forbes.com

Testing A Time-Jumping, Multiverse-Killing, Consciousness-Spawning Theory Of Reality

Andréa Morris

33–41 minutes

Nobel Laureate in Physics, Roger Penrose poses with his Nobel medal

POOL/AFP via Getty Images

“This retroactive idea. It has to be that,” says Nobel Prize-winning mathematical physicist Sir Roger Penrose, reflecting on a problem

about the building blocks of reality that has dogged physics for nearly a century. “Any sensible physicist wouldn't be perturbed by this,” he adds. “However, I'm not a sensible physicist.”

If Penrose isn’t a sensible physicist it’s because the laws of physics aren’t making sense, at least not on the subatomic level where the smallest things in the universe play by different rules than everything we see around us. He has reason to believe this disconnect involves a fissure that divides two different kinds of reality. He also has reason to believe that the physical process that bridges these realities will unlock answers to the physics of consciousness: the mystery of our own existence.

Penrose's contributions to math and physics are significant. He’s proposed a theory of sequential universes that existed before the big bang, traces of which seem to be penetrating ours. He collaborated with Stephen Hawking on the Penrose-Hawking singularity theorems, identifying points in the universe, *singularities*, where the gravitational forces are so intense that spacetime itself breaks down catastrophically.

For decades, Penrose has been working with anesthesiologist Stuart Hameroff on a theory of consciousness called *Orchestrated Objective Reduction* (Orch OR). Penrose primarily handles the physics of Orch OR whereas Hameroff handles the biology. Their theory was formulated as a response to serious gaps in established scientific frameworks spanning physics, neuroscience and psychology. All, some or none of the hypotheses in this theory might prove out experimentally.

The Theory Starts With A Tiny Collapse

The smallest bits of matter in the universe are quantum particles. Quantum particles exist in multiple possible states at once. This is called a particle's *superposition*. A *wave function* is a mathematical term that describes the particle's superposition. A wave function can collapse, causing a particle's many possible states to reduce to a single, fixed state. *Wave function collapse* is important for reality as we know it. It's because of collapse that when we look at something with our naked eye, we see one thing. In the realm of big things, the world described by classical physics, we don't see one thing as multiple possible things all at once.

The Connection Between Collapse And Consciousness

When scientists measure a particle, it seems to collapse to one fixed state. Yet no one can be sure what's causing collapse, also called *reduction of the state*. Some scientists and philosophers even think that wave function collapse is an elaborate illusion. This debate is called *the measurement problem* in quantum mechanics.

The measurement problem has led many physicists and philosophers to believe that a conscious observer is somehow acting on quantum particles. One proposal is that a conscious observer causes collapse. Another theory is that a conscious observer causes the universe to split apart, spiraling out alternate realities. These worlds would be parallel yet inaccessible to us so that we only ever see things in one single state in whatever possible world we're stuck in. This is the *Multiverse* or *Many Worlds* theory. "The point of view that it is consciousness that reduces the state is really an absurdity," says Penrose, adding that a belief in *Many Worlds* is a phase that every physicist, including

himself, eventually outgrows. “I shouldn't be so blunt because very distinguished people seem to have taken that view.” Penrose demurs. He politely but unequivocally waves off the idea that a conscious observer collapses wave functions by looking at them. Likewise, he dismisses the view that a conscious observer spins off near infinite universes with a glance. “That's making consciousness do the job of collapsing the wave function without having a theory of consciousness,” says Penrose. “I'm turning it around and I'm saying whatever consciousness is, for quite different reasons, I think it does depend on the collapse of the wave function. On that physical process.”

The Missing Force

What’s causing collapse? “It's an objective phenomenon,” insists Penrose. He’s convinced this objective phenomenon has to be the fundamental force: gravity. Gravity is a central player in all of classical physics conspicuously missing from quantum mechanics.

“There are a whole lot of people in this physics community who are trying to do quantum gravity,” says Penrose. “The sort of view, I gather, is that quantum mechanics is somehow more basic than gravitational theory and therefore you’ve got to bring gravity into the scheme of quantum mechanics.” With the majority of physicists wanting to bend gravity to accommodate quantum, Penrose pushes back. He sees some value in quantizing gravity, but he doesn’t think it should be the focus. “That’s not where physics should be going, not the experiments that should be done. It’s the other way around. It’s the influence of gravity on quantum mechanics. People don’t recognize fully enough that quantum mechanics is an inconsistent theory. It’s inconsistent with itself,”

says Penrose. “It’s not our understanding of quantum mechanics that has the gap, it’s the theory itself that has the gap.”

Penrose takes a hard pass on *Many Worlds* or ideas about conscious ghosts in the quantum machine as a way to bridge this gap. His bridge is neither an illusion nor a ghost. For Penrose, wave function collapse is a real, physical, objective phenomenon: a gravitational field can’t tolerate being in a quantum superposition, eventually collapsing the particle’s wave function. According to Penrose, gravity-induced wave function collapse involves a process that jumps the particle back in time, retroactively killing off possible quantum realities in under a second. This reality-annihilating backward-jumping makes it as though only one, fixed classical reality ever existed.

Sorry multiverses. But the death of multiverses allows for the birth of consciousness. Penrose’s theory proposes that each gravity-induced collapse causes a little blip of *proto-consciousness:* micro-events that get organized by biological structures called microtubules inside our brains into full-bodied awareness. A conscious observer doesn’t *cause* wave function collapse. A conscious observer is caused *by* wave function collapse.

From Incompleteness To Consciousness

Penrose’s interest in consciousness was inspired by a revolutionary mathematical discovery nearly a century ago. In 1931, mathematician Kurt Gödel revealed his incompleteness theorems—theorems of mathematical logic that show there are statements in mathematics that must be true even though they can’t be proven. Gödel’s incompleteness theorems, and

Goodstein's theorem sometime later, made an indelible imprint on Penrose. He took from these theorems that there's a unique property of the physical universe giving rise to conscious *understanding*. This is our human ability to understand truths that cannot be derived from the rules that gave us those truths. In other words, the rules allow us to ascertain truths beyond the rules. The ability to understand Gödel and Goodstein's theorems means there's something about our conscious understanding that is not confined to computational boundaries. Since all theories of physics are computational, Penrose believes something must be happening in the reduction of the quantum state that gives rise to non-computational understanding. "All I have are all the theories we know in physics. Computational, computational, computational. I mean, you've got to find room for this thing," says Penrose. He confirms that this thing that physics has to make room for is *understanding*.

Faster Than The Speed Of Light

Quantum weirdness doesn't stop at a thing existing in multiple possible states all at once. Quantum behaviors also seem to defy the laws of physics. Like the law that nothing can travel faster than the speed of light. When two quantum particles get close enough, their wave functions become entangled. Once entangled, you can separate the particles across the universe and anything you do to one particle instantly affects the other. If you make a measurement on one particle, collapsing its wave function, it immediately determines the state of the other particle, even if the other particle is located across the universe. Einstein called this *spooky action at a distance* because it seemed to suggest information was traveling

from one particle to another, faster than the speed of light. The 2022 Nobel Prize in physics was awarded to the team that proved entangled quantum particles do affect each other instantaneously even though they don't send a signal faster than the speed of light. “The quantum reality is, in some sense, not so fixed in spacetime,” says Penrose.

Backward Time-Jumping

According to Penrose, entangled particles merely appear to scientists as though they are affecting each other instantaneously. “It’s not even instantaneous. It’s *more* than instantaneous,” says Penrose, who sees collapse as a sort of boundary. On one side is the classical reality we know, where things are in one single state in space and time. The other side of the boundary is quantum reality where space, time and possibilities have a lot more freedom. Wave function collapse is something like a gateway between quantum and classical realities. “It's how quantum and classical physics relate to each other. It’s huge,” says Penrose.

The price to traverse realities is charged to classical reality’s timeline. Countless experiments show the collapse reduces multiple quantum states. Experiments also show this effect is instantaneous. But the effect may only *seem* instantaneous to us because the destruction of multiple quantum realities retroactively alters the classical reality timeline. In other words, classical reality retroactively emerges from the wave function collapse of quantum reality. Penrose calls this effect, aptly enough, *retro-activity.* It clears a path for making quantum behavior consistent with Einstein's theory of special relativity. Penrose thinks these backward time jumps are the only way a superposition can

collapse into a single, fixed state and still remain consistent with results from experiments in both quantum physics and classical physics.

Special relativity says time passes at different rates depending on your frame of reference. This is called *time dilation*. Experiments show that time dilation is a natural part of how time works. “There isn’t a universal time,” says Penrose. The average person and even other scientists may be skeptical about the idea of retro-activity. It may sound like science fiction for anyone unaccustomed to thinking about general relativity, special relativity and a universe where past, present and future already exist in a four-dimensional block. “I’ve been thinking about it, not since I’ve been in the cradle exactly,” says the 92-year-old, “but certainly a long way back.” In his 1989 pioneering book on consciousness, *Emperor’s New Mind*, Penrose first proposed the idea of a retroactive effect. In the book, he cautions that we may err when applying the physics of time to our conscious perception of time. He writes that consciousness is the only phenomenon in modern physics that requires time to flow at all.

Penrose’s ideas about retro-activity as an explanation for quantum anomalies are only recently gaining traction. *Retrocausality* is the proposal that a measurement in the present can change a particle’s properties even before the measurement was made. “You need this distinction between the two realities,” says Penrose. Classical reality and quantum reality are fundamentally different realities. He adds that even the notion of before and after may be incoherent in quantum reality.

Why might gravity-induced wave function collapse produce non-computational consciousness? Consciousness “could be non-

computable *because* it's retroactive," says Penrose.

Conscious Choices

For Penrose, this retro-active process helps explain how athletes make rapid decisions under extreme time constraints. "I used to play a lot of ping pong," says Penrose. "If I suddenly decide I want to shoot the ball this way rather than that way, I consider I'm making that decision consciously. Now that's far less than half a second." The process of taking in sensory information, making a decision and then acting, is a relatively lengthy physiological process. Decisions that involve a rapid reaction time are thought to be made unconsciously. According to cognitive psychology and neuroscience, the sense afterward that we made a conscious choice is an illusion. Penrose could never swallow this explanation. "Your conscious internal experience might be a kind of quantum reality," offers Penrose. He suspects we may, on some level, be conscious of all the possible realities that get retroactively annihilated in under a second.

"The argument is that there would be something in quantum superposition between this action and that action—somewhere at the earlier stage in the brain when these two procedures are in quantum superposition," says Penrose. "So the quantum state would contain both those alternatives. And then, when you decide to do one, it retroactively goes back." Jumping back and overwriting multiple quantum choices makes it *as if* there was only ever one, fixed classical choice. "Conscious experience happens in quantum reality. And classical reality is retroactively determined by that," says Penrose. He's quiet for a moment before gently voicing a concern that people might misinterpret what he's saying

about retro-activity, but mainly because he's still working out the details and potential paradoxes himself. "It's too easy for people to speculate in ways which are almost certainly wrong," says Penrose before emphasizing that retro-activity can only happen along the past light cone. The past light cone is a cone-shaped region in spacetime that represents every single past event that could have influenced a particular event. If retro-activity happens, it happens within these parameters.

The Critics

Penrose doesn't shy away from lobbing bold ideas into the public square of scientific debate before he's worked out all the details. In turn, the scientific community doesn't shy away from piling on when someone in their camp goes rogue. Penrose recalls giving a talk at the California Institute of Technology on his heterodox ideas in cosmology. Physicist Richard Feynman attended so he could heckle Penrose. Over the course of the talk, Feynman grew intrigued by what Penrose was saying. When another physicist heckled Penrose, Feynman turned in his seat and told the heckler to shut it and let the man speak.

Today, Penrose gets accused of making unsupported connections between strange phenomena in quantum mechanics and the mystery of consciousness. "People complain to me 'he's just saying, here's a mystery, there's a mystery, therefore they're the same thing.' That's not what I'm saying," says Penrose. "I can see why they complain that way. It's not that." Over the next hour he describes alternative theories and gives reasons for why he doesn't think they're credible. It's unclear to what extent he's driven by the reasoning of his own theory or by the implausibility of

any alternatives. He suggests that the only other good alternative might be a theory that no one has thought of yet. As things stand, he feels that both classical physics and quantum mechanics are extraordinary theories. Both have proven to be extraordinarily precise when tested. So Penrose is writing a chapter in modern physics that he hopes will unite them: “I think measuring the collapse of the wave function is the most important experiment anybody should do and not many people are trying.”

His polite skepticism and genial demeanor belies an unflagging determination to see his own ideas either proven out or falsified. There are three core hypotheses to be tested experimentally:

1) gravity causes wave function collapse

2) the collapse involves retro-activity

3) consciousness comes out of this process

Testing Gravity-Induced Wave Function Collapse

In 2022, a group of scientists ran an experiment and published a subsequent press release claiming they disproved Penrose's theory by disproving a prediction made by physicist Lajos Diósi. Diósi and Penrose had a similar timescale for how long it would take gravity to collapse the wave function. Their ideas were folded together and coined the *Diósi-Penrose model*. “Diósi’s model has some problems, very serious problems, which is that it doesn't conserve energy,” says Ivette Fuentes, a physicist at University of Southampton and Oxford Fellow. Diósi and Penrose agreed that gravity causes wave function collapse. They also agreed about how long it would take. For Diósi, however, gravity-induced wave function collapse involved radioactive heating. The 2022

experiment did not find radioactive heating, thereby disproving Diósi's theory. For Penrose, there is no radioactive heating because the collapse involves retro-activity. There were other issues with the experiment. "One of the things Roger predicts is that if you have a particle in a superposition, a massive particle in a superposition, it will collapse," says Fuentes. "But the [Diósi] experiment doesn't have a superposition. The experiment was one big mass not in a superposition."

Solids like mirrors, levitated nanobeads and diamonds are traditional materials for testing wave function collapse. Fuentes has a unique, non-solid approach. She cools atoms to the absolute lowest temperature possible on earth, turning them into a new state of matter resembling a gas. This kind of matter is called *Bose-Einstein Condensates* (BECs). Fuentes' work with BECs caught Penrose's attention and the two began collaboration on an experiment using BECs to test the first stages of gravity-induced wave function collapse called *the shaking of the building*. When testing a quantum particle in BECs, "the system behaves very differently and it's very sensitive to gravity," says Fuentes.

Like Penrose, Fuentes embraces the inclusion of consciousness in physical theories, as long as physical theories provide an explanation for what consciousness actually is. From the time she was in high school, Fuentes wanted to understand how consciousness emerged from the interaction of atoms and molecules. In the 1990s, there was not a single scientific discipline where consciousness was considered a serious area of study. Family members in science and medicine advised her to go into psychology or neuroscience, two areas proximal to her interests. Fuentes had a sense that answers to her questions weren't going

to be found in those fields, so she became a physicist. Now she designs out-of-the-box ways of testing problems about our understanding of the universe. Increasingly, this path seems the surest route back to her original question. “We're at the brink of some sort of shift or change in which we will have to incorporate mind and consciousness to make a fuller picture, a better picture,” says Fuentes adding, “I do think we need a change. And I do think that it involves having mind as part of the equation. And maybe, by this shift, we'll be able to understand why we were banging our heads not being able to bring quantum mechanics and general relativity together.”

Penrose and Fuentes teamed up with quantum physics experimentalist Philippe Bouyer at University of Amsterdam to design the BEC experiment. They’ve raised $2 million USD from global philanthropists. The project needs an additional $4 million. Once funded, the experiment will take approximately five years to complete.

If gravity-induced wave function collapse can be proven with BEC experiments, Penrose still needs to prove this process involves retro-activity and consciousness. He has ideas about testing for retro-activity using the Italian Space Agency’s mirrored disco-ball-like LARES satellite. Still, neither satellites nor BECs have anything to say about consciousness. If BECs are systems sensitive enough to test for gravity’s influence on quantum particles, Penrose thinks human beings might be physical systems sensitive enough to test for consciousness registering retro-activity.

Retroactivity In Psychological Experiments

“Am I the last survivor of the team?” asks Dennis Keith Pearl, statistician and co-author of a 1979 experiment led by late psychologist Benjamin Libet. Libet is best known for his seminal research that seems to show that our choices to act are too slow to be made consciously. The brain "registers" the decision to make movements before we consciously decide to move. Libet studies are controversial because they seem to do away with free will. Penrose isn’t too concerned with free will, but he does believe our choices are made consciously, not unconsciously, regardless of whether or not they’re free. Decades ago, physicist Erich Harth, a colleague of Penrose, brought Libet’s 1979 experiment to Penrose’s attention. Harth thought it may contain evidence that the brain is registering retro-activity. Retro-activity could give us the fractions of a second we need to salvage conscious choice. Harth included an interpretation of the Libet study in his book *Windows On The Mind*.

Pearl was a graduate student in 1979 and the youngest on Libet’s research team, which included California senator Dianne Feinstein’s husband, neurosurgeon Bertram Feinstein. “Too bad you weren't asking me 10 years ago,” says Pearl as he struggles to remember details from a half-century-old experiment. “I had a box full of all the original records from my work with Ben,” says Pearl. “I had lots of notes from Ben and original graphs and things like that.” Pearl had never been contacted about his work with Libet, despite the fact that Libet names Pearl in his written defense of his research, at one point writing in the journal of *Consciousness and Cognition* to “take up any statistical difficulties with Dennis Pearl.” Boxes of materials and raw data were tossed out during a move a decade ago. Now Pearl carefully inspects the

graphs that Harth constructed, graphs interpreted from the 1979 study. “I think everything that [Harth’s] got on this graph is correct in terms of what's reported,” says Pearl.

He’s drawn to Penrose’s use of probabilities in consciousness. He recalls a Libet experiment that he thinks might be of interest to Penrose. Libet stimulated a subject with a short burst of stimulus, and asked the subject if they felt it. The subject would report they did not. So Libet would ask the subject to hazard a guess. An ultra-short burst of stimulus that wasn’t likely to be felt resulted in sheer random guesses. As the bursts extended in duration, the subject would continue to report they couldn’t feel anything. However, guesses started to improve with accuracy until guesses were 100% accurate.

“[Libet] sent me some data and I looked at the curve and said, you know, these guys are getting it right,” says Pearl, recalling the conversation with Libet about a smooth probability curve from unconsciousness towards consciousness. “There's a fuzziness of time. That fuzziness is more on a probability scale. It's moving toward complete awareness, but in the meantime, there's some sort of a semi-foggy kind of period,” says Pearl, cautioning that he’s thinking about this as a statistician, not a neuroscientist or a physicist. He combs through papers trying to find the study where these results were published. Ultimately, he can’t. He wonders if it never made it into a publication because the experiment was only done on two patients.

Pearl takes another look at Harth’s graph. This time, something jumps out at him: the timescale from the infamous Libet clock. In the 1979 experiment, the duration of stimulus was timed precisely but not the subject's response. The timescale is an imperative

detail. Without it, evidence for retro-activity in the 1979 experiment never existed. Left in its place isn't a fixed classical state so much as an open question: Harth's mistaken interpretation of retro-activity in the Libet experiment doesn't undermine the retro-active hypothesis in physics. In fact, remove the Libet clock and there's nothing in physics preventing retro-activity from jumping even further back in time. So the question remains—if backward time jumps are happening, would it impact how we observe reality? And would that impact psychology studies in unexpected ways?

"Our results, there's something weird happening, and we're trying to get to the bottom of it," says cognitive scientist Marc Buehner, co-author of the study *Human Vision Reconstructs Time to Satisfy Causal Constraints* published in the journal *Association for Psychological Science.* "The visual system reorders the evidence, as it comes in," says Buehner. Imagine a game of pool. The white cue ball hits a yellow ball and a yellow ball then hits a purple ball into the corner pocket. There's a causal chain of white hitting yellow causing it to hit purple into the pocket. Buehner's study shows that at least sometimes, our visual system lies to us about this causal order. Buehner and his team conducted experiments where an ABC causal sequence is presented to subjects out of order. Instead of ABC, the researchers mixed up the sequence so C moved inexplicably before B. Subjects saw this ACB disordered sequence but reported an ABC order, despite repeat viewings of the out of order sequence.

"It's basically as if the visual system actually reverses it. So it turns ACB into ABC," says Buehner. "This weird stimulus as a whole, for reasons that are still not really quite known to us, creates an expectation of this causal event. So the expectation is that it

should be ABC, and that expectation clashes with reality," says Buehner. Interpreting sensory information from the environment to create a mental representation of the world involves a process we're not aware of. It's automatic and not consciously controlled. "What we demonstrated in this paper is that perception actually changes," says Buehner. The researchers ruled out a false memory of what the subjects just saw, called *post perceptual distortion* or *reinterpretation*. The effect also can't be explained by lapsed attention, or rapid, jerky eye movements we make when we shift our gaze, called *saccades*. "So you could say, oh it's just another one of those visual illusions. Because I asked you afterward, it's kind of like a post fiction. So you try to make sense of it. There's this weird thing you try to make sense of," says Buehner. "Except that's not what's happening. We could show that you actually perceive the motion onset in the B stimulus as later and the motion onset of the C stimuluses earlier. So you actually perceive a reversal live—as it happens."

An underlying assumption in perceptual science is that the brain uses sensory input to create mental representations of the world that correspond to what's actually happening out there. This is referred to as *veridical representations*—mental pictures that align with reality. Studies like Buehner's would suggest that either assumptions about the brain might be wrong, or assumptions about reality. "I'm not sure that I would necessarily want to make grand claims that potentially results are driven by some kind of like, you know…" Buehner presses the air with his fingers, "tapping into quantum mechanics. But if that's what's behind it, hey, that'd be super cool. But I want to be cautious." Buehner adds that it would be good to know if physics is doing something weird

that’s responsible for unexplained results in psychological experiments.

Could Consciousness Dethrone Spacetime?

Is it outrageous to imagine developments in physics could upend findings in cognitive science? “All of my colleagues, and again, these are my friends and they're brilliant, but they believe that space and time are fundamental and that brain activity causes conscious experiences,” says Donald Hoffman, cognitive scientist and author of the book *The Case Against Reality: Why Evolution Hid the Truth from Our Eyes*. Hoffman rejects Orch OR’s depiction of reality along with every other physical theory. He thinks the long-standing barrier between classical physics and quantum mechanics is because we’re assuming space and time are fundamental. “Spacetime—we thought it was the final reality. It turns out it's just a trivial data structure and there are much deeper and much more fascinating structures entirely outside of spacetime,” says Hoffman.

He echoes Nima Arkani-Hamed, a theoretical physicist at the Institute for Advanced Study at Princeton university who says spacetime is doomed. Hoffman’s research suggests that the underlying assumptions in perceptual science, neurophysiology and psychology are wrong—the brain does not use sensory input to create accurate mental representations of reality. Hoffman ran simulations using evolutionary game theory and observed that evolution selects for fitness over truth. According to Hoffman, we perceive a completely false reality that is far more practical for survival, useful illusions that lead us far afield the truth-seeking path.

The alternative theory Hoffman proposes is that conscious entities are fundamental entities that exist beyond spacetime. These entities are us. And we are also avatars of a single conscious entity that Hoffman calls the “conscious aleph infinity agent.” We interact with each other via an interface whose format is spacetime. For Hoffman, what’s really going on outside of conscious awareness is so complex, involving non-spacetime dimensions numbering in the trillions or quadrillions. Our simple human minds created an ultra-compressed version of reality stripped of details that would break our brains—if we actually thought with our brains, which Hoffman sees no convincing evidence for.

Hoffman is critical of theories of consciousness like Orch OR. “There's not a specific conscious experience that they can explain. Not one,”says Hoffman. Whereas modern physics has mostly omitted consciousness from theories of reality, Hoffman believes consciousness is the starting point for a theory of reality. He claims to start with a mathematically precise theory of consciousness from which physicists can derive reality. “I'm not going to stipulate all of the other stuff that they stipulate,” says Hoffman, who considers each and every conscious experience fundamental. The taste of chocolate ice cream and an infinite variety of experiences are irreducible and fundamental.

“What I think science has taught us that spiritual traditions didn't understand,” says Hoffman, “is that imprecise theories don't get you anywhere or they can get you in trouble. You can start fighting with each other and be dogmatic and kill each other because you disagree on descriptions. Once you start having mathematically precise descriptions you're forced to really look at your

experiments carefully," says Hoffman, whose theory is based on Markov chains. A Markov chain is a mathematical construct, a system that undergoes transitions from one state to another according to certain probabilistic rules where nothing about the past affects the probability of the future. "The math is absolutely essential to the correct interpretation or more useful interpretations of the experiments," says Hoffman.

Hoffman's math leads him to conclude that we are avatars of a *superconscious* or *arch-conscious agent*. The arch-conscious agent puts us avatars through the paces of an infinite number of experiences, no matter how joyous or horrific, so that the arch-conscious agent can experience everything. Hoffman also warns against overidentifying with our *self*, because the *self* is an avatar. What's more: "You are not any particular experience. You are the potential in which those experiences arise and disappear. That's what you really are in your essence. You transcend any particular experience because you are that potential," says Hoffman.

Hoffman's theory of consciousness resonates with many spiritual narratives, suggesting a unifying force exploring all of its potential. Because of this, it confronts significant ethical questions, grappling with notions like whether we, at the most fundamental level, are a powerful conscious force willingly subjecting ourselves and others to the most painful, terrifying and tragic experiences just to satiate a gluttonous drive for experience. Its intriguing alignment with spiritual philosophies means Hoffman's theory faces the same daunting challenge of explaining the existence of evil and suffering. Hoffman's theory is quite popular. His interview with Lex Fridman has over 6.4 million views on YouTube. "Spacetime is over. It's not fundamental in any sense. It's not like we have to go

do smaller things inside spacetime. We have to go entirely outside of spacetime," says Hoffman.

"Okay, I'm the conservative person," laughs Penrose upon learning of Hoffman's view. Penrose is a physicalist. Whatever consciousness is, he's convinced it can be explained by the laws of physics, and he's fairly confident our current theories give us at least some idea of what those laws are. "It's hugely tempting to go off in a wild direction," says Penrose, highlighting the risky business of trying to account for consciousness scientifically. He raises a concern that throwing around mathematical terminology can make a theory seem more credible than it is. Experiments are the anchor for any scientific theory. Hypotheses must be tested and the model subjected to experimental falsifiability to qualify as a scientific theory. It must have the potential to be disproven in order to distinguish itself from pseudoscience. According to Penrose, there's a risk of getting caught up in the beauty of a precise mathematical theory. "I think it's dangerous," says Penrose, "It could be that there's a deeper beauty which tells you why the thing you thought was true is not true." Given the track record of experimental success for both classical physics and quantum mechanics, and the lack of evidence needed to replace all of physics with a conscious agent, Penrose doesn't see the rush to flip the table on spacetime. "It's just that the laws of physics may be more puzzling than we think they are," says Penrose.

Can Artificial Intelligence Ever Be Conscious?

When it comes to the suddenly salient question of whether or not AI could be conscious, Penrose draws again from Gödel and Goodstein's theorems. Computer science is built on formalized

systems. They're confined by computation. For Penrose, AI built on classical computers today isn't capable of true understanding or consciousness. After some consideration, he adds a caveat when it comes to quantum computers: "You put wave function collapse into its process somehow…"

For an in-depth discussion about this theory, including Penrose's *Hemingway Paradox*, watch the interviews with Penrose that were the basis for this reporting:

**This article has been updated. A previous version reported that Benjamin Libet was the first recipient of a Nobel Prize in Psychology. However, the Klagenfurt Virtual Nobel Prize has no relation to the Nobel Prize from the Swedish Nobel Foundation