Wonderfest 98: The Bay Area Festival of Science

Wonderfest is a non profit educational organization dedicated to the promotion of science. Wonderfest 98 was sponsored by San Francisco University High School with support from The Branson School, Yahoo!, Inc., Chiron Corporation, and Wind River Systems. Wonderfest is directed by Tucker Hiatt.

P.O. Box 887
39 Fernhill Road
Ross, CA 94957

Other Wonderfest transcripts and audio tape programs are also available. Please visit us at:

www.wonderfest.org

DO WE UNDERSTAND CONSCIOUSNESS?

John Searle and Walter Freeman compared notes at a Wonderfest panel at the University of California, Berkeley, 16 August, 1998

John Searle[1]

The answer to the question for this session - Do we understand consciousness? - is, 'No', but I won't just send you home with that answer. The real question is *why* don't we understand it and what are we going to do about it? The question of consciousness, which I regard as the leading question in the biological sciences today, is: How does it work? How, in fact, does the brain do it?

It's an odd fact that, until quite recently, was not regarded as a respectable scientific question at all. As recently as twenty years ago, when I first got interested in this field, if you went to a cognitive science conference or started talking to neurobiologists, and you urged them to study consciousness, most of them thought either it wasn't a suitable topic for scientific investigation at all, or else perhaps it was, but we weren't anywhere near ready to approach it. I think both views are mistaken, and I want to use most of my time here removing certain philosophical objections to the serious scientific study of consciousness.

The first obstacle is that a lot of people will tell you that we cannot even define consciousness. We need to distinguish here between an *analytic* definition, which comes at the end of the investigation when we know how a phenomenon works, and a *common-sense* definition, which just identifies the target. For example, the analytic definition of water is H20 but it took a long time to figure that out. In the meantime we have a pretty good working definition - it is this colorless liquid that runs in streams and comes out of faucets, and so on.

With consciousness, we are at the common-sense stage, not at the analytic stage. It is easy to give a common-sense definition of consciousness so I'm going to do it. Consciousness consists of those states of sentience, or feeling, or awareness, which begin in the morning when we awake from a dreamless sleep and continue throughout the day until we fall into a coma or die or fall asleep again or otherwise become '*un*conscious'. (On this definition dreams are a form of consciousness, although, of course, they are qualitatively different and have quite different characteristics from ordinary waking consciousness.)

So that is the target. We have to realize there is nothing fancy about it. I'm not talking about the transcendental awareness of the self or altered states of consciousness or states of self-consciousness or all that stuff you get in German philosophy... We are just talking about the sorts of feelings you have and your dog has and other kinds of higher primates have. On this definition we don't know how far down the phylogenetic scale it goes. Some people say no animals have consciousness, but they haven't met my dog, Ludwig. There is no question that Ludwig is conscious. But when it gets down to snails and termites, I don't know. I leave it to the experts and we just don't know enough to worry about that question. So that is our target. That is what we are talking about: these states of feeling, sentience and awareness - something it feels like. Now what's the problem?

[1] The session opened with short presentations by each speaker intended for a general audience. These are reproduced here. Readers already familiar with John Searle's and Walter Freeman's basic viewpoints may prefer to go straight to the discussion on p.7 below.

The first question is, how does the brain do it? How do these billions of neurons produce consciousness? (Not everybody thinks this is the right level of investigation. Some want to go lower, to the synapses, while others - such as Walter here - think we should look at much larger arrays or clouds of neurons rather than at single cells.) So here is the puzzle: how does the interactivity of these microelements cause our states of feeling or sentience or awareness?

When you get into that you are in the area of the traditional mind-body problem. I have to say a few words about that, because I think that essential progress in this field is impeded, to some extent, by these traditional philosophical problems, which can be stated like this:

Given that there are these two fundamentally different kinds of phenomena in the universe - mental phenomena, such as consciousness, and physical phenomena, such as cells in general and neurons in particular - how can there ever be a causal connection between the two? What is the relation between the mind and the body? The first step towards getting the answer to the problem of consciousness is to overcome that traditional question, and I think it is not hard to overcome it if we just remind ourselves what we actually know. It seems to me that we know at least the following.

All our states of consciousness are caused by neurobiological processes in the brain. We don't know exactly how it works but we know that the brain does it. The currently standard textbook theories are that the causally efficacious feature is variable rates of neuron firings relative to different neuronal architectures. That is a remarkable idea so let me emphasize it. It is this: all of your thoughts and feelings - everything from worrying about your income tax to trying to remember your mother-in-law's telephone number or suffering the *angst* of post industrial society - all of that is caused by neurobiological processes in the brain. We are talking about biological phenomena.

That leads to the next question: what is the consciousness that they cause? It seems to me that again we are talking about biology, about biological phenomena - the existence of conscious states of sentience or awareness - and that they go on at a higher level of organization in the brain. *Consciousness is a level of organization in the brain.* We are not talking about something that gets squirted out by the brain. It is not as if consciousness were a separate fluid or substance. We are just talking about a feature of the brain when it is in a certain sort of state.

So the solution, I think, to the traditional mind-body problem is rather swift. Once we get out of the idea that consciousness is something mysterious and not a part of the ordinary physical biological world we all live in, then it seems to me the following two propositions present a solution to the traditional mind-body problem:

- All of our conscious states are caused by lower-level neurobiological processes in the brain.
- Those states are themselves features of the brain.

Now when I say that consciousness is a biological process, a lot of people think that I am into neuronal chauvinism or have a fetish for carbon-based molecules, or some such nonsense. But I am not saying that we might not be able to produce an artificial device that was also conscious using some other means. My point is that until we know how the brain does it we are in a poor position to try to do it artificially. This leads me to discuss some standard mistakes.

One standard mistake, which is very much a part of the contemporary intellectual environment, is to think that we shall have solved the problem of consciousness when we have built a conscious computer. That what we've got to do is just get the right program and then the program will be sufficient to guarantee consciousness. You'd be surprised how widespread that view still is. The argument is that, since thinking is a matter of information processing, and since the whole point of inventing computers was as information-processing devices, then the way to solve the problem of conscious thinking is to program your computer so that it is conscious.

The number of mistakes in that set of propositions is huge, but the essential one can be identified quite swiftly; that is the mistake of failing to realize the difference between *computation* as a formal symbolic mathematical process going on in the computer and *consciousness* as a quite specific form of awareness or sentience going on in the brain. The program of the computer is defined entirely in terms of formal symbolic manipulation, usually thought of as the manipulation of 0s and 1s. It is a terrific technical achievement that you can do this at the rate of millions of operations per second, but the point remains that in the computer the steps are purely formal, whereas the brain is a causal mechanism that causes consciousness by quite

specific - though still largely misunderstood or not-understood - biological processes. The mistake of thinking you can do this with a computer is the mistake of supposing that the *syntax*, the formal symbols of the computer, are sufficient to guarantee the presence of the *semantics*, or mental content, that goes on in actual brains.

I proved this almost two decades ago with a very simple thought experiment. Imagine you are carrying out the steps in the computer program for something you don't really understand. I don't understand Chinese, and I imagined that I would go through the steps in a computer program for answering questions in Chinese. Now, you can imagine that the programmers get so good at writing the program, and I get so good at shuffling the symbols, that my answers to the questions in Chinese would be indistinguishable from my answers to questions in English. However, there is a difference. I understand English perfectly but I do not understand Chinese at all. The point of the thought experiment is that if I do not understand Chinese on the basis of implementing the computer program, then neither does any other digital computer solely on that basis, because no digital computer has anything I do not have.

This is not a weakness of computation - it is part of its strength. Your pocket calculator does not have to worry about income tax in order to shuffle the symbols to print out the right arithmetical answer on the display. The reason is that the calculator is a purely formal symbol-manipulating device.

Now the point of the argument is to show something we knew all along: that the syntax of the computer program is not sufficient to guarantee the presence of actual mental content.

Let me finally point to what I take to be the right direction for research to take. Recently I reviewed a number of books on how the brain causes consciousness and the short answer is that nobody knows how the brain causes consciousness. All of these efforts fail for one reason or another. But I am delighted to see that the *project* is going forward. The project is to try to figure out which specific features in which specific neuronal architectures actually cause the entire sequence of thoughts, feelings, sentience and awareness that constitutes our conscious life. We don't know the answer, but I regard it as the most exciting question in science today. And who knows - we may even get to solve it in our lifetimes.

Walter Freeman

I am a card-carrying biologist; therefore, I want to present my credentials by showing you some of these neurons that we are talking about *(Figure 1).* The cells are stained by the Golgi technique, which picks out a few cells and stains them in their entirety to show up the axons and dendrites, which make the connections between them. These increase with the age of the subject - newborn / three months / six years / adult - and the good news is that they go on making new connections. So there is hope for John and myself as we go on into our 70s that we can still learn something new, because it is the *connections* that make all the difference.

John has emphasized - and I agree with this thoroughly - that brains are responsible in some sense for consciousness and that it has an important role to play. I want to explain this ultimately in terms of neurons, but that is not quite enough. We have to ask how and in what way are they organized. John has further pointed out - and I think it is a very important thing to say - that consciousness is not some 'thing' that is located 'somewhere' in the brain. It is a mode or style of organization involving the cells. If you see a ballerina performing a pirouette, you do not ask whether the pirouette is in her ankles or her knees or her hips: it is the whole body, the pattern of motion, which is important. So also, consciousness is not located in the brain stem or cortex or frontal lobes; it is a style of organization, not only of the whole brain but also of the whole body. You can carry that even further and say it is the style of organization of a person in their social context. I think the social philosopher John Dewey was right on target when he

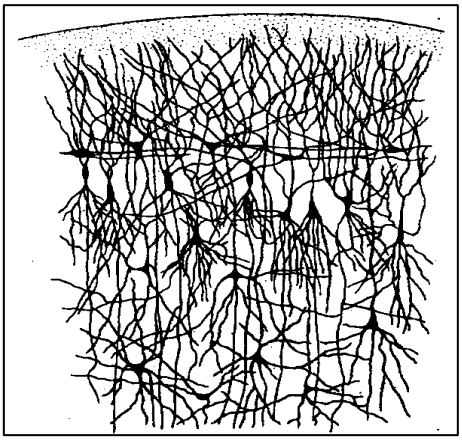

Figure 1.

Neurons stained by the Golgi technique, which shows up the axons and dendrites, which make the connections between them..

said that consciousness is meaningless unless it is defined in the context of a society.

Now let's bring it down to specifics. Some people think that consciousness is demonstrated when you perform an action such as raising your arm - that it is your conscious mind which is controlling your neurons and making you do this - but there are some problems with that view. In the first place, it does not account for involuntary or unintended actions. Nor does it explain the following kind of experience, recounted by psychologist William James, but familiar to all of us. In the morning he is lying in bed, telling himself to get up, but nothing happens. Some time later he finds himself out of bed and pulling on his socks or brushing his teeth, but has no recollection of how he got there. Clearly there are actions which brains take for which there is no conscious preparation or anticipation.

It gets worse. Neuroscientists, for the past forty-five years, have known that there are electrical events in the brain which can be recorded on the scalps of ordinary human volunteers. When they ask a subject to perform some action, such as raising an arm, they find that about one second before the arm goes up there are electrical changes in brain activity which provoke this action. Furthermore - and this is where it gets interesting - these changes occur before the subject is aware that he is about to perform the action. They even tell the experimenter which arm the subject is going to raise before the subject himself has decided what he is going to do.

This poses a very interesting philosophical problem. What is going on in the brain? Who is this little character inside your brain who is pulling your strings? Where is it located? The answer is straightforward; it is you, yourself, the real 'you' inside. Further, your knowledge about yourself comes from recognizing what it is that you do. You learn by acting and then investigating, evaluating and analyzing the consequences of your action. In some respects this is not new, because we have always known that knowing yourself involves looking at what you have done to tell what sort of person you are.

A lot of philosophers on the contemporary scene seem to have difficulty with this notion of something other than the conscious mind being in control. Davidson, for example, refers to this kind of behavior as 'incontinent'. The point is that - whatever this agency is 'inside' - consciousness is not responsible for the action. Consciousness is responsible for learning from the consequences of the action, for 'mopping up' afterwards: for analyzing, justifying, explaining, rationalizing, and then setting up circumstances such that as you get older, you are not going to make the same mistakes again. So consciousness has a role, but it is not to initiate action. It is to prepare the self in the light of experience - and that experience must be globally integrated. That is what is going on with consciousness.

Now you can see the ambiguity with the term 'cause'. Do brains cause consciousness? Does consciousness cause brains to act? Well indirectly, certainly, the nature of actions that are taken is dressed in the form of the modifications, which have taken place in the course of learning, when people stop and reflect. But consciousness does not *initiate* action, it *withholds* action: you stop and think and consider and that is when your consciousness level goes up.

Now my question is: why is it that over the past three hundred years, everyone from Descartes to Sigmund Freud has been peddling this line that the ego is in charge? I think the answer has to do with the innovation that Descartes brought in, which is algebra, to introduce the notion of linear causality. And that was reinforced with Leibnitz and Newton by the invention of calculus, which introduced dynamics but *linear* dynamics.

I want to show an example of how most of my colleagues think about the brain in terms of linear causal relationships. Here is a schematic brain *(Figure 2)* which will show the main outline of what I am talking about. The idea is that you have motor cortex that is sending axons down into the motor nuclei and up the muscles to initiate action. And somewhere now causality plays a role, because consciousness must come into play to influence those neurons that are doing the firing. Some people put it into the frontal lobe, some into the brain stem, some into the thalamus or other location - take your choice because all of them are wrong. There is the further point that there are axons, which come down to the amygdaloid nucleus. These are supposed to add the emotional tone to your consciousness, that will shape what comes out in terms of grimaces, trembling, blushing, and so forth.

Now we cannot let it rest there. This is clearly linear causality - one action after another in sequence - but then consciousness has to be explained. The way people do that is to present a stimulus which goes into the spinal cord, then up to the thalamus and on to the sensory cortex, and this is known as information processing. This is also a linear causal chain. The big problem now is: what kind of miracle is going to occur

between the sensory cortex and consciousness - wherever it is - that is going to close the loop and put something back out again?

Well, that is the essence of the picture that most of my colleagues carry around. The problem I have with it is not so much in the physiology, which on the whole is pretty good, as with the way it is all put together as a form of linear causality - causal chains. The problem with that is: how is it that you and I - in this kind of causal chain mechanism - can make choices? How is it that we can decide to act or not to act?

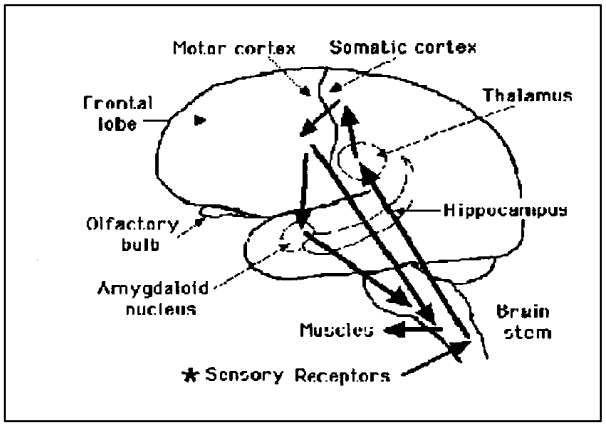

Figure 2. A schematic brain

This leads me to the kind of description which physicists are currently using to describe their complex systems, when linear causality does not make sense, and so might also apply when we come to neuron populations of the kind I have indicated, where the density of connections goes off the wall. You have the emergence of large populations of neurons and networks of populations and these have the capability of organizing themselves, and they do it with chaotic dynamics.

Now the key place to look in the nervous system for this kind of integration is very deep in the interior of the cerebral chemistry *(Figure* 3). I show this as a sketch, as though you were looking with x-ray eyes inside the brain. You will see this collection of structures which forms a ring. This is a central aspect of self-organization: you must have interaction, feedback on the grand scale. That kind of system is capable of generating patterns within itself. This is what happens when you wake up in the morning: you undergo what some call a 'state transition', in which your brain jumps from one global pattern to another one, and then to another, and another... In our recordings of brain activity, we can see these patterns jumping every tenth of a second from one state to the next. That is the basis for the emergence of action.

The typical picture which most psychologists and neurobiologists follow is that of a stimulus-response paradigm, in which the initiation of an action is by an external event of some kind, or is a combination of environmental influences and the pre-structuring of the genes which determine the connections in the nervous system. What this nature/nurture debate leaves out is the third element: self-determination. That is what neurobiologists should be concerned with. How does that take place? Basically this comes about because perception is not the passive registration that the computer model offers. Rather, it begins with the emergence in the limbic system of a search. The brain decides on its goal and it decides which kind of sensory input it wants. It sends its motor communicators out to move the hands, eyes, ears, etc. into positions for the sensory input it wants, and it sends messages at the same time to all the sensory cortexes to prime them for what it is that the brain is expecting, only to let in that which is relevant and valuable. We call that selective attention. The result is the interaction of the brain with its surroundings under its own terms - and that leads then to consciousness. The action itself is by intention, which need not be accompanied by consciousness, so consciousness is certainly not responsible for the initiation but for the mopping up afterwards.

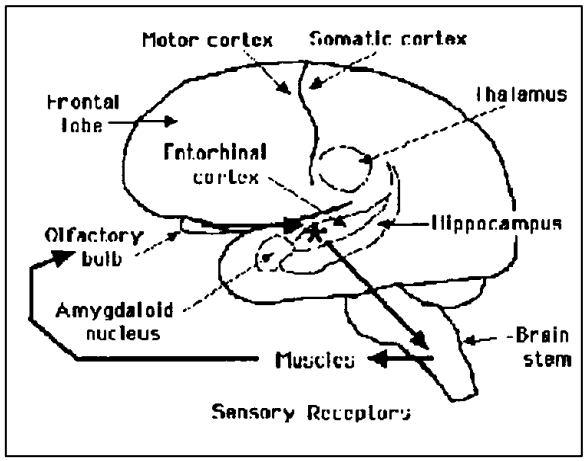

Figure 3. A schematic 'active' brain.

So now, what I propose is that most of neuroscience is based on the linear-causality model and this fact is the root of the problem. What we need to be talking about is what the physicists are calling circular causality.

Discussion

John Searle: I don't have anything to disagree with in Walter's presentation - except, of course, that he hasn't solved the problem of consciousness! He tells us that large clouds of neurons are engaged in self-organizing activity according to the mathematics of chaotic dynamics, which is different from the kind of linear causation traditional among the biological sciences - although, as he points out, these ideas are more common in contemporary physics.

But that is only one hypothesis among many held by different brain scientists. How is Walter so sure that his is the right one? What is the mechanism? How does it work? We still haven't got over the hump from neuron firings to feelings.

Walter Freeman: Well, John, I took a feather from your cap and did not include all the details. When you said you had solved the mind-body problem in five minutes, I thought maybe I should solve the consciousness problem in another five!

Searle: The beauty of the mind-body problem is that there never was a problem. It was a philosophical mistake that we got from Descartes and other thinkers. So I didn't *solve* the mind-body problem, I *dissolved* it: I showed there really was no such problem. But there *is* a really interesting scientific problem of consciousness: how does the brain do it? I have read a lot of books on this recently, and it's interesting to me as a philosopher to find that the neurobiologists hate each other as much as philosophers hate each other! Everyone has their own theory and there's a lot of hostility - professionally, it is quite interesting to see it from an outsider's point of view. But it does not look like we have a mature science of consciousness. We have a way to go yet.

Freeman: I would say that in some respects consciousness is a goal - a Holy Grail some, people have called it. My remarks were intended essentially to downgrade it to some extent, to some facet that is certainly important, but is not the main line of action. In fact, my picture is that the real key concept is intentionality. You have written extensively about this. What's at issue, I think, is a change of scene. Basically most of my colleagues adopt the picture, which the cognitive scientists have put forward for a number of years, that minds consist of collections of representations. These are thoughts, feelings, beliefs, etc. *about* the world, and the real problem with entertaining that view is how to attach the real world to those symbols. Or, *vice versa*, How do you attach labels to objects in the real world?

My view is that that concept is dead wrong. Brains do not have representations in them, they have meanings, and these meanings consist of global operators that mediate between the neurons and the objects in the outside world, which includes the muscles of the body. So directing remarks at consciousness is less productive than it would be to look at its prior state - the thing that gives it its origins - which is intentionality. One problem is that most scientists - and I think many philosophers - don't make the distinction between intent and motive. Lawyers do. They clearly understand that intent is what you are going to do, or have done, and motive is what made you do it. I think the significant neurobiological question is: how do we get the emergence of those goal-states, which then sooner or later will result in consciousness? How do you address that, John?

Searle: When I wrote my book *Intentionality* I thought that I could investigate the subject without bothering too much about consciousness, because intentional states can be either conscious or unconscious. I now think that was a mistake. I do not think you can have two independent investigations into intentionality and consciousness. In order for an agent to have intentionality at all - to have *meaning* in the extended sense which Walter is using - it has to be at least the *kind* of agent that can be conscious (even if for some reason, such as brain injury, it is not *actually* conscious at a given time). The difficulty

with saying that goal-directed behavior is what really matters is that you can have *apparently* goal-directed behavior in all sorts of situations - the digestive system, for instance - but there's not *mental* reality there. In order to get the specific mental reality of intentionality with intentional causation you have got to have the sort of agency, the sort of creature, that can be conscious.

Now Walter says - I think correctly - that in contemporary cognitive sciences there is too much of an obsession with representation, and that that is a carry over from the computer model. He is right in saying that is the wrong picture, but when he says that what we really ought to have is meanings - well, you are still back with representations, albeit of a different kind, because meanings are the mechanisms whereby the brain connects with the real world. So we are still stuck with intentionality, and with the representational character of intentionality, even if we get rid of the broad conception of the mind containing a lot of sentences in computer language. Then there is still the necessity to distinguish the real intentionality of mind from all kinds of *ersatz* intentionality that you find in evolutionary processes, or in digestion, or even in computers. And the connection, I think, comes with consciousness, so that only a system that is capable of consciousness is capable of genuine intentionality.

Freeman: The problem I see with basing this on consciousness is the difficulty of having no objective test of consciousness. This is an acute problem among clinicians, for example. They often have no way of telling whether a patient is conscious or not other than asking the person afterwards: were you aware? Now the problem here is that you cannot do that with animals. I know that John believes *his* dog is conscious, and I think that *mine* is, but I have my doubts about his!

I think that one should not downgrade the significance of this step beyond representation, because the old approach has so dominated the field, and still does for most people. They go in and look for representations of objects and features and so forth, and I think that is very stultifying and leads to all kinds of unsolvable problems. In my experience of looking for representations I was unable to find them because they were so slippery. They were continually changing. I would get these brain images - the patterns of brain activity - and, even though the stimulus was the same and the response was the same, the animal had learned something in the meantime and that had changed the pattern. So striving for an understanding of these patterns as meaning for the animal, rather than as representations of objects that the investigator has deter- mined, seems to me a very important step for researchers to take.

This is why I keep emphasizing that representations exist in objects but not in brains. On the other hand, meanings exist in brains but objects don't have meanings. So when we go into the study of what corresponds to, or correlates with, consciousness, we do not look for 'consciousness neurons' or for 'hot spots' in the frontal lobe, like a lot of people do with brain imaging. Consciousness is a style of organization that involves an entire hemisphere.

Searle: The point about representation is, I think, most obvious in the case of memory. A very traditional view of memory is of a kind of storehouse of mental representations. You want to remember Herbert Hoover, so you go back and get the Herbert Hoover dossier out of your storehouse of memory. That is clearly a misconception of memory, which we now understand to play a much more creative, active part. You don't react to an existing store: you create a present memory-experience on the basis of strengthened neural connections or whatever it is in the brain. That much is clearly right.

But the representing task of the brain is essential for our survival, and that is what I am calling intentionality. I don't see that any mileage is gained by saying, 'Well, it is meanings, not representations,' because meanings do represent. The only point is that you need to get rid of the idea - and this we are agreed about - that representations consist of a set of sentences or other symbolic devices stored in the brain. But that consciousness serves to relate the organism to the environment, in a way that enables the organism to have information about the environment and to act on the environment, that seems, to me, essential.

The point about language is that there is one level of description of words at which they are just phonetic sequences, but there is another level of description at which they are perfectly meaningful English words. If they were not, you would not be able to understand me.

Freeman: That is part of the problem. When you utter the words, people get an image in their mind - essentially expressing something they already know. This has been my difficulty in trying to convey that when I look at these brain images of animals; they don't correspond to the representations. I thoroughly agree with you there must be some correspondence between what is going on inside a person's head and the outside world - otherwise you would fail in the world, suffering one disaster after another. The question is: what is the nature of these neural organizations? There is some correspondence, and neurons are doing it, but how? And here, I think the approach to that question with the vast array of techniques that we have in our hands now is essentially being wrongly done. People are using the devices for brain-imaging - looking for the red spot which potentially shows where consciousness lies and where meaning lies, where the dancing-lessons are located, or the French lessons. You agree that is nonsense, but how *do* we go about it? I think that we have to redefine our language and go for some new terms, or at least get some terms that are not as loaded as 'representation' is.

Searle: You mentioned earlier that there is always a problem about knowing whether something (or someone) is conscious - unless you can ask them - and what is the nature of the conscious state they are in. I think this is often used as an excuse in neurobiology for not investigating consciousness, and I regard it as a serious mistake. I see this just like any other epistemic problem. Use your ingenuity. Try every weapon that comes to hand - anything that works. This is not the only time in the history of science that we have had unobservable phenomena crying out for investigation. So my feeling would be: yes, it is sometimes difficult to find out whether a patient is conscious and indeed, it is often difficult to find out what is going on in the conscious state - but that is a challenge and not an insuperable obstacle. It is just something we have to overcome, and indeed, any neurobiological textbook does just that, trying to figure out what sort of symptoms the patient is having.

Freeman: It is easy enough to say that, but it still leaves the question of how to relate the brain-pattern to whatever it is going on inside the person's head. We really don't know. The same holds for artificial intelligence, if indeed we are ever able to achieve what is still not a very clear goal. It is something beyond smart washing machines, but it's not clear what. I recall a discussion on this between a psychologist and an engineer, and the engineer asked the psychologist if he could name anything that the brain could do that a machine could not do...

Searle: I think that question is based on a traditional philosophical mistake, and it is a mistake that has been an obstacle to research into consciousness. The assumption of a contrast between the brain and a machine already contains the elements of a mistake, because the brain *is* a kind of biological machine. So the question, 'Can a machine be conscious?' has already been answered: we have such a machine inside our skulls. Then it becomes a fruitful question whether or not we can duplicate - not just simulate but duplicate - artificially the causal mechanisms that produce consciousness. That is then a well-defined question, rather than this vague philosophical idea that there is some obstacle to a machine ever being conscious. We are precisely conscious machines.

Now the paradox is that, although the machine you buy in the store is indeed an electronic machine, the actual processes that the computer undergoes are not, as such, machine processes. They are abstract mathematical processes that we implement on the machine. So there's a certain irony in the idea that the computer is a machine but the brain is not a machine. If you consider the processes involved, the *essence* of the brain is a machine whereas the computer is a machine that happens to implement a set of processes that are not machine processes - they are abstract mathematical processes.

Freeman: Certainly a digital computer has severe limitations working with numbers because it cannot do chaos. And chaotic dynamics is one of the central features of the operation of the brain, because this is where novel patterns can come about. We all talk about cause-and-effect and trial-and-error, but where do these trials come from, and how do they cause errors? For that it is necessary to look at the dynamical system, which is never going to operate twice the same. That is taking us pretty far afield from existing machines, but it certainly points us in the direction of making such a device. I have no wish to see such a

thing made - I should prefer people to try and fail - but I think it is important to get a better comprehension of how consciousness emerges in our brains and what kind of role it plays.

I see the real root of the difficulty in this business of causation. John keeps coming back to the idea of brains causing things to happen. That is where I have a problem, because the definition of causal relations in terms of circular causality - where you don't have a linear cause-and-effect relationship - really makes a major change in the thought processes, which go not just into building a device but rather into understanding the one we now have (in our head).

Searle: The problem is not with the notion of causation but with a mistaken set of ideas about causation, which mostly come from David Hume. He said that causation precisely consists in sequences of regular events discretely organized in time. In my view that is *one* kind of causation but there are all kinds of other forms of causation. For instance, this table supports the glass standing on it - and in that sense causes it not to fall to the floor - but that is not linear causality. It has to do with the molecular structure of the two solid objects. Once you see that causation is a matter of how the world works, what makes things happen, what is responsible for something happening, then you see that this linear Humean form of causation is only one kind. There are lots of others. What we want to know from a scientific study of nature is 'how does it work?' And that's what we want from a science of consciousness.

Questions

What is being done to draw on other cultures, such as the eastern traditions, in developing an understanding of consciousness?

Searle: I happen to have been on the same platform as the Dalai Lama when I gave a lecture on precisely this topic in Bombay, and to my horror I found that he made a lot of mistakes that I thought were peculiar to *our* intellectual tradition. His view sounded to me very much like dualism: he began his lecture by saying, 'Each of us is both a mind and a body,' and I gave a shudder. It ought to be comforting to us to know that the mistakes we make are widespread, they are not absolutely universal. I have an African friend who tells me that in his native African language, you cannot even *state* the Western mind-body problem. Now, I love that. That's what I'm trying to do: to get our language into that advantageous state.

My answer to your question is to use any weapon that comes to hand. Some people say, 'Well what do we care about these mystics and so on?' but I'm with William James on this one: use any information that comes to hand and stick with anything that works. So if there is something that we can learn from eastern mystics, or people that undergo all kinds of sensory deprivation, or hold completely different worldviews, then that's fine. Let's use what we can from them.

Freeman: I thoroughly agree with what John has just said. I think the whole area of altered states of consciousness and the variations that occur in worldview are extremely important to the understanding of the nature of consciousness. So people with other philosophies, oriental or otherwise, are very welcome at the table.

If a digital computer told you it was conscious, how would you decide whether it was lying?

Searle: It is very easy to program a digital computer so that if you ask it, 'Are you conscious?' it will answer 'Yes.' To my amazement, when I did the Chinese room argument, a whole lot of people said that a simple proof would be to ask the question in Chinese, 'Do you understand Chinese?' Well, I could give back a symbol that means 'I understand Chinese', but that seems to me quite useless and it shows nothing one way or the other. The question, 'Are you conscious?' addressed to a computer is irrelevant to deciding whether or not the computer is conscious.

So why is it relevant to a human being - say to one who appears to be coming out of a coma? Because in that case, we have a whole lot of background information, and the question is asked against that background. This is a human being exactly like us, who has a similar causal basis of consciousness to me. Now this goes for my dog, Ludwig. Why am I so confident that Ludwig is conscious? Is it just his behavior?

Well, we could make a robot that would rush out and wag its tail and jump all over me, so it is not just behavior- this is a big mistake we make, thinking it is just behavior that gives the grounds for attributing consciousness. Rather, it is behavior and causal structure. The reason I am absolutely confidant that Ludwig is conscious is that I can see he has mechanisms relatively like mine - there are his eyes, there are his ears. That's why I don't know what to say about grasshoppers and termites. I don't see the relevant similar causal structure.

So the basis on which we make the attribution of consciousness to others is not 'intelligent behavior, therefore consciousness' or 'answers "yes" to the question, "Are you conscious?"', but the acceptance of a certain relevant similarity in causal structure.

Freeman: If someone asks a digital computer, 'Do you have consciousness?' and the computer says, 'Yes - do you? - Prove it,' then that person is in difficulties.

Searle: I don't have to worry about whether I am conscious. If a digital computer - or indeed a roomful of neurobiologists - looked at me and said, 'You know, this guy is not conscious. He may look it but he isn't really,' I'm not going to worry about it and say, 'Gosh, perhaps they're right...

If we accept that consciousness is something to do with organization, why not apply this to the Chinese room? You are only a part of the set-up, so even though you do not understand Chinese, perhaps the whole thing does. I understand English, but no individual neuron in my brain divided up understands English.

Searle: This question, as you probably know, is one that is commonly raised about the Chinese room. The idea that *I* don't understand Chinese but the *whole room* does will not work for the following reason. In the thought experiment, I get the right Chinese answers to the Chinese questions because I know the rules for manipulating the symbols. I don't understand Chinese because I don't know the meanings of the words represented by the symbols. I have no way of getting from the syntax or symbols to the actual semantics or mental content. And if I have no way of getting from syntax to semantics then neither does the whole room. The argument is not answered by saying there is a difference between a single element and the whole system; of course there is, and I am not confusing them. The argument is that the syntax of the formal program is not sufficient to guarantee the semantics, the actual mental content of Chinese understanding.

In looking for an explanation of consciousness, how would we know when we had found it?

Searle: That is a beautiful question. I think in the end it would be like any other part of science. How do we know that we have an adequate theory of some disease, say? Well, for one thing, we get control. If you learn that to introduce a virus brings about the disease-system, and to remove it cures the disease, that confirms the theory. Let us suppose that, contrary to what Walter and I believe there was some simple neurobio-logical mechanism that was responsible for consciousness - call it *xyz*. Now suppose that by producing *xyz* in an otherwise unconscious patient, you could get the patient to consciousness; and that by suppressing *xyz* you could induce unconsciousness. That would provide the confirmation. If you are worried about how you would really know, you could take the place of the patient yourself.

Freeman: I think it is true that you have the opportunity here to develop what we don't yet have, which is an objective test for consciousness. There is already a real problem with people in what is called a vegetative state. There was a study just a couple of years ago of forty people in Los Angeles hospitals who were allegedly 'vegetables' and it turned out that seventeen of them - 43% - showed signs of consciousness. They were actually aware. We know also of so-called 'locked-in syndrome' - people who have severe damage from car accidents, for example - people who are totally paralyzed and yet they are aware, they know what is going on, although they are 'locked in' and unable to express it. This is a devastating kind of condition, as you might well imagine.

Now there is clear evidence from electrographic recordings of brain activity that awareness of consciousness is not localized in any one part of the body. Interestingly there is evidence from binocular

rivalry, where you can see two different images, one with each eye, and the two alternate back and forth. It used to be thought that this was gating at the thalamus, then that it was at the visual cortex. Now it is becoming apparent that the entire hemisphere becomes coordinated, becomes organized into a global pattern that supports the visualization, the experience of seeing one or other image, and that this flows back and forth between the two hemispheres. I think that if we pursue this further in the scientific realms of studying consciousness, we shall see some amazing things. Whether we can go beyond that to control - well, that is a marvelous goal, but let's start small.

Could you say something about how phenomenology relates to the mind-body problem of consciousness?

Searle: Much of our model of scientific investigation is that we think of science as reductionist, so the temptation is to think that we shall not have a scientific account of consciousness until we can get a reduction of consciousness to something else. Now I am extremely suspicious of that model and I read some of the literature on reductionism. The difficulty is that there are at least half a dozen different definitions of reduction available in the scientific and philosophy-of-science literature today. My own view - putting it in one sentence - is that we shall get a causal reduction of consciousness but we won't get an eliminative reduction. We won't get *rid* of consciousness.

Roughly speaking, reductionists divide into two kinds. The first gets rid of phenomena by showing that they never existed, so sunsets, for instance, got reduced to the rotation of the earth. The second kind explains phenomena in terms of a lower level without eliminating them, so, for example, we did not get rid of solidity when we reduced it to molecular structure. Now I want to say that you cannot do a reduction of *either* of these kinds on consciousness, because of the first-person subjective character of consciousness. You will get a causal explanation in terms of the neurobiological process but you will not show that there is nothing there but the neurobiological process because there is something more - the subjective feeling. There are various features - consciousness is not the only one - where you don't get a reduction according to the standard models.

Now, about phenomenology. It is the name of a movement in philosophy that tries to examine the character of our experience in order to produce grounding for our knowledge that will help us to understand how we relate to the world. My impression from the phenomenology that I've read is that most of it is pretty sloppy - but that is true of most things that are published, so I cannot fault it on that alone and my attitude to it is the same as with everything else: if it's useful, use it.

Freeman: I interpret phenomenology as the report of individual experience. I cannot really distinguish it from introspection, but as personal experience is what we are trying to explain, don't throw it out.

We seem not to have got away from the logical premise that one can reproduce consciousness with a computer. I would disagree with Professor Freeman about not being able to have a chaotic system in a computer if we include analogue as well as digital computers.

Freeman: A digital computer is limited by its use of rational numbers. An analogue device - which I would not call a computer - is working with continuous variables and that has the feasibility of reproducing some of the phenomena which one should like to put into a machine, which one would hope was going to show some of the phenomena and self-organization that we should like to observe. It just hasn't been done yet. I wouldn't say that it couldn't be done. I would have some trepidation about seeing it done, but that's essentially an ethical question. It gets back to another aspect of the word consciousness, which relates to conscience and means 'knowing together'. This gets to the social aspect of the problem, which I think should not be neglected.

Searle: I am grateful for this question because it enables me to make my position *vis-a-vis* computation absolutely clear. Here are the questions people ask: (1) Could a machine be conscious? The answer is trivially 'yes', because our brains are machines and they are conscious. (2) Could you build an artificial machine that was conscious? In principle, 'yes': if you knew how the brain did it you could duplicate it artificially. (3) Could a computer be conscious? Well, under some descriptions just about anything is a

computer, and my brain can compute: '2+2 = 4'. That is a computation so you just listened to a conscious computer. So trivially again, the answer is 'yes': computers can be conscious because my brain is a computer and I am conscious. But the *crucial* question is not any of these.

The crucial question is this: Is computation (as defined by Alan Turing in his definition of a Turing machine) by itself sufficient to guarantee the presence of consciousness? And the answer to *that* question is 'no'. Why? Because computation *as defined in the literature of computational science* is the abstract manipulation of formal symbols - usually thought of as 0s and 1s - and that by itself is insufficient to guarantee the presence of consciousness. That is what the Chinese room shows.

Now the Chinese room does not show that God could not decide to endow all Macintosh computers with consciousness! I think it extremely unlikely, but that is not my problem. I am just dealing with logical relations and, logically speaking; the formal syntax of the computer program is insufficient by itself to guarantee consciousness. What is lacking? What is lacking is *causation.* The brain is above all a biological - causal mechanism like the heart or the liver. Computation as defined in the literature does not mean a causal process.

The relation of a computer simulation of the brain to the brain is exactly the same as the relation between a computer simulation of the stomach and the stomach. You don't say, 'We've got the stomach program running on the computer, so stuff a pizza into the computer'! What the computer does is to give you a model or picture of a stomach, not a stomach, and that is what it does also for the brain.

The Dalai Lama has always emphasized as the essence of consciousness its spontaneity*, so should we be thinking in terms of a causal chain at all? Modern physics would seem to suggest that states of matter do not determine what happens next but only give a range of options. So perhaps what consciousness does is to take advantage of the uncaused possibilities?*

Searle: This is the kind of question where we're being asked to talk about quantum indeterminacy, the freedom of the will *and* the nature of consciousness - so let's do it. The problem you've put is, I think, one of the leading unsolved problems in our life, and I don't know how to solve it.

We are all aware of our own freedom. We are all aware of the gap between the causes of our actions and our actions. Walter says this doesn't work for raising your hand, but what about deciding whether to go back to California or to move to Alaska, or to run for office, or whatever? For those kinds of things you can't just look at a monitor and say, 'Ah yes, there's a neuron firing...' sometimes you *do* have this gap and this gap has a name: it is freedom, or freedom of the will. The difficulty is that there aren't any gaps in the brain. When you look at the brain you find the chaotic dynamics, or a sequence of neuron-firings, or whatever you like, but you don't find this kind of 'gappiness'.

Now a lot of people say the answer to all of this is in the quantum indeterminacy at a much lower level than that of neurons. But I have never been able to make the connection between quantum indeterminacy and the freedom of the will. It just seems to me that the statistical indeterminacy at the quantum level is irrelevant to the position of human freedom. So I don't know the solution to the problem, but I can point to the nature of the problem.

Our free will is ineliminable in that we cannot even *think* it away. You go into a restaurant and they give you the menu and you have to decide between the veal and the steak. You cannot say to the waiter, 'Look, I'm a determinist. *Que' sera sera,'* because even doing that is an exercise of freedom. There is no escaping the necessity of exercising your own free choice.

Freeman: I have one final comment to make on causation. You are looking for what is causing your behavior: that is essentially an attribute of how humans look at the world, how we understand the world. We act as our cause and we get an effect as a result. Then we extrapolate that to the outside world around us. This is where we can get into deep trouble, because we reason away our own freedom. That is absurd. It is the ultimate *reductio ad absurdum.*