Virtual Immortality

Why the Mind-Body Problem is Still a Problem

BY ROBERT LAWRENCE KUHN

Virtual immortality is the theory that the fullness of our mental selves can be uploaded with first-person perfection to non-biological media, so that when our mortal bodies die our mental selves will live on. I am all for virtual immortality and I hope it happens (rather soon, too). Alas, I don't think it will (not soon, anyway). I'd deem it virtually impossible for centuries, if not millennia. Worse, virtual immortality could wind up being absolutely impossible, forbidden even in principle.

This is not the received wisdom of optimo-techno-futurists, who believe that the exponential development of technology in general, and of artificial intelligence (AI) in particular (including the complete digital duplication of human brains in the near or mid term), will radically transform humanity through two revolutions. The first is the "singularity," when AI will redesign itself recursively and progressively, such that it will become vastly more powerful than human intelligence (superstrong AI). The second, they claim, will be virtual immortality.

AI singularity and virtual immortality would mark a startling, transhuman world that optimo-techno-futurists envision as inevitable in the long run and perhaps just over the horizon in the short run. They do not question whether their vision can be actualized; they only debate when will it occur, with estimates ranging from 10 to 100 years.

I'm skeptical. I think the complexity of the science is vastly underrated, and I challenge the philosophical foundation of the claim. Consciousness is the elephant in the room, though many refuse to see it. They assume, almost as an article of faith, that superstrong AI (post-singularity) will inevitably be conscious (almost *ipso facto*). They may be correct, but to make that judgment requires an analysis that is surely multifaceted and, I suspect, likely inconclusive.

Whatever consciousness may be, it determines whether virtual immortality is even possible. So I focus here on consciousness. First, however, there are two other potential obstacles to virtual immortality. I consider them briefly.

One is sheer complexity. What would it take to duplicate the human brain such that our first-person inner awareness, and all that it entails, could not be distinguished from the original?

Consider some (very) rough data for the human brain: it contains about 85 billion neurons (specialized nerve cells that convey electrical information); 100 to 1000 trillion synapses (small space between neurons, the junction across which information is transmitted by neurochemicals); one to five trillion glial cells (traditionally assumed limited to metabolic support for neurons, now suspected as also participating in brain functions); up to 1000 moments per second for positioning action potentials (the electrical spark of information in neurons); ten billion proteins per neuron (some of which form memories); innumerable 3-dimensional structural forms for proteins and their geometric interactions; various extracellular molecules (some of which may be involved in brain functions). The list goes on.

How much of all of this complexity is required for total virtual duplication such that the mental fullness of the original person can be said to exist? Who knows?

Granted, not all of the brain is needed for consciousness and its contents, and much of the machinery is metabolic. The bodily control mechanisms, such as regulating breathing and heart rate and digestion, will not be needed in non-biological substrates. On the other hand, contemporary philosophy of mind suggests that bodily sense is needed for normal cognition (i.e., "embodied brain" and "extended mind").

Take all the brain data together and consider all possible combinations and permutations that work to generate the more than 100 billion distinct human personalities who have ever lived (each of whom differs from moment to moment). I hesitate even to estimate the number of specifications that would be required. How could all these be accessed non-invasively, in sufficient detail, in real time, and simultaneously? The technologies exceed my imagination. But in principle, they are possible.

A second potential deterrent to virtual immortality is quantum mechanics, the inherent indeterminacies of which could make creating a perfect

mental duplicate problematic or even impossible. After all, if quantum events (like radioactive decays) are in principle non-predictable, how then would it be possible to duplicate a brain perfectly?

But quantum indeterminacies exist everywhere, in bricks just as well as in brains, so its special applicability to brain function and hence to virtual immortality is questionable. The crux of the issue is at which level in the hierarchy of causation, if any, does quantum mechanics make meaningful contributions to brain function and to consciousness? Certainly the vast majority of neuroscientists think quantum mechanics works only at bedrock levels of fundamental physics, way too low to play any special role at the higher levels where brains work and minds happen.

So while the sheer complexity of the brain would deter virtual immortality, and the indeterminacy of quantum mechanics might be an insurmountable obstacle to perfect duplication, the former would only delay its advent while the latter is deemed not relevant.

That leaves consciousness—that elephant in the room—around which optimo-techno-futurists have gathered to plan their virtual afterlife.

What is Consciousness?

Consciousness is a main theme of *Closer To Truth*, my public television series on science and philosophy, and among the subtopics I discuss with scientists and philosophers on the program is the classic "mind-body problem"—what is the relationship between the mental thoughts in our minds and the physical activities in our bodies/brains? What is the deep cause of consciousness? (All quotes that follow are from *Closer To Truth*: www.closertotruth.com.)

NYU philosopher David Chalmers famously described the "hard problem" of consciousness: "Why does it feel like something inside? Why is all our brain processing —vast neural circuits and computational mechanisms—accompanied by conscious experience? Why do we have this amazing inner movie going on in our minds? I don't think the hard problem of consciousness can be solved purely in terms of neuroscience."

"Qualia" are the core of the mind-body-problem. "Qualia are the raw sensations of experience," Chalmers continued. "I see colors—reds, greens, blues—and they feel a certain way to me. I see a red rose; I hear a clarinet; I smell mothballs. All of these feel a certain way to me. You must experience them to know what they're like. You could provide a perfect, complete map of my brain [down to elementary particles]—what's going on when I see, hear, smell—but if I haven't seen, heard, smelled for myself, that brain map is not going to tell me about the quality of seeing red, hearing a clarinet, smelling mothballs. You must experience it."

Can a Computer be Conscious?

To Berkeley philosopher John Searle, computer programs can never have a mind or be conscious in the human sense, even if they give rise to equivalent behaviors and interactions with the external world. (In Searle's "Chinese Room" argument, a person inside a closed space can use a rule book to match Chinese characters with English words and thus appear to understand Chinese, when, in fact, she does not.)

"But," I asked Searle, "Will it ever be possible, with hyperadvanced technology, for non-biological intelligences to be conscious in the same sense that we are conscious? Can computers have 'inner experience'?"

"It's like the question, 'Can a machine artificially pump blood as the heart does?" Searle responded. "Sure it can—we have artificial hearts. So if we can know exactly how the brain causes consciousness, down to its finest details, I don't see any obstacle, in principle, to building a conscious machine. That is, if you knew what was causally sufficient to produce consciousness in human beings and if you could have that [mechanism] in another system, then you would produce consciousness in that other system. Note that you don't need neurons to have consciousness. It's like saying you don't need feathers in order to fly. But to build a flying machine, you do need sufficient causal power to overcome the force of gravity."

Searle then cautioned: "The one mistake we must avoid is supposing that if you simulate it, you duplicate it. A deep mistake embedded in our popular culture is that simulation is equivalent to duplication. But of course it isn't. A perfect simulation of the brain—say, on a computer—would be no more conscious than a perfect simulation of a rainstorm would make us all wet."

Robotics entrepreneur (and MIT professor emeritus) Rodney Brooks agrees that consciousness can be created in non-biological media, but disagrees on the nature of consciousness itself. "There's no reason we couldn't have a conscious machine made from silicon," he said. Brooks' view is a natural consequence of his beliefs that the universe is mechanistic and that consciousness, which seems special, is an illusion. He claims that, because the external behaviors of a human, animal or even a robot can be similar, we "fool ourselves" into thinking "our internal feelings are so unique."

Can We Ever Really Assess Consciousness?

"I don't know if you're conscious. You don't know if I'm conscious," said Princeton neuroscientist Michael Graziano. "But we have a kind of gut certainty about it. This is because an assumption of consciousness is an attribution, a social attribution. And when a robot acts like it's conscious and can talk about its own awareness, and when we interact with it, we will inevitably have that social perception, that gut feeling, that the robot is conscious."

"But can you really ever know if there's 'anybody home' internally, if there is any inner experience?" he continued. "All we do is compute a construct of awareness."

Warren Brown, a psychologist at Fuller Theological Seminary and a member of UCLA's Brain Research Institute, stressed "embodied cognition, embodied consciousness," in that "biology is the richest substrate for embodying consciousness." But he didn't rule out that consciousness "might be embodied in something non-biological." On the other hand, Brown speculated, "consciousness may be a particular kind of organization of the world that just cannot be replicated in a non-biological system."

Neuroscientist Christof Koch, president and chief scientific officer of the Allen Institute for Brain Science, takes a strong philosophical stance based on his work as a neuroscientist. "I am a functionalist when it comes to consciousness," he said. "As long as we can reproduce the same kind of relevant relationships among all the relevant neurons in the brain, I think we will have recreated consciousness. The difficult part is, what do we mean by 'relevant relationships'? Does it mean we have to reproduce the individual motions of all the molecules? Unlikely. It's more likely that we have to recreate all the relevant relationships of the brain's synapses and the brain's wiring (today known as the 'connectome') in a different medium, like a computer. If we can do all of this reconstruction at the right level, this entity, this software construct, would be conscious."

Koch stresses that "experience" requires new, perhaps radical, scientific thinking. "You need to expand the traditional laws of physics," he told me. "In physics there is space, time, energy, mass. Those by themselves are sufficient to explain the physics of the brain. The brain is subject to the same laws of physics as any other object in the universe. But in addition there is something else. There is experience. The experience of pain. The experience of falling in love. And to account for experience, you need to enhance the laws of physics."

Radical Visions of Consciousness

A new theory of consciousness—developed by Giulio Tononi, a neuroscientist at the University of Wisconsin (and supported by Koch)—is based on "integrated information" such that distinct conscious experiences are represented by distinct structures in a specialized and heretofore unknown kind of space.

"Integrated information theory means that you need a very special kind of mechanism organized in a special kind of way to experience consciousness," Tononi said. "A conscious experience is a maximally reduced conceptual structure in a space called 'qualia space.' Think of it as a shape. But not an ordinary shape—a shape seen from the inside."

Tononi stressed that simulation is "not the real thing." To be truly conscious, he said, an entity must be "of a certain kind that can constrain its past and future—and certainly a simulation is not of that kind."

Koch envisions how Tononi's theory of integrated information could explain how experience—how consciousness—arises out of matter. "The theory makes two fundamental axiomatic assumptions," Koch explained. "First, conscious experiences are unique and there are a vast number of different conscious experiences. Just think of all the frames of all the movies you've ever seen or movies that will ever be made until the end of time. Each one is a unique visual experience and you can couple that with all the unique auditory experiences, pain experiences, etc. All possible conscious experiences are a gigantic number. Second, at the same time, each experience is integrated—what philosophers refer to as unitary. Whatever I am conscious of, I am conscious of as a whole. I apprehend as a whole. So the idea is to take these two axioms seriously and to cast them into an information theory framework. Why information theory? Because information theory deals with different states and their interrelationships. We don't think the stuff the brain is made out of is really what's critical about consciousness. It's the interrelationship that's critical."

I asked Koch if he'd be "comfortable" with non-biological consciousness.

"Why should I not be?" he responded. "Consciousness doesn't require any magical ingredient."

Mathematician Roger Penrose claims that consciousness is non-computable and that only a non-computational physical process could explain consciousness. He is not saying that consciousness is beyond physics, rather that it is beyond today's physics. "Conscious thinking can't be described en-

tirely by the physics that we know," Penrose said, explaining that he "needed something that had a hope of being non-computational." He focuses on "the main gap in physics": the contradiction between the continuous, probabilistic evolution given by the Schrödinger equation in quantum mechanics and the discrete, deterministic events when you make a measurement in classical physics—"how rules like Schrödinger's cat being dead and alive at the same time in quantum mechanics do not apply at the classical level."

Penrose argues that the missing physics that describes how the quantum world becomes the classical world "is the only place where you could have non-computational activity." But he admits that it's "a tall order" to sustain quantum information in the hot, wet brain, because "whenever quantum systems become entangled with the environment, 'environmental decoherence' occurs and information is lost."

"Quantum mechanics acting incoherently is not useful [to account for consciousness]," Penrose explains; "it has to act coherently. That's why we call [our mechanism] 'OrchOR', or 'orchestrated objective reduction'—the 'OR' stands for objective reduction, which is where the quantum state collapses to one alternative or another, and 'Orch' stands for orchestrated. The whole system must be orchestrated, or organized, in some global way, so that the different reductions of the states actually do make a big difference to what happens to the network of neurons."

So how can the hot, wet brain operate a quantum information system? A biological mechanism utilizing microtubules in neurons was proposed by Dr. Stuart Hameroff, an anesthesiologist, who then together with Penrose, developed their quantum theory of consciousness

"Objective reduction in the quantum world is occurring everywhere," Hameroff recognizes, "so proto-conscious, undifferentiated moments are ubiquitous in the universe. Now in our view when orchestrated objective reduction occurs in neuronal microtubules, the process gives rise to rich conscious experience." But, he asked rhetorically, "could your consciousness be downloaded into some artificial medium as the singularity folks have been saying for years, but without any progress whatsoever?" Hameroff thinks it is possible. "It could happen in an alternative medium that has the proper properties," he said, "perhaps artificial nanotubes made of carbon fullerenes. [Creating consciousness in non-biological media] can be done as long as you have enough mass superposition to reach threshold in a reasonable time."

Inventor and futurist extraordinaire Ray Kurzweil believes that "we will get to a point where computers will evidence the rich array of emotionally subtle behaviors that we see in human beings; they will be very intelligent, and they will claim to be conscious. They will act in ways that are conscious; they will talk about their own consciousness and argue about it just the way you and I do. And so the philosophical debate will be whether or not they really are conscious—and they will be participating in the debate."

Kurzweil argues that assessing the consciousness of other [possible] minds is not a scientific question. "We can talk scientifically about the neurological correlates of consciousness, but fundamentally, consciousness is this subjective experience that only I can experience. I should only talk about it in first-person terms—although I've been sufficiently socialized to accept other people's consciousness. There's really no way to measure the conscious experiences of another entity."

"But I would accept that these non-biological intelligences are conscious," Kurzweil concluded. "And that'll be convenient, because if I don't, they'll get mad at me."

AI Consciousness: Precursor of Virtual Immortality

It is my conjecture that unless humanlike inner awareness can be created in non-biological intelligences, uploading one's neural patterns and pathways, however complete, could never preserve the original, first-person mental self (the private "I"), and virtual immortality would be impossible. That's why a precursor to the question of virtual immortality is the question of AI consciousness. Can robots, however advanced their technology, ever have inner awareness and first-person experience?

I submit that the nature of the AI singularity would differ profoundly in the case where it is literally conscious, with humanlike inner awareness, from the case where it is not literally conscious—even though in both cases superstrong AI would be vastly more powerful than human intelligence and by all accounts they would appear to be equally conscious. This difference between being conscious and appearing conscious would become even more fundamental if, by some objective, absolute standard, humanlike inner awareness conveys some kind of "intrinsic worthiness" to entities possessing it.

For example, the first colonizers of the cosmos will likely be robots, eventually self-replicating robots, and whether such non-biological probes are

conscious or not-conscious could radically affect the intrinsic nature of such colonization. It would differ profoundly, I suggest, in the case where such robots were literally conscious, with humanlike inner awareness and thus experiencing the cosmos, from the case where they were not literally conscious with no inner awareness and not experiencing anything.

I agree that after superstrong AI exceeds some threshold, science could never, even in principle, distinguish actual inner awareness from apparent inner awareness, say in our cosmos-colonizing robots. But I do not agree with what usually follows: that this everlasting uncertainty about inner awareness and conscious experience in other entities (non-biological or biological) makes the question irrelevant. I think the question maximally relevant. Unless our robotic probes were literally conscious, even if they were to colonize every object in the galaxy, the absence of inner experience would mean a diminished intrinsic worth.

That's why, to explore the possible meaning of AI consciousness as well as to assess the real-world viability of virtual immortality, the deep cause of consciousness is critical.

Alternative Causes of Consciousness

Through my conversations (and decades of nightmusings) on the philosophy of mind, I can array nine alternative theories or causes of consciousness. (There are others, and different categorizations). Traditionally, the clash is between physicalism/materialism (No. 1 below) and dualism (No. 8), but such oversimplification may be part of the problem—the other six alternatives have standing.

1. *Physicalism or Materialism*. Consciousness is entirely physical, solely the product of biological brains, and all mental states can be fully "reduced" to (wholly explained by) physical states—which, at their deepest levels, are the fields and particles of fundamental physics. Overwhelmingly for scientists, physicalism/materialism is the prevailing theory of consciousness. To them, the utter physicality of consciousness is an assumed premise, supported strongly by incontrovertible evidence. "Eliminative materialism" is the boundary position that our common-sense view of the mind is misleading and that consciousness is in a sense an illusion. A preferred mechanism of physicalism/materialism is identity theory, where mental states literally *are* physical states. (Though the terms "materialism" and "physicalism" are generally interchangeable, materialism is older and connotes a more metaphysical or ontological meaning, whereas physicalism emerged in the early 20th century and conveys a more methodological or linguistic usage.)
2. *Epiphenomenalism*. Consciousness is entirely physical, solely the product of biological brains, but mental states cannot be entirely reduced to physical states (brains or otherwise), though mental states have no powers. The mind is entirely inert; our awareness of consciousness is real but our sense of mental causation is not. There is no "top-down causation"; our feelings that our thoughts can cause things are an illusion. In this manner, epiphenomenalism is a weaker form of non-reductive physicalism (see next). The classic analogy for consciousness as an epiphenomenon is "foam on a wave," always there but never doing anything.
3. *Non-reductive Physicalism*. Consciousness is entirely physical, solely the product of biological brains, but mental states are real and cannot be reduced to physical states (brains or otherwise). While mental states are generated entirely by physical states (of the brain), they are truly other than physical (i.e., mental states are ontologically distinct). A prime feature of non-reductive physicalism is "top-down causation," where the content of consciousness is causally efficacious—qualia can do real work. The mechanism of non-reductive physicalism is emergence, where novel properties at higher levels of integration are not discernible (and perhaps not even predictable) from all-you-can-know at lower or more fundamental levels. (There is a close relationship between non-reductive physicalism and property dualism—both recognize real mental states and yet only one kind of substance—but, as expected, some adherents of each reject the claims of the other.)
4. *Quantum Consciousness*. Consciousness is non-computational and relates to (or resides in) the fundamental gap between the quantum and the classical worlds. Consciousness is still explained by the physics of neurons, but a physics enlarged from that which we know currently. Though dismissed by most scientists, the claim is that these two great mysteries, consciousness and quantum theory, can be solved simultaneously.
5. *Qualia Force*. Consciousness is an independent, nonreducible feature of physical reality that exists in addition to (and probably not derived from) the fields and particles of fundamental physics. This heretofore unknown aspect of the world may take the form of a new, independent, fundamental physical law or force (fifth force?).
6. *Qualia Space*. Consciousness is an independent, non-reducible feature of physical reality that exists in addition to the mass-energy and space-time of fundamental physics. This heretofore unknown as-

pect of the world may take the form of a radically new structure or organization of reality, perhaps a different dimension of reality (e.g., "qualia space" as postulated by "integrated information theory").

7. *Panpsychism*. Consciousness is a non-reducible feature of each and every physical field and particle of fundamental physics. Everything that exists has a kind of inherent "proto-consciousness" which, in certain aggregates and under certain conditions, can generate real inner awareness. Panpsychism is one of the oldest theories in philosophy of mind (going back to pre-modern animistic religions and the ancient Greeks). It is being revived, in various forms, by some contemporary philosophers in response to the seemingly intractable "hard problem" of consciousness.
8. *Dualism*. Consciousness requires a radically separate, nonphysical substance that is not only independent of the physical brain but also apart from the physical world. This would mean that reality consists of two, ontologically distinct parts—physical and nonphysical substances, divisions, dimensions or planes of existence. (The two distinct parts account for the origin of the term "dualism"). While human consciousness would require, under dualism, both a physical brain and a non-physical substance (somehow working together), following the death of the body and the dissolution of the brain, this nonphysical substance by itself could maintain some kind of conscious existence. (Though this nonphysical substance is traditionally called a "soul"—a term laden with theological burdens—a soul is not the only kind of thing, or form, that such a nonphysical substance could be.)
9. *Consciousness as Ultimate Reality*. The age-old claim, rooted in some wisdom traditions, is that the only thing that's really real is consciousness—everything else, especially the entire physical world and all it contains (including physical brains), is derived from an all-encompassing "cosmic consciousness." Each individual instance of consciousness—human, animal, robotic or otherwise—is a part of this cosmic consciousness. Eastern religions, in general, espouse this kind of view. (See Deepak Chopra for contemporary arguments that ultimate reality is consciousness.)

"Functionalism" is the theory in philosophy of mind that mechanisms are more important than mediums, that what's critical is how mental states work, not in what substrates mental states are found. As long as the activities (functions) are conducive to creating consciousness, it does not matter whether the substrates are neural tissue or computer chips or anything else that can support or enable the same activities (functions). As such, functionalism would apply to the categories 1, 2, 3 and 4 above, but not to categories 7, 8 and 9. (I'm not sure about 5 and 6, which, pending details, can be argued either way.)

Will Superstrong AI be Conscious?

I'm not going to evaluate each competing cause of consciousness. (That would require a course, not an essay.) Rather, for each potential cause, I assess the implications for virtual immortality, asking whether true first-person survival is in principle possible.

But first, to prepare a systematic analysis, I address the related but less complex issue of whether non-biological intelligences with superstrong AI (post-singularity) could be conscious and possess inner awareness. To the extent that the case for non-biological intelligences to be conscious can be made, the case for virtual immortality improves. To the extent that the case for non-biological intelligences to be conscious is weak, the case for virtual immorality is weaker. So for each cause of consciousness, could non-biological intelligences become conscious? The list follows.

If physicalism/materialism explains consciousness entirely (without remainder), then it would be almost certainly true that non-biological intelligences with superstrong AI would eventually have the same kind of inner awareness that humans do. Moreover, as AI would rush past the singularity and become ineffably more sophisticated than the human brain, it would likely express forms of consciousness higher than we today could even imagine.

If epiphenomenalism or non-reductive physicalism is true, then it would be highly likely that non-biological intelligences could eventually be conscious —though the increasing reality of mental states attenuates (slightly, unpredictably) the likelihood of inner awareness—an argument that is itself countered by functionalism (if functionalism is true).

A similar line of reasoning holds if quantum physics is the key to consciousness—with one difference being that the physical constraints of manipulating myriad quantum events, with their inherent indeterminacies, would seem even more daunting.

If consciousness requires an independent, non-reducible feature of physical reality—qualia force or qualia space—then it would remain an open question whether non-biological intelligences could ever experience true inner awareness. (It would depend on the deep nature of the consciousness-causing feature, the qualia force or qualia space, and whether

this feature could be controlled by technology.)

If panpsychism is true and consciousness is a non-reducible property of each and every elementary physical field and particle, then it would seem likely that non-biological intelligences with superstrong AI could experience true inner awareness (because consciousness would be an intrinsic part of the fabric of physical reality).

If dualism is true and consciousness requires a radically separate, nonphysical substance not causally determined by the physical world, then it would seem impossible that non-biological intelligences, no matter how superstrong their AI, could ever experience true inner awareness. (An exception might be in the extremely remote condition that somehow the physical actions of the brain could exert causal force on the supposed nonphysical substance.)

If consciousness is ultimate reality (cosmic consciousness), then anything could be (or is) conscious (whatever that may mean), including non-biological intelligences.

Remember, in each of these cases, no one could detect, using any conceivable scientific test, whether the non-biological intelligences with superstrong AI had the inner awareness of true consciousness. (They would claim to, of course, and do so convincingly.)

In all aspects of behavior and communications, these non-biological intelligences, such as cosmos-colonizing robots, would seem to be equal to (or superior to) humans. But if they did not, in fact, have the felt sense of inner experience, they would be "zombies" ("philosophical zombies" to be precise), externally identical to conscious beings, but there's no mental content, there's nothing inside.

This stark dichotomy between conscious and non-conscious entities spotlights (a bit circularly) our probative questions about self-replicating robots that, unless we destroy ourselves or our planet, will eventually colonize the cosmos. Post-singularity, will superstrong AI *without* inner awareness be in all respects as powerful as superstrong AI *with* inner awareness, and in no respects deficient? That is, are there kinds of cognition that, in principle or of necessity, require true consciousness? The answer could affect what it means to colonize the cosmos.

Moreover, would true conscious experience and inner awareness in these galaxy-traversing robots represent a higher form of intrinsic worthiness, some kind of absolute, universal value (however anthropomorphic this may seem)? For assessing the fundamental nature of robotic probes colonizing the cosmos, the question of consciousness is profound.

Is Virtual Immortality Possible?

Can the fullness of our first-person mental selves (our "I") be digitized and uploaded perfectly to non-biological media so that our mental selves can live on beyond the death of our bodies and the destruction of our brains? Whether virtual immortality is even possible has never changed, of course; always it has been determined by the unchanging cause of consciousness. It's just that there is no consensus on what that cause actually is. Let's assess each of the nine alternatives with this question in mind.

1. *Physical/materialism*. If physicalism/materialism explains consciousness entirely (without remainder), then our first-person mental self would be (almost certainly) uploadable and virtual immortality would be attainable. The technology might take hundreds or thousands of years—not decades, as optimo-techno-futurists predict—but, barring human-wide catastrophe, it would happen.
2. *Epiphenomenalism*. If epiphenomenalism is true, then it is highly likely that some kind of virtual immortality would be attainable. The inert "foam" of consciousness should have little impact.
3. *Non-reductive Physicalism*. If non-reductive physicalism is true, then it is also highly likely that some kind of virtual immortality would be attainable. The causative power of mental states should not affect virtual immortality because a perfect duplication of the physical states would *ipso facto* produce a perfect duplication of the mental states.
4. *Quantum Consciousness*. If quantum consciousness is true, then it is likely that some kind of virtual immortality would be attainable. However, the indeterminacies, probabilistics and strangeness of quantum physics add a degree of uncertainty that cannot as yet be evaluated. The test, as with all potential causes of consciousness, is whether advanced technology can manipulate and control the cause of consciousness, and do so comprehensively and precisely. The quantum nature of consciousness, if true, would introduce unpredictability and perhaps undermine perfect duplicability.
   Note: The theory of functionalism would support virtual immortality for categories 1, 2, 3 and 4 above.
5. *Qualia Force*. If consciousness requires an independent, non-reducible feature of physical reality that may take the form of a new, independent, fundamental physical force, then it would be possible but remain an open question whether our first-person mental self could be uploadable. Virtual immortality would be less

likely with Qualia Force than it would be in 1, 2, 3 and (probably) 4 above, because not knowing much about this consciousness-causing new force, we would not know whether it could be manipulated by technology, no matter how advanced. But because consciousness would still be physical, efficacious manipulation and successful uploading would seem possible.

6. *Qualia Space*. If consciousness requires an independent, non-reducible feature of physical reality that may take the form of a radically new structure or organization of reality, perhaps a different dimension of reality (e.g., as postulated by "integrated information theory"), then virtual immortality would be possible, but it would be remain an open question whether our first-person mental self could be uploadable. Not understanding this consciousness-causing feature, we could not now know whether it could be manipulated by technology, no matter how advanced. If this qualia space could be in some sense directed by activities in the brain, with predictable regularities, then virtual immortality would be more likely.
7. *Panpsychism*. If panpsychism is true and consciousness is a non-reducible property of each and every elementary physical field and particle, then it would seem probable that our first-person mental self could be uploadable. There would be two reasons: (i) consciousness would be an intrinsic part of the fabric of physical reality, and (ii) there would probably be regularities in the way particles would need to be aggregated to produce consciousness, and if there are regularities, then advanced technologies could learn to control them.
8. *Dualism*. If dualism is true and consciousness requires a radically separate, nonphysical substance not causally determined by the physical world, then it would seem impossible to upload our first-person mental self by digitally duplicating the brain, because a necessary cause of our consciousness, this nonphysical component, would be absent. (An exception might be in the extremely remote case that somehow the physical actions of the brain could exert causal force on the supposed nonphysical substance.)
9. *Consciousness as Ultimate Reality*. If consciousness is ultimate reality, then consciousness would exist of itself, without any physical prerequisites. But would the unique digital pattern of a complete physical brain (derived, in this case, from consciousness) favor a specific segment of the cosmic consciousness (i.e., our unique first-person mental self)? It's not clear, in this extreme case, whether uploading would make much difference (or much sense).

Whereas most neuroscientists assume that whole brain duplication can achieve virtual immortality, Giulio Tononi is not convinced. According to his theory of integrated information, "what would most likely happen is, you would create a perfect 'zombie'—somebody who acts exactly like you, somebody whom other people would mistake for you, but you wouldn't be there."

So, in pursuit of virtual immortality, would a perfect digital duplication of a human brain generate first-person consciousness? Here are my (tentative) conclusions for each alternative: 1, surely; 2 and 3, highly likely; 4, somewhat likely; 5 and 6, possibly but uncertain; 7, probably; 8, no; 9, doesn't matter.

The Trouble with Duplicates

In trying to distinguish among these alternative causes of consciousness, and thus assess the viability of virtual immortality, I am troubled by a simple observation. Assume that a perfect digital duplication of *my* brain does, in fact, generate *my* first-person consciousness—which is the minimum requirement for virtual immortality. This would mean that my first-person self and personal awareness could be uploaded to a new medium (non-biological or even, for that matter, a new biological body). But here's the problem: If "I" can be duplicated once, then I can be duplicated twice; and if twice, then an unlimited number of times.

What happens to my first-person inner awareness? What happens to my "I"? Assume I do the digital duplication procedure and it works perfectly—say, five times. Where is my first-person inner awareness located? Where am I? Each of the five duplicates would state with unabashed certainty that he is "Robert Kuhn," and no one could dispute any of them. (For simplicity of the argument, physical appearances of the clones are neutralized.) Inhabiting my original body, I would also claim to be the real "me," but I could not prove my priority. (See David Brin's novel *Kiln People*, a thought experiment about "duplicates," and his comments on personal identity.)

I'll frame the question more precisely. Comparing my inner awareness from right before to right after the duplication process, will I feel or sense differently? Here are four duplication scenarios, with their implications:

1. I do not sense any difference in my first-person awareness. This would mean that the five duplicates are like super-identical twins—they are independent conscious entities, such that each, after his cre-

ation, begins instantly to diverge from the others. This would imply that consciousness is the local expression or manifestation of a set of physical factors or patterns. (An alternative explanation would be that the duplicates are zombies, with no inner awareness—a charge, of course, they would deny.)

2. My first-person awareness suddenly has six parts—my original and the five duplicates in different locations—and they all somehow merge or blur together into a single conscious frame, the six conscious entities fusing into a single composite (if not coherent) "picture." In this way, the unified effect of my six conscious centers would be like the "binding problem" on steroids. (The binding problem in psychology asks how our separate sense modalities like sight and sound come together such that our normal conscious experience feels singular and smooth, not built up from discrete, disparate elements). This would mean that consciousness has some kind of overarching presence or a kind of supra-physical structure.
3. My personal first-person awareness shifts from one conscious entity to another, or fragments, or fractionates. These states are logically (if remotely) possible, but only, I think, if consciousness would be an imperfect, incomplete expression of evolution, devoid of fundamental grounding.
4. My personal first-person awareness disappears upon duplication, although each of the six (original plus five) claims to be the original and really believes it. (This, too, would make consciousness even more mysterious.)

Suppose, after the duplicates are made, the original (me) is destroyed. What then? Almost certainly my first-person awareness would vanish, although each of the five duplicates would assert indignantly that he is the real "Robert Kuhn" and would advise, perhaps smugly, not to fret over the deceased and discarded original.

If Virtual Immortality, Then Colonize the Cosmos?

There's a further implication, and an odd one at that. Assuming that our superstrong AI, cosmos-colonizing robots could become conscious, I can make the case that such galaxy-traveling, consciousness-bearing entities could include you—yes you, your first-person inner awareness, exploring the cosmos virtually forever.

Here's the argument. If virtual immortality and superstrong AI consciousness are possible—and there is high correlation between the two—then human personality can be uploaded (ultimately) into space probes and we ourselves can colonize the cosmos!

I'd see no reason why we couldn't choose where we would like our virtual immortality to be housed, and if we choose a cosmos-colonizing robot, we could experience the galactic journeys through robotic senses (while at the same time enjoying our virtual world, especially during those eons of dead time traveling between star systems).

Would I Take the Plunge?

At some time in the (far) future, scientists will likely assure us that the technology is up and working. If I were around, would I believe the scientists and upload my consciousness? Moreover, entranced by what I assume will be commercial advertisements, would I select a cosmos-colonizing robot as my medium of storage so that I could spend my virtual immortality touring the galaxy? I might, if only because I'd be confident that duplication possibility 1 (above) is true and 2, 3 and 4 are false, and that the duplication procedure would not affect my first-person mental self one whit. (I sure wouldn't let them destroy the original, though the duplicates may call for it.)

So while all the duplicates wouldn't feel like me (as I know me), I'd kind of enjoy sending "Robert Kuhn" out there exploring star systems galore. (There's more. If my consciousness is entirely physical and can be uploaded without degradation, then it can be uploaded without degradation to as many cosmos-colonizing robots as I'd like—or can afford. It gets crazy.)

Whether non-biological entities such as robots can be conscious, or not, presents us with two disjunctive possibilities, each with profound consequences. If robots can never be conscious, then there may be a greater moral imperative for human beings to colonize the cosmos. If robots can be conscious, then there may be less reason for humans, with our fragile bodies, to explore space—but your personal consciousness could be uploaded into cosmos-colonizing robots, probably into innumerable such galactic probes, and you yourself (or your mental clones) could colonize the cosmos.

My intuition, for what it's worth, is that it's a pipedream. I deem virtual immortality for my first-person inner awareness to be not possible, and to be never possible, though in the (far) future duplicates may convince us otherwise. But confident in my conclusion, I am not.

For me for now, I'm convinced of only this: Virtual immortality, like the AI singularity, must confront the deep cause of consciousness. S