionism led him to place a central emphasis on the mind as mechanism for coordinating behaviour (itself seen as the body's response to environmental complexity). Spencer is, in many ways, the point of intersection between the two most important streams in the 19th-century British intellectual scene, namely the empiricist philosophical tradition of Locke and Hume and the new evolutionary theories of the biologists. Dewey, doyen of the turn-of-century pragmatists, viewed beliefs as instrumental guides to action. Both focused on the role that thought plays in enabling the body to respond to a complex and variable environment.

In his attempt to develop a coherent externalist theory of the mind, Godfrey-Smith is uncompromising in his dismissal of the sillier features of the alternative internalist (or constructivist) theories. These relativist views are apt to see the external world as a mere construction of the way we are socialised to use language. Biologists in particular have always found the narcissism inherent in the constructivist view frankly offensive.

Godfrey-Smith's aim, then, is not to provide a theory of the location of the mental in the physical (the mind-qua-brain stories we associate with much conventional philosophy of mind) but rather to elucidate the role of the mind in nature and what we use it for.

In the second half of the book, he tries to get to grips with this question with the aid of some mathematical models. Drawing on the biological concept of *fitness* maximisation, he attempts to show *that the* ability to manipulate the world mentally in order to be able to predict optimal responses to environmental vagaries would be selectively advantageous. He sees the processes involved as analogous to a Bayesian model of experimentation. Indeed, his whole argument hinges around the question of when it would pay an organism to evolve the (rather expensive) capacity to learn.

These models are mainly game-theory models of the kind that are commonly used in evolutionary biology to assess selection advantages when an organism has a choice of alternative ways to proceed. Since decisions of this kind depend on mechanisms for acquiring information about the environment, Godfrey-Smith supplements his analysis with models derived from signal detection theory.

This has been a long-overdue exercise, and one that will strike a chord with evolutionary biologists. My own quibbles with the book rest on the fact that he devotes far too little attention to the so-called "social brain" hypothesis — the increasingly influential view that brain (and hence mind) evolution within the primates owes its origins not to the need to handle environmental complexity but to the growing need to handle the infinitely more complex problems of the social world. It is not an exaggeration to say that the average human mind daily executes calculations in the social domain of an order of complexity not far off that required to understand quantum theory. Yet most of us wilt at the mere sight of even a simple Newtonian physics equation.

In the end, the success of this book will obviously depend on how well it can convince not biologists but psychologists and philosophers. I fear that many of them will find the second half of this book difficult to follow. That will be a pity but in the interests of genuine interdisciplinary exchange I can only encourage them to persevere.

Robin Dunbar is professor of psychology, University of Liverpool.