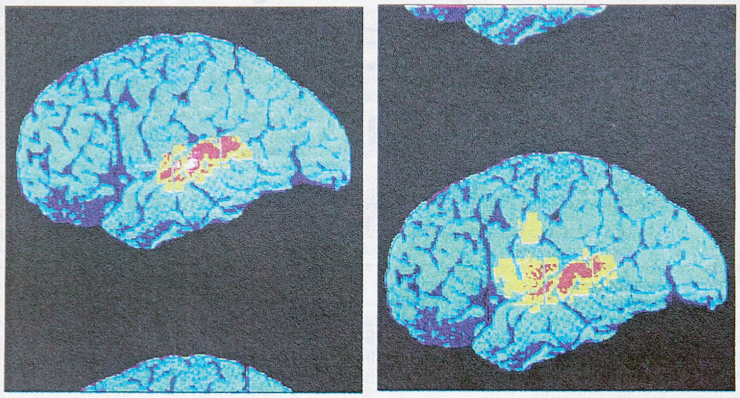
THE TIMES HIGHER APRIL 5 1996

Does consciousness emerge from quantum processes?

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

SLP



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

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.

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.

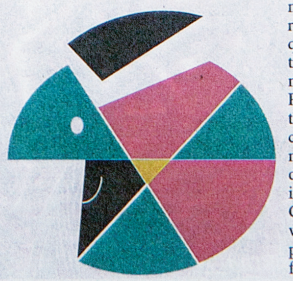
Microtubules are crystal-like protein cylinders that perform varieties of cellular chores within nerve cells including forming and regulating synaptic connections. The individual protein subunits ("tubulins") which make up the microtubules can switch between different configurations, known as "conformations", 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.

THE TIMES HIGHER
APRIL 5, 1996

Does consciousness emerge from quantum processes?

Stuart Hameroff and **Roger Penrose** 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 "proto-consciousness" (a "funda-men-tality") 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 modem thinkers argue that what makes up the universe is fundamental information with experiential properties. Perhaps neurobiological systems somehow access and organize 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 Schrodinger'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-organizing OR process could select the individual space-time geometry of experience.

Could self-organizing 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.

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.

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.

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 paralyzed, 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 Godel 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 Godel result implies for brain function. Additionally, the link between conscious experiences such as smelling cinnamon and the Godel 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 signaling 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 instandy 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.

Stuart Hameroff, MD, is a practicing clinical anesthetist and professor at the University of Arizona.
Sir Roger Penrose is Rouse Ball professor of mathematics, University of Oxford.
Patricia Churchland is director of the Experimental Philosophy Lab, University of California, San Diego.