The Journal of Mind and Behavior
Summer 2025, Volume 46, Number 3
Pages 60–84
ISSN 0271–0137

The Incompatibility of Special Relativity and Conscious State Duplication: A Logical Argument Against Computational Theories of Mind

Andrew Knight *
Brain-in-a-Vat Research

The possibility of algorithmic consciousness depends on the assumption that conscious states can be copied or repeated by sufficiently duplicating their underlying physical states. Whether or not this is physically possible has profound implications for physics, biology, computer science, and philosophy. By assuming that a conscious state supervenes on some sufficient underlying physical state, I show that the existence of the transtemporal identity commonly associated with consciousness is logically incompatible with the ability to copy or repeat conscious states. This incompatibility arises from relativistic constraints on spacelike and timelike separated physical instantiations of a conscious state. This result has significant implications for strong artificial intelligence, mind uploading, teleportation, Boltzmann brains, and other philosophical problems.

Either it is physically possible to copy or repeat a conscious state or it is not. Both possibilities have profound implications for physics, biology, computer science, and philosophy.

Whether or not consciousness can be copied is fertile ground for a multitude of troubling, if not fascinating, thought experiments. There's the duplication problem (Penrose, 1989): imagine we can teleport a traveler to another planet by creating "a precise duplicate of the traveler, together with all his memories, his intentions, his hopes, and his deepest feelings," but then we decide not to destroy the original copy? "Would his 'awareness' be in two places at once?" There's the

* I would like to thank Scott Aaronson, Kenneth Augustyn, Martin Bier, Gary Ihas, Annie Knight, Christopher Lind, Kyle Padgett, Brian Pitts, Igor Salom, and Paul Tappenden for their guidance, comments, and stimulating conversations. Correspondence concerning this article should be addressed to Andrew Knight, J. D, Director, Brain-in-a-Vat Research, 3104 Copperfield Rd. Greenville, NC 27834. URL: https://www.inavat.com/ E-mail: aknight@alum.mit.edu

simulation problem (Bostrom, 2003): if conscious awareness can be uploaded onto a computer, then how do we know we aren't simulated minds in simulated universes? In fact, because simulated universes are much less expensive, in terms of matter and energy, than actual universes, then if consciousness can be duplicated, we are almost certainly one of vast numbers of simulated copies. There's the problem of self-location (Elga, 2004): if a psychopath tells you that he has created an exact physical copy of you and will torture it unless you pay a hefty ransom, then you should pay the ransom unless you are absolutely sure that *you* aren't the copy. There's the problem of the Boltzmann brain (Carroll, 2017): if consciousness is just the result of atoms in a brain, what's to prevent a set of physically identical atoms, somewhere in the universe, from accidentally coming together in just the right way to create your brain? And what would that feel like?

Advances in artificial intelligence have renewed interest in questions of machine consciousness, computational theories of mind, and physical constraints on consciousness. Unlike purely philosophical approaches (Searle, 1990), I adopt a methodology that examines logical constraints imposed by physical theories we already accept — specifically special relativity (or, more broadly, physical limitations on the transfer of information). The incompatibility between relativity and conscious state copiability that I will identify in this paper provides a novel argument that complements other work demonstrating the potential non-computability of consciousness (Penrose, 1989) and physical constraints on computational theories of consciousness (Stoica, 2020).

Notice that each of the above problems (duplication, simulation, self-location, and the Boltzmann brain) is a direct consequence of the *copiability* or *repeatability* of conscious states.[1] If it turns out, for whatever reason, that conscious states cannot be copied, then these and other problems disappear. Given the rate that technology is advancing, it may be just a matter of time before each of the above scenarios is empirically tested. There are related problems that may never be testable such as, "Will a computer ever become conscious — and how would we know?" After all, there is no consensus on how to measure the existence or level of consciousness in an entity. Some say that consciousness depends on the ability to pass a hypothetical Turing test. Some say it depends on the level of complexity or information integration in neural networks. Some say it depends on certain brain activity. So how can we possibly learn anything about the nature of consciousness if it depends on a definition?

---

[1] Throughout this paper, I'll treat copiability (alternative spelling: copyability) and repeatability of conscious states, including the ability to physically reset a conscious state to an earlier state, as meaning the same because repeating a conscious state is akin to copying it at a later point in spacetime.

Further, just as there is no consensus on the definition of consciousness, there is no single term or phrase to represent it; others are mentality, sentience, self-awareness, and the subjective experience of qualia.

It is easy to get bogged down in the meaning of "conscious" and lose sight of the big picture. It's as if the question, "Will a ball near Earth experience a force of gravity?" has been preempted by, "What do you mean by 'ball'?" So let me rephrase: "If I am near Earth, will *I* experience a force of gravity?" My goal is to derive, if possible, an objectively correct prediction to several questions such as: will it ever be possible to upload my conscious awareness onto a computer so that I can outlive my physical death? Will it ever be possible to teleport a copy of myself to another planet? Will it ever be possible for a collection of atoms somewhere in the universe to accidentally come together in just the right way to create my conscious awareness? Will it be possible for me to experience these things? The answer is yes only if my conscious states are fundamentally such that they could be copied or repeated. Are they?

First, let's consider the question of whether it will ever be possible to upload me to a digital computer. To phrase it scientifically in the form of a falsifiable hypothesis, "It is impossible to upload me to a digital computer" would be falsified, for example, by my being conscious and aware while possessing a body consisting essentially of digital switches. But what do I mean by "me"? I mean my identity, but what is that? Questions like this lead many to reject from the realm of scientific inquiry anything involving identity or consciousness. But that's a mistake because science is about making predictions about our own experiences, and without a consistent sense of identity, *whose* experiences are we predicting?

In this paper I am going to argue that it will *not* be possible to upload me to a digital computer, among other conclusions. Here is a brief introduction to the structure of the argument.

I am going to assume, first, that my consciousness is created by some sufficient underlying physical state. To a physicalist (and perhaps most physicists), this is obviously true: if the "underlying physical state" includes the state of the entire universe, then clearly it creates my consciousness. But what I want to know is whether conscious states can be recreated by copying their underlying physical states, so I'm going to assume that there is some underlying physical state, significantly smaller and presumably easier to copy than the entire universe, that is by itself sufficient to create my consciousness. If so, then my consciousness is

necessarily created by a small local[2] physical state enclosed by some spacetime region such that my conscious awareness entirely depends on physical stuff inside that region and is independent of anything outside.

How small might such a physical state be? Let's not forget the original thought experiments that led to the philosophical problems outlined earlier: duplication of conscious states by copying one's body or brain. I think most physicalists would accept, then, that my consciousness is created by some relatively small physical state, a region in spacetime that encloses, for example, my brain. But if that's true, if within the confines of my brain is a physical state that is by itself sufficient to produce my consciousness, then if it were copied and instantiated elsewhere, such as in a brain-in-a-vat experiment on Mars, wouldn't that also produce *my* consciousness? Wouldn't I experience a conscious awareness due to both brains? And wouldn't the diverging conscious awareness of the brains (resulting, for example, from their different physical surroundings) be logically problematic?

One might try to circumvent this logical awkwardness by insisting that I wouldn't experience the consciousness of the brain-in-a-vat on Mars, that the brain-in-a-vat would be a separate but fully conscious person who thinks and acts like me and has my memories (as of the time of duplication), but also would not experience what I experience here on Earth. Or, perhaps, that I would experience the conscious awareness of the brain-in-a-vat but not that of the original brain on Earth. In other words, either I experience the Earth me or the Mars me, not both. But here's the problem: if the brain state is truly sufficient to produce my consciousness, then it cannot contain any information that would differentiate whether my conscious identity "flowed" through the Earth brain or the Mars brain. That implies only two possibilities: (1) the chosen physical state (in this case, the brain state) was not actually sufficient to create my consciousness, in which case we can simply limit ourselves to physical states that are adequate to create my consciousness and then determine the extent to which special relativity limits the existence of multiple copies of those physical states; or (2) there is no such thing as conscious identity or, if there is, it does not flow or persist through time. I will address possibility (1) in the section entitled, "Why Conscious States Cannot Be Copied." Regarding possibility (2), I will devote the section entitled "Questioning Identity" to a discussion of why objections based on identity are inadequate to counter the conclusions of this paper. Strictly speaking, the analysis in this paper need not depend on special

---

[2] I use the word "local" loosely here without commenting on the size or shape of that spacetime region or whether the resulting conscious state depends on relativistically nonlocal effects, superluminal signaling, or quantum entanglement within that region.

relativity to the extent that we can instead speak in terms of sufficient underlying physical states whose physical evolutions are bounded by regions of spacetime that are nonlocal to each other. Nevertheless, the use of special relativity clarified the arguments for me and will, I hope, similarly serve the reader.

Recent work at the intersection of philosophy of mind and physics has explored various constraints that physical theories may place on our understanding of consciousness, such as quantum mechanical constraints (Hameroff & Penrose, 2014). Most relevant to the present argument, Stoica (2020) presents a logical demonstration that mental states must be nonlocal, arguing that consciousness cannot be reduced to classical computation. While Stoica's approach differs methodologically, it reaches complementary conclusions about the non-computational nature of consciousness.

This paper contributes to this growing literature by introducing a novel argument based on the logical incompatibility between special relativity and the copiability of conscious states. Unlike approaches that begin with speculative assumptions about consciousness, I examine the constraints imposed by well-established physical theories when applied to conscious states with transtemporal identity.

Questioning Identity

Let's consider what it means to upload me to a digital computer. I think it means two things. First, that it's me, that I experience it. (Let's call this *existence* of identity.) If a scientist tells me that an exact copy of me has been uploaded to a computer, but I don't experience whatever sensations accompany living inside a computer, I will respond that I have not been uploaded. "But I can't tell the difference between you two," the scientist may respond. "Yes, but I can," I would reply, "and you have failed to upload me." Second, that my identity would persist; that as the computer physically evolves in time according to the laws of nature, the consciousness it creates would continue to be mine. (Let's call this *persistence* of identity.) In other words, if the computer begins in some initial physical state S_1 that produces me in conscious state C_1, and after a while physically evolves to physical state S_2 that happens to produce conscious state C_2, that conscious state C_2 will still be mine, and I subjectively experience an evolution from conscious state C_1 to C_2. And if, due to a random quantum event, the computer were to evolve to physical state S_2' (instead of S_2) that happens to produce C_2' (instead of C_2), then conscious state C_2' will still be mine and I subjectively experience an evolution from conscious state C_1 to C_2'. I don't expect the scientist to guarantee that I will live forever — after all, the physical evolution of the computer could simply stop producing consciousness — but I

do expect that as long as the physical evolution of the computer continues to produce conscious states, *I* will be the person experiencing them, not someone else. In other words, if a scientist claims to have uploaded me to a digital computer, if I don't subjectively experience (in the form of evolving conscious states) the physical evolution of the computer over time, then that scientist has failed. Let me combine existence and persistence of identity into one concept: temporal continuity[3] of identity (or *transtemporal identity*). Whether there really is such a thing, and whether it matters, will be addressed in the following sections.

I will argue in this section that consciousness includes transtemporal identity, an assumption I'll explicitly include in a later section as "transtemporal identity of conscious states." To elaborate on this assumption, a conscious state is a state experienced by *someone*; that is, a person (whether instantiated in human or other biological or nonbiological form) experiencing a conscious state has an identity and, consequently, one conscious state cannot be experienced by more than one person. This does not imply that two people, Alice and Bob, couldn't experience a conscious state that seemed or felt identical or nearly identical. Rather, Alice experiencing that conscious state would experience it as Alice — i.e., the conscious state experienced by Alice is identified with Alice — and she wouldn't momentarily confuse herself with Bob or experience the state as Bob. So if a particular physical state produces a conscious state, and that physical state includes (among other things) light sensors that are absorbing photons having a wavelength of 680nm, then the produced conscious state may correspond to someone's subjective experience of seeing the color red, in which case we can ask *who* is experiencing that conscious state and whether that person is seeing red. The following subsections detail several arguments in support of the claim that conscious states are associated with transtemporal identity.

Boltzmann Brains and Bad Science

The Boltzmann brain problem is much bigger than the weirdness of the occasional human brain fluctuating in and out of existence. In cosmological models in which Boltzmann brains dominate, not only is any brain much more likely to be a random fluctuation, but any complete person, world, or even galaxy is much more likely to be a random fluctuation. Are you a Boltzmann brain? You might think the answer is no because you can see a vast world around you and you seem to have

---

[3] This temporal flow is often called a "stream" of consciousness, but many have argued that consciousness consists of a series of discontinuous blips of awareness that are perceived as continuous. This distinction is irrelevant to the present analysis.

many years' worth of memories that are consistent with a long-lasting, physical-law-abiding world. However, the overwhelming majority of people who would make such claims are, in these cosmological models, random fluctuations. They are "Boltzmann observers" who have no right to believe that any of their observations are reliable or that their scientific predictions will come to fruition.

This is a serious problem. Science, after all, is about collecting evidence and making predictions based on falsifiable hypotheses. However, if a model ostensibly based on science predicts that the foundation of science itself is not reliable—then what? Carroll (2017) eloquently sums it up: "it's overwhelmingly likely that everything we think we know about the laws of physics, and the cosmological model that predicts we are likely to be random fluctuations, has randomly fluctuated into our heads." In other words, if a model that is based on physical evidence in the world around us predicts that that very evidence is illusory, then we should reject such a model as "cognitively unstable."

A similar fate befalls denial of transtemporal identity because the foundations of science fundamentally depend on it. "A ball falls toward Earth at $9.8 m/s^2$" is a hypothesis that, if incorrect, can be falsified through experimentation. But that requires a scientist to prepare the experiment, then perform it, then collect and analyze data, and then determine whether the hypothesis has been falsified. This series of events, all of which are necessary in scientific investigation, occurs in a temporal order, and if the scientist did not or could not expect to be the *same* person throughout, then she would have no reason to believe that she could test falsifiable hypotheses. She would have no reason to trust science. Therefore, any theory or model that rejects the temporal continuity of identity should be rejected as rendering science internally inconsistent. Further, while transtemporal identity is a necessary ingredient to scientific inquiry, it is not itself subject to scientific scrutiny because *no experiment can falsify it*. Therefore, there is no scientific rationale to doubt that consciousness is associated with transtemporal identity.

Fissioning Philosophers

Philosophers have had much to say about both the existence and persistence of identity, but their analyses almost exclusively depend on the assumption that conscious states can be copied. The typical philosophical analysis of identity goes something like this: *Identity is produced by the brain; the brain is just a chunk of matter that can be moved around, copied, replaced, and so forth, at will; therefore, the notion of identity is problematic.*

For instance, Parfit (1971) begins his analysis with, "We suppose that my brain is transplanted into someone else's (brainless) body, and that the resulting person has my character and apparent memories of my life. [I shall here assume] that the resulting person is me." He then proceeds, in thought experiment, to slice up the brain, trade brain sections with other people, sever

connections between hemispheres, and so forth, to arrive at various conclusions (such as the "fission" or "fusion" of identities) that are at odds with transtemporal identity. Zuboff (1990) also starts with replication of a brain that "is a precise duplicate of yours in every discriminable respect..." and concludes that so-called fission — in which "one subject *can,* in a single next moment, experience two differing non-integrated contents of experience" — is the necessary outcome. These "brain-in-a-vat" style thought experiments have dominated philosophical theories of mind and identity, not to mention science fiction, for decades, and represent the best *prima facie* arguments against transtemporal identity.

However, an argument that relies on premise *X* cannot logically be used to disprove an argument for *not X* (written ¬*X*). My goal in this paper is to show, among other things, that conscious states cannot be copied. Because the typical philosophical arguments depend on the in-principle copiability of brains, mental states, etc., then not only are they inadequate to counter the arguments in this paper, they are entirely logically inapplicable. In other words, the strongest arguments against transtemporal identity depend on the assumption that conscious states can be copied, but this is the very notion that I am attempting to refute. Therefore, any notion or theory of identity that depends on the copiability of conscious states (or the copiability of brains in conjunction with the assumption that brains produce consciousness) is simply irrelevant to the present analysis as impermissibly circular.

Universal Belief

Why fight the notion of transtemporal identity in the first place? After all, it's what we all implicitly believe. Indeed, the universality of the belief in transtemporal identity is itself an indicator of the significant empirical evidence we have to support it. Every decision we make about the future depends on our belief in identity. Let's say that a travel agent was to offer you the following options:

1. For $10,000, you will spend a week at a resort in Tahiti;
2. For $6,000, an exact physical copy[4] of you will be created and will spend a week at a resort in Tahiti while your current body is anesthetized;

[4] I have no idea what this phrase means, nor does any other physicist, despite its ubiquitous use in the philosophy literature. An exact copy in classical phase space is impossible because infinite precision does not exist in the physical world; an exact copy in quantum mechanical Hilbert space is impossible because of quantum no-cloning (Wootters & Zurek, 1982); and so on. But for the sake of this section, let's pretend there is such a thing.

3. For $2,000, a really good copy of you — one that looks, acts, and talks just like you but is not an exact copy — will be created and will spend a week at a resort in Tahiti while your current body is anesthetized.

In all three cases, the underlying question you'll no doubt ask yourself is, "What will *I* experience?" If you believe that an exact physical copy of you inevitably includes your identity, then you might be tempted to go with the second option, given that the $4,000 in savings could be enjoyed by you at a later date. And if you believe that a reasonably good copy of you includes your identity, then you might be tempted by the third option. In other words, your potential choice between the latter two options depends on (your beliefs about) whether your identity can flow to a copy of you, not on whether you have an identity. When choosing, you are, I presume, not thinking about the neural firings corresponding to pleasure sensations in a brain that is similar or identical to yours; you are thinking about whether *you* will experience those pleasure sensations.

Not only is universal belief in transtemporal identity evidenced by the choices we make, that belief seems to be constantly supported by consistent observations. When I choose to eat an apple on the expectation that it will be the same me who, moments later, tastes an apple, I am also not surprised, moments later, that I am tasting an apple, given my belief that it was the same me, moments earlier, that decided to eat an apple. We each experience thousands of similar confirmations every day that our conscious identities endure over time. Don't these confirmations serve as extremely compelling empirical evidence that consciousness includes transtemporal identity? We can reject that evidence, of course, but at the exorbitant cost of rejecting the very foundations of scientific investigation.

Much Ado About Nothing

The most likely objections to the arguments in this paper will target transtemporal identity. But even if I haven't convinced you that science without identity is unscientific, or that philosophy without identity is logically inapplicable to this paper, or that belief in transtemporal identity is universal, it does not much matter because the issue of identity is a red herring.

One goal of this paper is to determine whether or not conscious states can be copied, where a particular conscious state is identified with a particular person. That is, I want to answer questions like, "Is it possible in principle to upload me to a computer?" or "Is it possible to teleport me to a distant galaxy?" To do so, I will assume that a person's conscious state, where that conscious state corresponds to that person's transtemporal identity, is created by some sufficient underlying physical state, and then show that the existence of multiple copies instantiated nonlocally in

spacetime leads to a contradiction. But if a person's conscious state does *not* correspond to a transtemporal identity (whether because identity is an illusion, or identity fissions, or whatever), then we can avoid a logical contradiction but nevertheless come to the related conclusion that there is nothing physical that can be copied that will produce a conscious state of *that person*.

Stated bluntly: if there is something physical that can be copied that causes *me* to be uploaded to a computer, then what are the consequences? (That is a question I will explore in the following section.) But if there is *not* something physical that can be copied that causes *me* to be uploaded to a computer, then we have the answer: (my) conscious states cannot be copied.

Why Conscious States Cannot Be Copied

Assumptions

In this section, I'll lay out the structure of my arguments as well as some of the assumptions I make. I'll start with a generalization of various questions broached earlier:

A scientist claims that he is able to upload me to a digital computer. Do we have reason to doubt his claim?

Let's assume that the scientist's claim is true and then follow it to its logical consequences. First, he is claiming to upload *me*. If he is right, then when he uploads me, I will subjectively experience being conscious and aware yet embodied not by a human body but by a collection of physical objects (such as digital switches) configured as a general purpose computer executing software code. Next, because software is simply a set of instructions that can inherently be copied and executed on any general purpose computer, what happens if he uploads me to a second computer? Will I subjectively experience both? What would happen if those two experiences, responding to different stimuli or inputs, start to diverge? It might seem easy to solve the problem by adding an *ad hoc* requirement that only the first software copy can be executed, but how could Mother Nature possibly enforce this rule if the two computers are spacelike separated — i.e., causally nonlocal — so that the temporal ordering of events is relative to an observer?

Now, instead of software, the scientist claims that he is able to create a copy of me in a different spacetime position. He is not merely claiming that he is able to make a reasonably good physical copy of my brain, or an exact copy of my brain, or a reasonably good physical copy of my body, or an exact copy of my body. Rather, he is claiming to be able to copy me, identity and all. The same problems arise as in the case of uploading me to a computer. I will argue that the demands of special relativity are inconsistent with the ability to freely copy

a conscious state associated with a transtemporal identity because the following three assumptions lead to a contradiction:

1. **Supervenience of conscious states on sufficient physical states ("Supervenience"):** A conscious or mental state C_1 supervenes on sufficient physical state S_1 such that instantiation of physical state S_1 is sufficient to create conscious state C_1 of a person.
2. **Transtemporal identity of conscious states ("Identity"):** Conscious state C_1, created by sufficient physical state S_1, is a conscious state of a person having an identity. If physical state S_1 physically evolves over time to state S_2 that is sufficient to create conscious state C_2, then state C_2 will be the same person (i.e., having the same identity) as state C_1.
3. **Copiability of conscious states ("Copiability"):** A copy of physical state S_1, which is sufficient to create conscious state C_1, can be instantiated elsewhere in spacetime in a manner that does not prevent other instantiations from evolving.

Note that if state S_1 depends on spatiotemporal location — i.e., if that state cannot be instantiated elsewhere in spacetime — then the assumption of Copiability is trivially false. To use the philosopher's lexicon, two physical systems are not "numerically identical" if they exist at different locations in spacetime, no matter how physically similar they may be. So if one already accepts that numerical identity is necessary for the same person to experience the same conscious state, then it follows that conscious states associated with transtemporal identity cannot be copied or repeated.

Note also that Copiability depends on Supervenience; Copiability can be false on its own, but if Supervenience is false, then Copiability must also be false because if there is no physical state the instantiation of which is sufficient to create a certain conscious state, then there is nothing physical that can be copied and instantiated elsewhere to produce that conscious state. In other words, ¬Supervenience → ¬Copiability. The same is true of the Identity assumption as written; it can be false on its own, but ¬Supervenience → ¬Identity. (A dualist or non-physicalist approach to consciousness, which would deny Supervenience, would not assume that transtemporal identity of conscious states depends on the sufficiency of underlying physical states.) Therefore, logically, if the conjunction of these three assumptions leads to a contradiction, then either Identity or Copiability (or both) is false.

The phrase "physical state" is meant in the broadest sense possible. Classically, the physical state of a system of matter might be fully described in phase space — i.e., in terms of the positions and momenta of each of its constituent particles. However, we know that the classical description of the universe is only an approximation and fails to correctly describe and predict extremely small systems,

the purview of quantum mechanics. Quantum mechanically, however, it can't be said that any particular particle in a system even *has* a position or momentum unless and until measured; and while its position can be measured to nearly arbitrary precision, quantum uncertainty guarantees a trade-off in the precision to which its momentum can be simultaneously measured. Further, the quantum mechanical description of a particle can't be separated from that of other particles with which it is entangled, making the quantum mechanical prediction of a physical system larger than a few hundred atoms essentially impossible. Quantum mechanics may itself turn out to be an emergent approximation of a more fundamental ontology.

Therefore I do not intend to limit "physical state" to any known description. Instead, by assuming that conscious state C_1 supervenes on sufficient physical state S_1, I simply mean that there is something physical about the universe that gives rise to that conscious state. Of course, there's no guarantee that a given physical state will create a conscious state — probably very, very few will — but if a conscious state exists then it depends entirely on some underlying physical state. I also assume that a given conscious state might be created by more than one physical state. It may be the case that conscious state C_1 arises from any member of some large set $\{S_1, S_1^*, S_1^{**}, ...\}$ of underlying physical states, and that instantiating any of these physical states will produce the same conscious state C_1. And because state S_1 is sufficient to create state C_1, any larger physical state $S > S_1$ that includes state S_1 is also sufficient to create state C_1.

The "Supervenience" assumption may be rejected by dualists but should seem reasonable to physicalists and materialists. For any conscious state, there is a sufficient physical state whose instantiation guarantees creation of the conscious state without regard to information, influence, or causation beyond or nonlocal to that physical state. While the arguments in this paper apply to a sufficient physical state of any size, they are more elucidating when applied to a small or minimal (but still sufficient) physical state S_1 that produces conscious state C_1, although we need not limit ourselves to the absolute minimum or smallest state S_1 that produces the person's consciousness. For instance, if it turns out that the human brain[5] is sufficient to create a person's consciousness, then certainly that brain *plus* a body would also create the person's consciousness, as would that brain plus a body plus Earth plus the Andromeda galaxy. But if our goal is simply to create the person's consciousness as simply and efficiently as possible, we may as well focus on just the brain. So when I assert that conscious state C_1 is created by sufficient physical state

[5] Specifying physical state S_1 as "the human brain" is quite sloppy because we really need to specify the features (e.g., relationships among cells, molecules, electrons, etc.) that define state S_1 and give rise to conscious state C_1. If, for example, state S_1 depends only on how certain cells act like digital switches, then S_1 can be specified and instantiated in the form of a digital computer. If, however, state S_1 depends on quantum effects, then quantum no-cloning may prevent specification of S_1 at all.

S_1, I don't necessarily know what kinds of information or physical attributes specify that physical state S_1. I also don't know how big state S_1 must be so that it is sufficient to create state C_1. Is it the local state of certain neural connections in one's brain? The state of one's entire brain? One's body? The planet? The universe? We don't need to know any of these things to accept the assumption that for a given conscious state, there is some underlying physical state that is sufficient to create it. Although, if it turns out as a matter of fact that the smallest physical state S_1 sufficient to create a conscious state C_1 happens to involve particles throughout the entire universe, then clearly that would prevent state C_1 from being physically copied.

The "Identity" assumption was discussed at length previously. Further, as pointed out above, Identity depends on Supervenience. In combining Supervenience and Identity into a conjunction assumption ("S∧I"), we must be careful to limit the following analysis to underlying physical states that are sufficient to create a conscious state of a person *having a transtemporal identity.* If there is a physical state S_1 that satisfies Identity, then it also satisfies Supervenience. In other words, if S∧I is true, then instantiation of that physical state S_1 will necessarily create conscious state C_1 of a person having a transtemporal identity. If instantiation of some arbitrary physical state S_0 does *not* necessarily create a conscious state of a person having a transtemporal identity, then either Identity is false or the selected physical state S_0 is not sufficient, in which case a larger or different physical state should be chosen that *is* sufficient to create such a conscious state. So, in the following analysis in which S∧I is assumed true, physical states are chosen to guarantee that their instantiation creates conscious states associated with transtemporal identity.

Finally, the "Copiability" assumption inherently assumes that state S_1 is smaller than all of spacetime, but more importantly that it is small enough that instantiation "elsewhere" in spacetime does not immediately causally interfere with the natural physical evolution of other instantiations. We can loosely think of state S_1 as being a "local" state enclosed by some surface in spacetime, and conscious state C_1 as being created by that local physical state S_1. Because state S_1 is sufficient to create state C_1, it will do so even if other nonlocal instantiations of state S_1 also create state C_1.

Spacelike Separated Copies

I will now show how the assumptions of Supervenience and Identity (S∧I) conflict with Copiability. With reference to Figure 1, a sufficient physical state S_1 on which conscious state C_1 supervenes exists at a Point 1 in spacetime. By S∧I, a conscious person having an identity determined by state S_1 exists at Point 1 — let's call her Alice and designate the person at Point 1 as $Alice_1$. For simplicity, assume that the information specifying state S_1 is read and then the state is destroyed or reconfigured soon afterward so that state C_1 no longer exists. (This is not a requirement but does simplify the analysis. If it turns out that the mere existence of the information necessary to produce state S_1 is itself adequate to produce state C_1, or if a physical state

fundamentally is information, then it is impossible to destroy S_1 without also preventing any future instantiations of C_1.) The light cone of Point 1 is shown; events outside the light cone are called *spacelike* (or nonlocal) and cannot have any cause-effect relationship, temporal relationship, or information connection to the event at Point 1; events inside the light cone are called *timelike* and do have a temporal relationship to Point 1.

A copy of state S_1 is instantiated at Point 2 (shown with its light cone) and allowed to physically evolve to a physical state S_2' that is sufficient to create conscious state C_2'. (For the moment we will ignore what may be happening at Point 3.) The physical evolution of state S_1 over time is subject to physical laws and depends on both random quantum events[6] and chaotic amplifications of initial conditions such that state S_1 could evolve over some time period to any one of a large set of mutually exclusive states, among which S_2, S_2', and S_2'' are possibilities. By SΛI, a conscious person exists at Point 2, and that person is Alice — let's call her Alice_2. Alice_1 and Alice_2 are the same person; Alice at Point 1 is Alice at Point 2, and she will subjectively experience the evolution from C_1 to C_2' because of a temporal continuity in her identity.

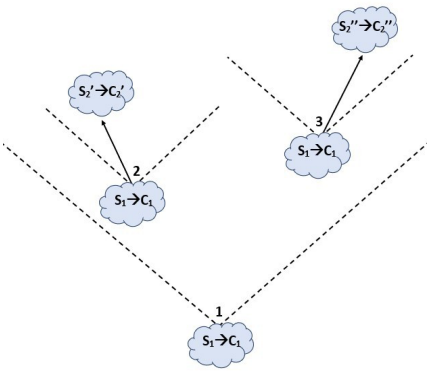

Figure 1: Spacelike separated instantiations of physical state S_1.

Figure 1 also shows a Point 3 (with its light cone), with Points 2 and 3 timelike to Point 1 but spacelike to each other. A copy of state S_1 is instantiated at Point 3 and allowed to physically evolve to physical state S_2'' that is sufficient to create conscious state C_2'', where C_2'' is consciously distinct from C_2' — that is, Alice would be able to tell them apart. (For the moment we will ignore what may be happening at Point 2.) By SΛI, Alice is created at Point 3 — let's call her Alice_3.

---

6 Whether or not quantum mechanics has anything to do with consciousness, it is clear that conscious states can correlate to (amplified) quantum events, such as the "click" a scientist consciously notices when using a Geiger counter to measure decays of radioisotopes.

Again, $Alice_3$ is the same person as $Alice_1$ and she will subjectively experience the evolution from C_1 to C_2'' because of a temporal continuity in her identity.

Now, imagine for the moment in Figure 1 that copies of state S_1 are instantiated at both of Points 2 and 3, but that $S_2' = S_2'' = S_2$ and $C_2' = C_2'' = C_2$ (i.e., they are not consciously distinct states). Note that Points 2 and 3 are timelike to Point 1, which is a necessary condition if the information necessary to produce the copies at Points 2 and 3 depends on information collected at Point 1. Further, Points 2 and 3 are spacelike to each other so that there is no fact about simultaneity or temporal ordering between them, nor can information pass from one to the other. While Point 2 appears closer in time to Point 1, and thus earlier than Point 3, in fact the interval in spacetime is the metric $s = \sqrt{c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2}$. Which of Points 2 and 3 occurs earlier in time is relative to an observer.

We already know that if state S_1 was not instantiated at Point 2, then Alice exists at Point 3 and she will subjectively experience the evolution from C_1 to C_2 because of a temporal continuity in her identity. However, even when S_1 is instantiated at Point 2, Alice exists at *both* Points 2 and 3, she has precisely the same subjective experience at both points, and she will subjectively experience the evolution from C_1 to C_2. This must be true because we already know that Alice exists in conscious state C_1 at Point 2. But the creation of physical state S_1 at Point 3 is spacelike to creation of physical state S_1 at Point 2, which means that from the perspective of Point 3, there is no fact about the creation of state S_1 at Point 2. The creation of S_1 at Point 2 can have no physical effect at Point 3. For the same reason that Alice exists and experiences an evolution from C_1 to C_2 due to the creation of state S_1 at Point 2, she exists and experiences an evolution from C_1 to C_2 due to the creation of state S_1 at Point 3. This may seem philosophically odd, but (so far) there is no physical contradiction. Tappenden (2011) explains: "How can two doppelgangers zillions of lightyears apart whose simultaneity we know, from Special Relativity, is entirely relative to an inertial frame, how can they share a single mind?" His answer: no need for causal connection. We should be careful, however, not to imagine different Alices who happen to have very similar experiences; $Alice_1$, $Alice_2$, and $Alice_3$ are all the same Alice. By S∧I, $Alice_1$ necessarily experiences the evolution of C_1 to C_2 due to physical evolutions of instantiations of physical state S_1 at both Points 2 and 3.

Finally, let's return to the original configuration of Fig. 1, in which $S_2' \neq S_2''$ and $C_2' \neq C_2''$ (i.e., they are consciously distinct states). It makes no difference which of the two systems diverges because the divergence of one is relative to the other. For instance, physical state S_1 might include Alice's body in which her eyes are closed, with state S_1 at Point 2 in intergalactic space oriented so that she faces one direction and state S_1 at Point 3 oriented elsewhere in intergalactic space and facing the opposite direction. Clearly, upon opening her eyes, her

view and resulting conscious experience will differ, corresponding to a difference in underlying physical states.

In that case, both $Alice_2$ and $Alice_3$ exist, but the underlying physical evolutions of the two physical states S_1 diverge. For the sake of clarity, assume also that the entirety of the evolutions of states S_1 (at Point 2) to S_2' and S_1 (at Point 3) to S_2'' are spacelike to each other. Given that Alice exists at both Points 2 and 3, what does Alice actually experience in Figure 1? At both Points 2 and 3 she experiences state C_1 arising from state S_1, but then does she subjectively experience an evolution to C_2' or C_2''? There are exactly three logical possibilities:[7] she experiences neither, both, or one or the other.

Is it possible that Alice experiences neither? Remember that Alice is created independently by instantiations at both Points 2 and 3; the instantiation evolving from Point 3 cannot be affected by the instantiation evolving from Point 2. If Alice is created by the instantiation at Point 3 and experiences an evolution from C_1 to C_2'', then nothing that happens in the evolution of state S_1 at Point 2 to S_2' can affect Alice's experience from C_1 to C_2''. Therefore, she cannot experience neither.

Perhaps Alice experiences both? What would it feel like for Alice to subjectively experience consciously distinct evolutions from C_1 to C_2' and C_1 to C_2''? Or, more colloquially, what would it feel like for her to experience diverging streams of consciousness? Before we indulge in speculations about higher planes of consciousness or fissioning identities, we should recognize the problem with locality. Note that the experience of $Alice_2$ in Fig. 1 depends only on the underlying local physical evolution of S_1 to S_2', because state S_2' is sufficient to create state C_2', while the experience of $Alice_3$ depends only on the underlying local physical evolution of S_1 to S_2'', because state S_2'' is sufficient to create state C_2''. $Alice_3$ cannot experience an evolution of conscious states that depends on a physical evolution of S_1 to S_2', because S_2' is spacelike to S_2'', just as $Alice_2$ cannot experience an evolution of conscious states that depends on a physical evolution of S_1 to S_2''. Therefore, Alice cannot experience an evolution of conscious state from C_1 to *both* C_2' and C_2''.

Finally, is it possible that Alice experiences one or the other? The problem here is that because the underlying physical systems are spacelike separated, there is no means or criterion by which nature can independently select one instantiation or the other for the continuity of Alice's identity. For instance, imagine that Mother Nature simply required that the "first" instantiation of state S_1 would assume Alice's identity; that won't work because Points 2 and 3 are

---

[7] These three options perfectly mirror the options suggested by Saunders (1998): "Nothing, both, or else just one of them?" He concludes that only the third alternative is possible but does not consider the effects of nonlocality.

spacelike to each other so there is no observer-independent fact about which happens first. Perhaps Mother Nature could choose the point having the shortest spacetime distance to Point 1. However, again, the spacelike separation of the points prevents the passage of any signal between Points 2 and 3 to "stop" the creation of Alice's conscious state C_1 at one of them.

Therefore, having ruled out the only three possibilities, the copiability of conscious states to spacelike locations in spacetime leads to a contradiction with SΛI. One might object that the above argument depends on the instantiation and evolution of spacelike-separated copies of state S_1; perhaps there is no problem when the copies are instantiated locally to each other. This objection fails. If there is something fundamentally impossible about the spacelike instantiation of copies, then it must be found in their instantiation, not in their relative locations or the extent to which they are separated. Further, the following section will show that timelike instantiations of copies of state S_1 are equally problematic.

Timelike Separated Copies

The spacelike instantiations discussed above are problematic because of a lack of causal connection. I will now show that a similar contradiction arises when we consider two timelike instantiations of state S_1 because they would require impermissible backward-in-time causation.

Figure 2 is comparable to Figure 1, the primary difference being that Point 3 is timelike to (and occurs after) Point 2, which is timelike to (and occurs after) Point 1. For the same reasons, if we ignore what may be happening at Point 3, then $Alice_2$, who is the same person as $Alice_1$ at Point 1, exists at Point 2 and will subjectively experience the evolution from C_1 to C_2'. And if we ignore what may be happening at Point 2, then $Alice_3$, who is the same person as $Alice_1$ at Point 1, exists at Point 3 and will subjectively experience the evolution from C_1 to C_2''.

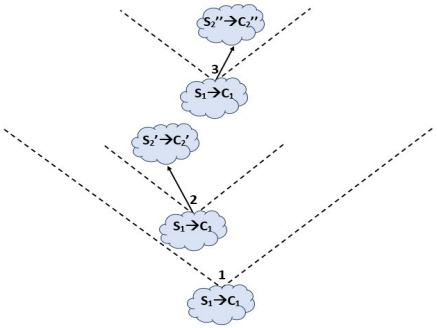

Figure 2: Timelike separated instantiations of physical state S_1

Now, imagine for the moment in Figure 2 that copies of physical state S_1 are instantiated at both of Points 2 and 3, but that $S_2' = S_2'' = S_2$ and $C_2' = C_2'' = C_2$. By SΛI, Alice exists and has precisely the same subjective experience at *both* Points 2 and 3 and will subjectively experience an evolution from C_1 to C_2. To the extent that any conscious state, when it is experienced, is subjectively experienced as occurring "now," states C_1 at Points 2 and 3 must be experienced identically and therefore must be subjectively experienced simultaneously. So, even though Point 3 objectively occurs after Point 2, Alice's conscious experience (which includes the subjective experience of "now") is the same at both points. After all, if they were subjectively experienced as different times, such different experiences would manifest themselves in different conscious states. While this may seem philosophically odd, again there is no contradiction. Zuboff (1990) explains: "This experience [across brains] of being you, here, now, would be numerically the same whenever, as well as wherever, it was realized."

Unlike in the previous subsection in which the copies of state S_1 are spacelike separated, $Alice_3$ is in fact created after $Alice_2$, in which case a causal connection is possible in principle. However, state S_1 is already assumed (by SΛI) to be sufficient to create Alice in conscious state C_1, so the instantiation of state S_1 at Point 3 must necessarily create Alice in state C_1, independently of her creation at Point 2 (or any other event prior to Point 3). In other words, the assumptions of Supervenience and Copiability prevent the kind of causal connection that would invalidate these arguments. For example, consider the case in which creation of $Alice_2$ causally prevents the creation of $Alice_3$; if that were true, then creation of conscious state C_1 at Point 3 would depend on facts that transcend physical state S_1, thus contradicting Supervenience. Therefore, if state S_1 is sufficient to create Alice in state C_1, then instantiation of S_1 at Point 3, and its corresponding creation of $Alice_3$, cannot be altered by the existence of $Alice_2$. For the same reason that Alice exists and experiences an evolution from C_1 to C_2 due to the creation of state S_1 at Point 2, she also exists and experiences an evolution from C_1 to C_2 due to the creation of state S_1 at Point 3. Again, we should be careful not to imagine different Alices who happen to have very similar experiences; it is the same Alice.

Finally, let's return to the original configuration of Figure 2, in which $S_2' \neq S_2''$ and $C_2' \neq C_2''$ (i.e., they are consciously distinct states). In that case, both $Alice_2$ and $Alice_3$ exist, but the underlying physical evolutions of the two states S_1 diverge. Given that Alice exists at both Points 2 and 3, what does Alice actually experience in Figure 2? At both Points 2 and 3 she experiences state C_1 arising from state S_1, but then does she subjectively experience an evolution to C_2' or C_2''? Like in the case of spacelike separated copies of state S_1, there are exactly three logical possibilities: she experiences neither, both, or one or the other.

Regarding neither, I have explained why instantiation of state S_1 at Point 3, and its evolution to state S_2'', will necessarily produce Alice subjectively experiencing state C_1 to C_2'' independently of instantiation of state S_1 at Point 2. Therefore, it cannot be neither.

Regarding both, the problem here is that Alice experiences the evolutions of C_1 to C_2' and C_1 to C_2'' as simultaneous (from her subjective perspective), because they both begin at the same conscious state C_1. Therefore, for Alice to experience both conscious evolutions, it must be the case that the conscious experience of $Alice_2$ depends not only on the underlying physical evolution of S_1 (at Point 2) to S_2', but also on the underlying physical evolution of S_1 (at Point 3) to S_2''. However, if Alice is to experience conscious states in real time with their underlying physical states, this scenario would require impermissible backward-in-time causation, because what she experiences prior to Point 3 would depend at least in part on events occurring *after* Point 3. Further, because of physical indeterminism (e.g., chaos and quantum mechanical randomness), it cannot be known at Point 2 whether, or when, additional copies of state S_1 will be instantiated. Therefore, it would be impossible for Alice to experience both an evolution of C_1 to C_2' and C_1 to C_2'' if she experiences them in real time with the underlying physical evolutions.

The only possible way for Alice to experience both evolutions would require two facts: first, that Alice's conscious states are experienced *after* (i.e., not in real time with) the physical states that produced them; and second, that Alice could not experience any conscious states until the instantiation of further copies of state S_1 is no longer possible. In other words, Alice would only be able to experience conscious state C_1 created by multiple copies of state S_1 after the final copy of state S_1 was instantiated. Thus, Alice can only subjectively experience consciousness after which time her conscious states can no longer be copied or repeated — in which case Copiability is false.

Finally, regarding whether Alice will experience one evolution or the other, I have explained why instantiation of state S_1 at Point 3, and its evolution to state S_2'', will produce $Alice_3$ subjectively experiencing state C_1 to C_2''. Therefore, if Alice is to experience conscious states in real time with their underlying physical states, whether or not $Alice_2$ will experience the evolution from C_1 to C_2' depends on the *future* fact of whether $Alice_3$ is created by instantiation of state S_1 at Point 3, another impermissible backward-in-time causation. Thus, for Alice to experience one evolution or the other, not only would Alice's conscious states necessarily be experienced long after the physical states that produced them have disappeared, but Alice could not experience any conscious states until the instantiation of further copies of state S_1 is no longer possible. Thus, Alice can only subjectively experience consciousness after which time her conscious states can no longer be copied or repeated — in which case Copiability is false.

Therefore, having ruled out the only three possibilities, the copiability of conscious states to timelike locations in spacetime leads to a contradiction with S∧I. In conjunction with the conclusion of the subsection on Spacelike Separated Copies, it follows that the conjunction of Supervenience, Identity, and Copiability is false.

Discussion

The conjunction of the assumptions of Supervenience, Identity, and Copiability leads to a contradiction, whether physical copies of a conscious state are created spacelike or timelike. Therefore, because both Identity and Copiability depend on Supervenience, then either Identity or Copiability (or both) is false. Logically, Identity → ¬Copiability and, equivalently, Copiability → ¬Identity.

That is, if a conscious state of a person having a transtemporal identity is created by a physical state, then that conscious state cannot be copied or repeated. (I neglect the possibility that something about the universe prevents those copies from evolving differently. However, if something about the universe prevents physical copies of conscious states from evolving to consciously distinct states, then the philosophical problems that motivated this paper are mooted anyway.) Equivalently and alternatively, if a conscious state can be physically copied or repeated, then it is not the conscious state of a person having a transtemporal identity. In this section, I will discuss a variety of implications of this result.

Direct Implications

The following implications depend only on the logical incompatibility between Identity and Copiability (and not on whether the assumption of Identity is true). Earlier, I posed the following problem: *A scientist claims that he is able to upload me to a digital computer. Do we have reason to doubt his claim?*

I think the objectively correct answer to this question, based on the previous analysis, is yes. The possibilities of mind uploading and computer consciousness depend on whether consciousness is fundamentally algorithmic — that is, whether consciousness can be reduced to a finite set of input-dependent instructions. Because an algorithm can be executed in the form of software on any general-purpose computer independently of its underlying physical substrate, any instantiation of an algorithm can be copied with no fundamental limitation on the number of copies that can be executed; it can also be repeated on the same computer by resetting that computer and then executing the software again. Therefore, if a computer can support a conscious state C_1, then the underlying sufficient physical state S_1 creating that conscious state is inherently copiable, such as by executing the

relevant algorithm on a general purpose computer located elsewhere in spacetime. Because Copiability → ¬Identity, that copiable conscious state C_1 cannot be associated with a transtemporal identity.

But now we have a problem. I certainly *believe* that I have a transtemporal identity, and on this basis alone I would doubt the scientist's claim even if I am factually wrong. However, he is claiming to upload me to the computer. If I do not actually have a transtemporal identity, then his claim to upload me is both nonsensical and false because if he actually succeeds in physically configuring a computer to create a conscious state, that state lacks an identity and thus cannot in any sense be regarded as *my* conscious state.

In other words, the scientist's claim is logically incompatible with Identity → ¬Copiability. The scientist's claim to upload *me* to a computer, so that I can expect to experience the conscious state C_1 produced by the computer, necessitates that I have transtemporal identity and that conscious state C_1 is one of mine; but if that is true, it could not have been created by a digital computer executing software because software is inherently copiable.

So if conscious states can be physically copied or repeated, then it is at the expense of transtemporal identity. I can then conclude that no scientist will be able to upload me to a digital computer, because either conscious states cannot be copied or repeated or because there is no transtemporal identity that equates the "pre-uploaded me" to the "post-uploaded me." For similar reasons, I can conclude that *my* conscious states cannot be copied or repeated, such as by duplicating my brain or body, either because there is something fundamentally physically unclonable about conscious states or because those conscious states lack transtemporal identity, in which case they cannot be mine and it is logically nonsensical to discuss what *I* might consciously experience due to those brain/body duplicates.

A similar analysis instantly renders moot the previously mentioned philosophical problems. Regarding the duplication problem, either it is not physically possible to create a "precise duplicate of the traveler," or else these copies do not create "his" awareness. The simulation problem depends on uploading one's conscious awareness onto a computer, which conflicts with Identity → ¬Copiability as discussed above. Regarding the self-location problem, just as a scientist cannot upload me to a digital computer, he cannot physically create multiple copies of me (i.e., physical copies of conscious states associated with my transtemporal identity), in which case I have no difficulty in self-locating. Finally, the Boltzmann brain problem also runs afoul of Identity → ¬Copiability: either these brains are not sufficient copies of my brain (or whatever underlying physical state produces my consciousness) or the conscious states they create lack transtemporal identity, in which case there is nothing that I would experience by their creation.

Indirect Implications

In deriving the logical incompatibility between Identity and Copiability, I have shown that not more than one can be true. Which is more likely to be true? In the section entitled "Questioning Identity," I argued that the Identity assumption is very strongly supported, in which case Copiability is false. Of course, if Copiability turns out to be true, then Identity is false and we would need a very good explanation for the universality of belief in transtemporal identity. Further, as scientists determining whether Copiability is true, we might consider these questions:

1. What empirical evidence do we currently have in its favor, such as evidence that an existing computer or artificial intelligence is conscious?
2. Because Copiability → ¬Identity, what empirical evidence would serve to validate Copiability given that *all* scientific inferences (and the foundations of science in general) depend on a scientist's transtemporal identity to make those inferences?

I think reasonable answers to the above questions will seriously undermine the possibility that Copiability is true instead of Identity. For that reason, I'll now discuss an implication in which Copiability is false. Because an algorithm that is executable on a general purpose computer can be readily copied and executed on another general purpose computer elsewhere in spacetime, if a particular conscious state cannot be copied or repeated, then execution of an algorithm cannot produce such a conscious state. This conclusion applies equally to classical and quantum computers. Of course, such an implication does not rule out a form of algorithmic consciousness that lacks identity, although it would imply a different understanding of the notion of computer consciousness. For instance, in speaking to a computer system producing a conscious state lacking identity, one could not meaningfully tell it, "I am going to make a copy of *you* on another computer." This leads inevitably to the conclusion that Strong Artificial Intelligence ("Strong AI") is either false (cf. (Searle, 1990)) or that the consciousness produced by artificial intelligence lacks transtemporal identity.

Empirical Approaches

While the logical argument presented in this paper may or may not directly generate empirical predictions, it does suggest directions for empirical research that might indirectly test its conclusions. If consciousness indeed depends on physical states that cannot be copied due to quantum no-cloning or other physical constraints, then we might expect to find evidence of quantum effects in biological systems associated with consciousness.

Landmark work by Hameroff and Penrose (2014) proposes not only that consciousness may depend on quantum interactions in the brain and

entanglements beyond the brain, but also suggests the specific means and method by which such quantum superpositions and interactions may be relevant.

An empirical approach to testing the non-computational nature of consciousness might also examine whether proposed measures of consciousness, such as integrated information (Tononi et al., 2016), correlate with measures of quantum coherence or other non-classical properties in neural systems. If consciousness depends on physical properties that cannot be copied, we might expect to find that neural correlates of consciousness exhibit features that cannot be fully captured in classical computational models.

Explanatory Hypotheses

Assuming, as in the subsection on Indirect Implications, that Copiability is false, why might it be fundamentally physically impossible to copy or repeat conscious states? For example, the statement, "Consciousness is created by the brain," might very well be true, but if Copiability is false, then there must be something physical about the brain (or, specifically, a physical state encapsulated by the brain that is sufficient to create consciousness) that prevents it from being copied. Why might this be the case?

For physical state S_1 that is sufficient to create conscious state C_1, it may be the case that there is no way to adequately measure physical state S_1, or if there is, that there is no way to recreate it elsewhere in spacetime. Some potential reasons may include: the information necessary to specify the physical state is contained in distant particles that cannot be reached; attempting to measure the information inherently destroys it, for example, because of reduction of a quantum wave function; the information cannot be measured with adequate precision due, for example, to quantum uncertainty; the state includes quantum entanglements with distant particles that cannot be measured or recreated; and/or the state is too large to measure or instantiate elsewhere.

The assumption of Supervenience does not depend on physical state S_1 being a quantum state, nor do the arguments in this paper depend on any relationship between consciousness and quantum mechanics. Having said that, many of the above explanatory hypotheses for the physical inability to duplicate conscious states invoke quantum mechanics. This should not be entirely surprising; the only known physical mechanism that prevents the instantiation of multiple copies of the same entity is quantum no-cloning, a no-go theorem that asserts the impossibility of creating an identical copy of an unknown quantum state (Wootters & Zurek, 1982). It may well be the case that whatever physical state S_1 is required to create conscious state C_1 depends on quantum information that prevents it from being copied. Aaronson (2013),

for example, points out that if a conscious chunk of matter is "unclonable for fundamental physical reasons," then that unclonability could be a consequence of quantum no-cloning if the granularity a brain would need to be simulated at in order to duplicate someone's subjective identity was down to the quantum level.

Further, the non-repeatability of a conscious state implies irreversibility, and the concept of irreversibility pervades physics: classically in the form of the Second Law of Thermodynamics and increasing entropy, and quantum mechanically in the form of decoherence (Haroche, 1998), which occurs "as soon as a single [correlated] quantum is lost to the environment." Since there will always be those who claim that any quantum mechanical process is reversible in principle by acting on the system with the reverse Hamiltonian, we could limit ourselves to "any fact of which the news is already propagating outward at the speed of light, so that the information can never, even in principle, be gathered together again in order to 'uncause' the fact" (Aaronson, 2013). If we indeed live in a universe with nonnegative curvature, as is currently believed, and given that almost the entire night sky is black, we can regard virtually every photon emitted into space as irretrievable. Thus if it turns out that consciousness can only result from irreversible processes, that irreversibility may itself be manifested in a physical state that cannot be read or copied because some of the information specifying it is embedded in photons streaming through space at the speed of light.

Conclusion

This paper has demonstrated that the conjunction of three seemingly reasonable assumptions — Supervenience, Identity, and Copiability — leads to a contradiction when examined within the constraints of special relativity. This finding has profound implications for our understanding of consciousness and its relationship to physical reality. The conclusion that conscious states with transtemporal identity cannot be copied challenges fundamental assumptions in computational theories of mind and artificial intelligence research. Consider the following common argument: *conscious states are produced by the brain; the brain is a just a chuck of matter; it is copiable in principle; therefore, copying a conscious state is merely a technological problem.*

The allure of this faulty logic is irresistible and has led physicists and philosophers into a minefield of intractable problems and seeming paradoxes. More careful scrutiny of the assumptions, particularly in light of physical limitations imposed by special relativity on copies of conscious states instantiated at different points in spacetime, yields a contradiction: there is something about consciousness that prevents it from being copied or repeated at will.

References

Aaronson, S. (2013). The ghost in the quantum Turing machine. *arXiv preprint arXiv:1306.0159*.

Bostrom, N. (2003). Are we living in a computer simulation? *The Philosophical Quarterly*, *53*(211), 243–255.

Carroll, S. M. (2017). Why Boltzmann brains are bad. In S. Dasgupta, R. Dotan, & B. Weslake (Eds.), *Current controversies in philosophy of science* (pp. 7–20). Routledge.

Elga, A. (2004). Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research*, *69*(2), 383–396.

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory. *Physics of Life Reviews*, *11*(1), 39–78.

Haroche, S. (1998). Entanglement, decoherence and the quantum/classical boundary. *Physics Today*, *51*(7), 36–42.

Parfit, D. (1971). Personal identity. *The Philosophical Review*, *80*(1), 3–27.

Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.

Saunders, S. (1998). Time, quantum mechanics, and probability. *Synthese*, *114*(3), 373–404.

Searle, J. R. (1990). Is the brain a digital computer? In *Proceedings and addresses of the American Philosophical Association* (Vol 64, No. 3, pp. 21–37). American Philosophical Association.

Stoica, O. C. (2020). Are mental states nonlocal? *Mind and Matter*, *18*(2), 175–218.

Tappenden, P. (2011). A metaphysics for semantic internalism. *Methaphysica*, *12*(2), 125–136.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461.

Wootters, W. K., & Zurek, W. H. (1982). A single quantum cannot be cloned. *Nature*, *299*(5886), 802–803.

Zuboff, A. (1990). One self: The logic of experience. *Inquiry*, *33*(1), 39–68.