Quantum consciousness

Nothing in science is as mysterious as quantum mechanics—except, perhaps, the mechanics of the mind. Now genius-of-all-trades Roger Penrose says the two are intimately connected.

By David H. Freedman

PHOTOGRAPHS

by Michael Llewellyn

IN A SMALL LECTURE ROOM AT Penn State, Oxford mathematician and physicist Roger Penrose lectures to a packed crowd of colleagues and students. The soft-spoken 62-year-old professor is one of the world's leading experts on general relativity and quantum mechanics, the two complex theories that explain just about everything that happens in our universe. His talk today, however, appears to concern a small bundle of tubes, which he hastily sketches on the blackboard. ▪ The bundle could be any one of a dozen different exotic entities from Penrose's menagerie of mathematical and physical objects—"supertwistors," "wormholes," and "worldtubes"—all space-time oddities now second nature to his audience.

Yet this particular bundle turns out to be an unfamiliar one to his listeners, even though it is, in fact, far more mundane. It is an arrangement of protein structures found in all living cells. According to Penrose, these structures could play a very special role in the universe: they may enable the brain, which is essentially a clump of the same sort of matter that makes up rocks and stars, to generate the mind, that intangible, unbounded entity that provides us with an inner voice, imagination, emotions, thought, and our very sense of self.

Although the subject may seem a little far afield for this gathering, Penrose contends it is well within the group's purview. The quest for the ultimate laws of nature has taken physicists to such wondrous locales as the interiors of massive black holes and the unimaginably small islets of matter conjured up in particle accelerators. Penrose maintains that the trail ultimately snakes closer to home, right through the three and a half pounds of grayish goo jiggling in our skulls. To understand the mind, he says, you need new physics—and, almost paradoxically, uncovering this new physics may very well depend on new conceptions of mind.

Penrose first advanced the argument for a deep, if somewhat vague, connection between the mind and physics in his 1989 surprise best-seller, *The Emperor's New Mind*. In that book he suggested that consciousness is created by some mysterious quantum mechanical phenomenon that takes place in brain cells. Unfortunately, brain cells seem an improbable locale for quantum mechanical antics. The well-known weirdness of quantum behavior appears almost exclusively in isolated subatomic particles, and it easily becomes masked in large and crowded systems of atoms, such as exist in ordinary matter—and cells. At the time, Penrose was unable to provide any hints as to how that conflict might be resolved. But during the past year he has found a way. Penrose can now point to a component of brain cells that appears to be an ideal conduit for quantum mechanical phenomena. That component, known as a microtubule, is Penrose's nominee for the physical root of consciousness.

Surprisingly, Penrose's insight about microtubules was inspired not by an article in one of the leading journals of neuroscience but through an out-of-the-blue encounter with a free-thinking Tucson, Arizona, anesthesiologist named Stuart Hameroff. Though the colorful if obscure Hameroff and the reserved but celebrated Penrose make an unlikely team, the partnership has produced what may be the most explicit theory of the physical basis of consciousness ever put forth—a theory that, if proved right, could shake up fields as diverse as physics, biology, computer science, and philosophy.

Penrose is a self-described "dabbler." He says it's a trait he picked up from his father, a doctor who actively pursued interests in psychology, philosophy, and mathematical puzzles. Unable or unwilling to keep his mind focused on what other people might consider more than sufficient material for a single career, Penrose has continued to collect new specialties for the past several decades. After receiving his Ph.D. in mathematics from Cambridge University in 1957, he briefly took up computer science, moved back into physics, and came to focus first on quantum mechanics and then on general relativity during faculty stints at Princeton and Syracuse universities before signing on at Oxford.

Along the way Penrose came to think about the mind. In particular, he wondered whether or not a computer could be programmed to acquire something akin to consciousness. Artificial-intelligence researchers have already created programs that seem to capture at least the flavor of all of the mind's unconscious activities, including the work of the five senses, muscle control, and instinct. Such programs allow robots to find and pick up blocks, computers to answer questions about auto repair, and cartoonlike "artificial life" creatures to mate, find food, and otherwise live out their lives on a video screen. However, researchers haven't a clue about how to get a computer to intuitively assess the truth in a subtle argument or see the humor in a joke, to feel the emotional impact of music, philosophize about the meaning of life, or come up with counterintuitive solutions to unfamiliar problems. In short, they have no idea how to invest a computer with those aspects of mind that seem clearly conscious—that allowed Descartes to declare, "*Cogito, ergo sum.*"

Why are such processes so elusive? It may just be, as most artificial-intelligence researchers assert, that simulating them on a computer requires programs far more complex than any yet devised. But Penrose finds this explanation unsatisfactory. For one thing, research suggests that most brain cells are preoccupied with such unconscious tasks as processing and storing images and controlling muscles, and that only relatively small portions of the brain are dedicated to the sorts of tasks we associate with conscious thought. Such evidence runs counter to the notion that consciousness emerges from a more complex version of the same sorts of brain processes that give rise to unconscious thought; if it did, one might expect it to account for the lion's share of brain matter.

Besides, if consciousness were no more than a program—even a horrendously complex one—why wouldn't artificial-life researchers or neuroscientists have gained at least a tiny insight into its nature? The reason, Penrose concluded, is that the "quality of understanding and feeling possessed by human beings is not something that can be simulated computationally"; that is, it simply cannot be broken down into a series of steps, a sort of recipe, that when followed on a computer will result in a reasonable imitation of the real thing.

The notion of noncomputable processes is not unfamiliar to mathematicians and computer scientists. One particularly well-known and striking example of such a process comes from the mathematics of tiling, which concerns the ways in which different sets of flat shapes, or tiles, can or can't be arranged to cover an infinite flat surface without leaving gaps. That certain shapes—squares or triangles or hexagons—can do so seems intuitively obvious. But curiously, mathematicians have proved that it's impossible to devise a computer program—a general set of rules—that can predict whether tiles of any given shape can completely cover a plane. (Penrose himself has explored this problem, and out of his investigations, in 1973, he discovered a pair of diamond-shaped tiles that could completely cover a surface, but only in an infinite variety of never-repeating patterns.)

ROGER PENROSE

The Oxford mathematician and physicist has "dabbled" in everything from tiling problems to wormholes to other universes: now he's zeroing in on the quantum mechanics of the mind.

If the question of whether certain tiles can cover a floor is noncomputable,

> PERHAPS NERVE SIGNALS START OFF IN A QUANTUM MECHANICAL MISHMASH OF STATES THAT ALLOWS FOR THE SIMULTANEOUS EXISTENCE OF COUNTLESS BILLIONS OF DIFFERENT PATTERNS.

then might not the task of evaluating an object's beauty as well as any other chore of consciousness be the same? Penrose was sure they were. But if consciousness is noncomputable, then whatever process in the brain that gives rise to consciousness must also be noncomputable. This conclusion has an unsettling and inevitable implication: presumably, whatever happens in the brain obeys the laws of physics, and if one is going to keep religion and metaphysics out of the picture, all the known laws of physics are computable. According to these laws, every physical process in the universe—from atomic collisions to galactic collisions—can be flawlessly simulated, at least in principle, on a computer. That being the case, Penrose decided the brain must incorporate a physical process that simply isn't covered by the known laws of physics. Consciousness, he concluded, is rooted in new physics—that is, in laws not yet discovered or formulated. Furthermore, he thought he knew where to look for them: in the weird underworld of quantum mechanics.

Quantum mechanics is an unrelentingly strange theory. Among other things, it tells us that an electron or another denizen of the subatomic world tends to exist in a multitude of states all at once: it is simultaneously here and there, moving fast and slowly, spinning one way and the other. But at the moment the electron interacts with ordinary matter or energy—when it smacks into the molecules in a detector, for example, or is bombarded by a beam of light—the disturbance somehow causes the electron to "choose" a single state. At that point it behaves exactly as one would expect a minuscule billiard ball to behave. Real-life billiard balls never exhibit multiple personalities because quantum mechanical weirdness is generally apparent only in objects that are roughly the size of an atom or smaller and that exist in "quiet" environments, isolated from the random jostling of other particles and forces.

Every measurement ever made has supported this bizarre picture, so physicists regard it as gospel. Yet there is no widely accepted explanation of *how* a particle happens to choose a single state when it is disturbed. As far as most physicists are concerned, it just does. Of course, they wouldn't mind if a good explanation just happened to emerge, even from as unlikely a place as the study of consciousness. But in all probability, any such explanation would also require a fundamental change in our understanding of quantum mechanics—just as Einstein's general theory of relativity necessitated a fundamental change in our understanding of gravitation. And even a small change could be a shocking development for a theory that has withstood so many rigorous tests.

Penrose has always been uncomfortable with this gap in the quantum mechanic's view of the world and has long argued that the theory must be modified to account precisely for the process of choosing a single state. But in thinking about the need for a noncomputable physical process in the brain, it occurred to him that he might be able to kill, or at least graze, two birds with one stone. What if the process of quantum mechanically choosing a state was noncomputable? Then it would be a viable candidate for the physical process that gives rise to consciousness. "Sometimes," he says, "the dabblings seem to come together for no clear reason."

According to standard neuroscience, the brain processes information through the pattern of electrical impulses that the brain's nerve cells, or neurons, swap with one another. Perhaps, mused Penrose, these signals start off in a quantum mechanical mishmash of states that allows the simultaneous existence of countless billions of different patterns; out of this quantum mechanical mix, one pattern turns up that fulfills the task at hand—it "clicks"—and that's the pattern that becomes a conscious thought.

Penrose does not address exactly how the brain knows just when a solution is "right." It doesn't, of course; but somehow, amid the buzz of unconscious activity that's always going on behind the scenes, occasionally a thought or inspiration or feeling emerges from the general background noise and pushes into our conscious awareness. Just why that state gets "picked" as opposed to some other may have to do with some sort of match between patterns in the mind. Beyond that, for now at least, no one can say.

This quantum mechanical choosing among many states is the model for consciousness that Penrose presented in *The Emperor's New Mind*. (The argument actually appears in the last few pages of the book; the rest of the work is a brilliant primer on the physics, information science, neurobiology, and other fields needed to grasp the argument.) The book raised a maelstrom of dissension from virtually every side of the issue. Artificial-intelligence researchers fumed that Penrose's elegant discourse on noncomputability failed to provide convincing evidence that consciousness is noncomputable. So what if intuition, insight, and self-awareness seem mysterious to us now? Forty-five years ago, they said, we didn't even have computers, and today they teach children and play chess at the grand master level: who knows what computers will be thinking 45 years from now? As for physicists, the idea of modifying quantum mechanics simply because the process of choosing one state out of multiple states seems arbitrary is one that attracts few admirers.

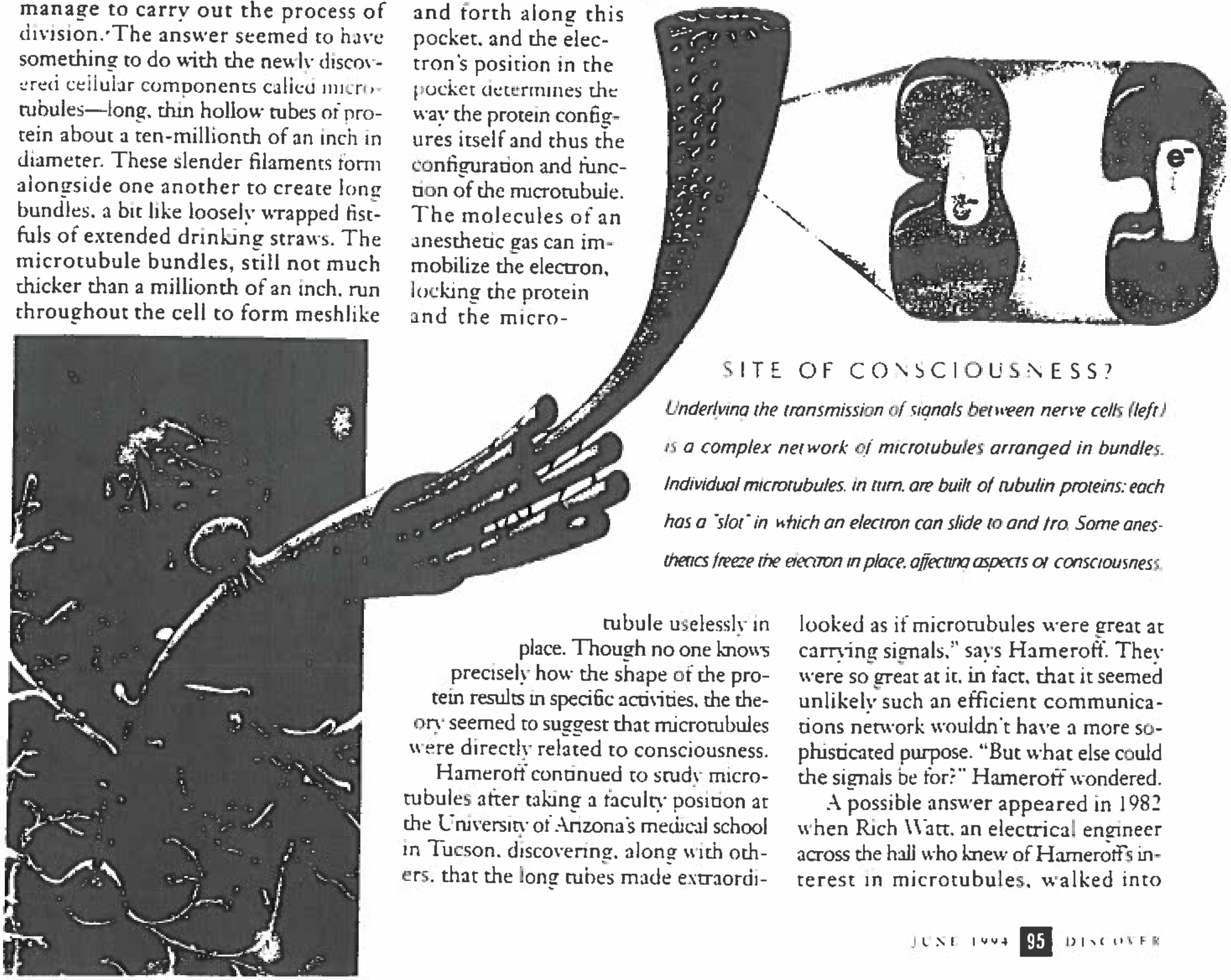
In early 1992 Penrose set out to write a follow-up book that would support his arguments. But the more

John Karapelou

progress he made, the more he felt nagged by one question: How could nerve impulses—the packets of electrical energy that neurons swap among themselves when they fire—be quantum mechanical? Nerve impulses are flying around in one of the noisiest environments imaginable: the brain is a dense structure of cells bustling with chemical and electrical activity. Yet it's exactly this kind of interaction with surrounding matter and energy that tends to drown out quantum mechanical behavior. "I was entirely uncomfortable with the idea that neuron firing could be a quantum event," says Penrose. But where else besides nerve impulses could quantum mechanics play a meaningful role in thought?

As a student at Hahnemann Medical School in Philadelphia in the early 1970s, Stuart Hameroff was captivated by the question of how cells manage to carry out the process of division. The answer seemed to have something to do with the newly discovered cellular components called microtubules—long, thin hollow tubes of protein about a ten-millionth of an inch in diameter. These slender filaments form alongside one another to create long bundles, a bit like loosely wrapped fistfuls of extended drinking straws. The microtubule bundles, still not much thicker than a millionth of an inch, run throughout the cell to form meshlike networks. These microtubule networks serve as a sort of skeleton to the cell, lending it structure and creating pathways for the transport of chemicals within the cell. But most intriguingly, when a cell is about to divide, the bundles dissolve; then the microtubules reform in new configurations that pull the cell apart in exactly the right place. They behave like cellular traffic cops, directing the complex process of division.

Hameroff also had an interest in the mysteries of consciousness and intelligence, and when he began his internship at Tucson Medical Center, he leaned toward specializing in neurology. But a colleague in anesthesiology wooed him away by telling him of a curious finding: gaseous anesthetics like ether or halothane, which can "turn off" consciousness without otherwise significantly impairing brain function, appeared to work by somehow temporarily crippling the microtubules in neurons. The protein molecule that makes up a microtubule has a sort of pocket along its length; a single electron can slide back and forth along this pocket, and the electron's position in the pocket determines the way the protein configures itself and thus the configuration and function of the microtubule. The molecules of an anesthetic gas can immobilize the electron, locking the protein and the microtubule uselessly in place. Though no one knows precisely how the shape of the protein results in specific activities, the theory seemed to suggest that microtubules were directly related to consciousness.



SITE OF CONSCIOUSNESS?

Underlying the transmission of signals between nerve cells (left) is a complex network of microtubules arranged in bundles. Individual microtubules, in turn, are built of tubulin proteins; each has a "slot" in which an electron can slide to and fro. Some anesthetics freeze the electron in place, affecting aspects of consciousness.

John Karapelou

Hameroff continued to study microtubules after taking a faculty position at the University of Arizona's medical school in Tucson, discovering, along with others, that the long tubes made extraordinarily good conductors of physical vibrations, or sound waves. Using computer models to simulate microtubule behavior, he found that a vibration introduced into one end of the tube could propagate unchanged throughout the length of its hollow, water-filled interior. Furthermore, he found that disturbances in neighboring microtubules displayed a "coherence"—that is, a vibration in one microtubule could start another one vibrating in exactly the same fashion, just as a vibrating tuning fork can get a nearby tuning fork vibrating. Eventually, a twitch propagating down one microtubule could be passed on to entire bundles of microtubules vibrating in synchrony, and perhaps even right through cell membranes to microtubules in neighboring cells.

Hameroff suspected that this property was related to the microtubules' traffic-cop function in the cell: if they were in charge of organizing behavior in the cell, then they'd need to communicate with one another, and for that they'd need an accurate, rapid signaling system. "It looked as if microtubules were great at carrying signals," says Hameroff. They were so great at it, in fact, that it seemed unlikely such an efficient communications network wouldn't have a more sophisticated purpose. "But what else could the signals be for?" Hameroff wondered.

A possible answer appeared in 1982 when Rich Watt, an electrical engineer across the hall who knew of Hameroff's interest in microtubules, walked into

Hameroff's office and showed him a photograph taken through an electron microscope. "What is it?" asked Watt. "It looks like a microtubule," answered Hameroff instantly. "Look again," said Watt. The photograph was actually of one of the microelectronic switches that make up computer chips. Hameroff realized what had been lurking in the back of his mind for some time: microtubules made up some sort of information-processing network.

Hameroff spent much of the next ten years developing a theory of how this signal-carrying property of microtubules could enable a network of them to serve as a computer within a brain cell. All a computer chip does is carry electric pulses around a railroad-yard network of pathways interconnected by transistors serving as tiny switches. That's one way the brain functions, too, with neurons serving as the switches. But Hameroff strongly suspected that microtubule networks in the cell could also play the role of such a switching yard, directing vibrational pulses along certain paths within the cell and also among cells; by arranging and interconnecting themselves in the right ways, they'd direct vibrational signals here and there just as wires direct the flow of electric signals. Since such a network could exist within a single brain cell—a brain cell that was itself part of a computing network—a microtubule network would in effect be a computer within a computer.

STUART HAMEROFF

The University of Arizona anesthesiologist found a quantum mechanical structure in the brain, but he lacked a theory of consciousness to apply it to; and then he met Roger Penrose.

If the brain's network of neurons is itself a computer, why would it need to have each individual neuron serve as its own computer? According to Hameroff, the brain's conventional neural network alone is simply far too underpowered even to account for such tasks as a person's ability to walk into a room and instantly recognize every object in it. Though this may seem entirely unremarkable to us, it is in fact a near-miraculous feat of information processing; even a dozen of the world's largest supercomputers couldn't come close to replicating it. Neurons acting as relatively simple switches couldn't possibly provide this level of computing, Hameroff contended. The additional brainpower, he claimed, is provided *within* each neuron. At this point, Hameroff wasn't even concerned with consciousness, per se. He was simply impressed that the brain seemed to require more computing power than neurons could provide, and microtubules signaling each other through sympathetic vibration seemed to offer a plausible additional mechanism.

When the brain is attempting to solve a problem, according to Hameroff, it handles the enormous amount of processing required on two levels: at the level of microtubules swapping vibrational signals, and at the level of whole neurons swapping electric signals. Hameroff proposed several mechanisms for linking these two somewhat distinct processes. Microtubule signaling could provide a lower level of processing that determines when neurons will fire as the last stage of processing, for example. Or the two processes could work in concert, trading signals back and forth in a continuous dialogue. Or microtubules could pass signals from neuron to neuron in *response* to neuronal firing, providing a scheme for fine-tuning the firing patterns, much as an artillery spotter tells the soldiers at the guns how to adjust their aim after each shot.

Hameroff's theory, innovative as it was, did not get much attention from the neuroscience mainstream. It didn't help that his enthusiasm and imagination led him to suggest in talks and papers that microtubules could be artificially grown and harnessed to create "nanorobots" programmed to perform medical services in the bloodstream of patients, or to form gigantic artificial brains placed in orbit around Earth.

Meanwhile, research in the late 1980s by biophysicists started to produce intriguing hints that microtubules, by virtue of their tiny dimensions and tubular form, had some unique quantum mechanical properties. Normally, any pulse of vibrational or other energy in the brain couldn't exist in a quantum mechanical mixture of states, because all the matter and activity in the brain would disturb it and instantly cause it to choose a single state. But according to some researchers' calculations, a microtubule could insulate a pulse from the hubbub; the pulse could travel along the microtubule oblivious to the noise around it, without having to interact with the molecules in the microtubule wall. And as long as the pulse wasn't forced to choose a single state, it would be free to explore simultaneously any number of possible patterns within and among microtubules. (Even crossing among cells wouldn't necessarily disturb the pulse; after all, one tuning fork can set another vibrating even if the intervening air is noisy with other signals.) Hameroff suspected this new aspect of microtubules might open up yet more possibilities, but he didn't understand enough about quantum mechanics to see how.

Then, in 1992, he got around to reading *The Emperor's New Mind*. Though he had to struggle through some of the sections on physics, he grasped enough of the material to recognize that whereas he had a quantum mechanical structure in the brain without a theory of consciousness to apply it to, Penrose had a theory of quantum consciousness that lacked an appropriate biological structure. "I figured I really ought to get in touch with this guy," Hameroff recalls.

Ever since the publication of *The Emperor's New Mind*, Penrose has received a great deal of mail; much he politely characterizes as "unreasonable." When he first started reading the letter from a Tucson anesthesiologist, he was not particularly encouraged. But as he read on, and then examined the enclosed papers, his interest grew. "Some of the things that Stuart is willing to put into print are sort of, um, far out," he says. "But even if your ideas sound like science fiction, one doesn't want to throw out the baby with the bathwater. There seemed to be something really important here." Hameroff ended his letter by informing Penrose that he was planning to visit London during an upcoming trip to Europe. Penrose wrote back: Come on down.

On a brisk, damp fall day in 1992, Penrose met Hameroff at the train station in Oxford. Back at his office, in a small clearing in the spectacular clutter just large enough for two chairs, the two sat by a window overlooking a garden. Hameroff fielded questions from Penrose for more than two hours: he explained what microtubules were, how they were configured in the brain, the ways they seemed to be able to pass on signals. The two discussed calculations showing that the microtubules' insulating properties would allow vibrational pulses to quantum mechanically explore multiple pathways. As Penrose listened, his concerns about Hameroff slowly evaporated. "In person, Stuart was much more careful to point out which of his ideas were on the wild side, and which ones weren't, than he was in print," he says. "It's the other way around with most researchers."

Though Penrose had seemed thoroughly intrigued by microtubules, Hameroff wondered after being dropped off at the train station whether Penrose would do much with the idea. "It seemed like sort of a long shot," Hameroff says. It wasn't until two weeks later that he learned Penrose had announced to an audience—just days after their meeting—that thanks to Hameroff, he had finally located a plausible site for the roots of consciousness.

Even as modified to include Hameroff's ideas, Penrose's theory is far from a done deal. For one thing, it is still vague on many crucial points. For example, Hameroff is still trying to determine how thoughts are represented by signaling patterns in the microtubules, and how these patterns trigger, supplement, or modify the firing of nerve impulses. And Penrose does not yet seem to be close to saying precisely how it is that a quantum mechanical mixture of patterns goes about "choosing" one.

Of course, the theory could simply be plain wrong. "There are so many places one could have turned off in the wrong direction," Penrose concedes. There is no hard evidence yet that the mind is based on a noncomputable process, for example, or even that such a process exists in quantum mechanics or anywhere else in physics. Essentially, what he knows is that the brain, in creating mind, seems to be doing something noncomputable and that the mechanism behind quantum mechanical "choosing" is so suggestively incomplete that it may well also involve noncomputable processes—and that there's a nice overlap between the two that seems to converge at microtubules.

But he does have ideas for confirming at least parts of his theory. His assertion that there is a gap in quantum mechanical theory because it can't account for the process of choosing a single state would be tremendously bolstered if he could come up with a prediction about the process that can't be wrung from conventional theory. If the current theory were complete, it would be able to provide, at least in principle, all predictions that can be made. If he can make a prediction that doesn't come from the theory, then the theory is incomplete. Penrose and other physicists believe that just such a prediction may, in fact, be possible. They have proposed an addition to quantum mechanics that would describe how much time, on average, a given particle in a given environment would last in a quantum mechanical mixture of states before choosing a single state—a question on which conventional theory is mute. "I believe these times could very possibly be measured," he says. "I'm talking to experimentalists about the possibility."

CELLULAR SCAFFOLDING

Until recently biologists viewed microtubule networks like these primarily as passive skeletal underpinnings for cells; now researchers suspect they are also able to actively transmit vibrational signals

Courtesy N. J. Hirokawa

Hameroff, for his part, is trying to come up with a scheme for experimentally determining how microtubules process patterns of signals. To do that, he is trying to design a device with two microscopic prongs: one would introduce a tiny electric jolt at certain points in a microtubule network; the other would detect any jolts that emerged at some other point. Determining where jolts emerge would provide some indication of how microtubules interconnect and direct signal flow as well as how these arrangements change over time. Eventually, Hameroff would like to determine how microtubule signal patterns represent information, the way a computer chip represents information with long series of 1's and 0's. "If we can decode the patterns," he says, "we might be able to connect microtubules to a computer and swap information back and forth."

But even in the absence of experimental evidence, Penrose has a very good feeling about the theory. "It's the old Sherlock Holmes argument," he says. "After you've eliminated the impossible, whatever remains, no matter how improbable, is the truth. I'm 90 percent sure these ideas are basically correct." After a moment of thought, he adds: "Well, maybe 80 percent. The Sherlock Holmes argument can be a dangerous one."

The quantum consciousness model, as an exquisite bonus, could in principle help lead the way to a new theory of physics that could repair a serious hole (to some) in quantum mechanics. Penrose believes such a theory will come to light sooner or later, whether or not he is right about consciousness—but it could happen sooner if the model proves correct.

What a great practical joke nature will have played on us if all the thinking that has gone into uncovering the ultimate laws of the universe turns out to reveal that one of the biggest clues was woven all along into the very fabric of thought itself.