Scientific Research Publishing

Open Journal of Philosophy, 2026, 16(3), 343-368
https://www.scirp.org/journal/ojpp
ISSN Online: 2163-9442
ISSN Print: 2163-9434

Embodying Consciousness in Brains and Machines I: Mental Monism

Peter B. Lloyd

Independent Researcher, Inverness, United Kingdom
Email: peter@peterblloyd.com

How to cite this paper: Lloyd, P. B. (2026). Embodying Consciousness in Brains and Machines I: Mental Monism. *Open Journal of Philosophy, 16,* 343-368.
https://doi.org/10.4236/ojpp.2026.163021

Received: April 24, 2026
Accepted: June 1, 2026
Published: June 4, 2026

Abstract

Mental monism is the philosophical theory that ultimate reality comprises only conscious minds. (It is usually identified with subjective idealism, but there are differing readings of that term, so "mental monism" is preferred.) This is a relatively neglected framework for examining the embodiment of consciousness in physical objects such as brains and computers. There is, however, growing interest in mental monism as other frameworks have faltered in seeking to solve the "Hard Problem" of consciousness. It is therefore timely to re-examine the case for mental monism, and to investigate the possible ways in which the embodiment of conscious minds, natural or artificial, could be modeled within this framework. This paper defends mental monism, and an accompanying paper examines the constraints on models of embodiment within that framework, proposes a biologically plausible model of embodiment in brains, and suggests an extension to embodiment in machines. The conscious mind is characterized as a nonphysical and nonlocal system comprised only of volitions and phenomenal experiences, subject to scientifically discoverable natural laws, and capable of processing information independently of the linked brain tissue.

Keywords

Hard Problem, Consciousness, Mental Monism, Idealism, Artificial Consciousness, Conscious AI

1. Introduction

The Hard Problem of consciousness (Chalmers, 1996), which asks how conscious minds relate to the brains in which they are embodied, has a long and contentious history. More recently, the AI (Artificial Intelligence) industry has addressed the parallel question of consciousness embodied in machines. The AI community has widely adopted computational functionalism, in which consciousness is identified

with, or supervenes on, information processing irrespective of the substrate that processes the information. As that position is widely seen as problematic in the philosophical community, it is timely to present a defense of an alternative view, comprising (a) theories that assign fundamental reality to consciousness and (b) theories that assign substrate specificity to the interface between the mind and its embodiment.

The term "conscious artificial intelligence" is often used interchangeably with "artificial consciousness" but this conflation rests upon a contentious assumption that having conscious experiences requires intelligence. Everyday experience points to the contrary. Having any conscious experience, such as seeing phenomenal red, or feeling pain, is a given, and not the outcome of intelligently solving a problem.

The "embodiment" of a conscious mind in a physical object is a condition where the conscious states affect, and are affected by, the physical states of the object. That definition excludes epiphenomenal consciousness. Human minds are obviously not epiphenomenal, otherwise we could never talk about conscious experiences. (Nor, *a fortiori*, could we meaningfully state that we possess epiphenomenal consciousness, as the term "consciousness" would be incapable of achieving a reference to consciousness.) Likewise, an artificial conscious mind that is of interest will affect, and be affected by, the states of the machine in which it is embodied. So, we should not say that an object "is conscious" or "has consciousness", as such expressions assume that being conscious is a property of a physical object, like its temperature. Rather, we should say only that it "embodies" consciousness, which leaves open the manner of that embodiment and whether the consciousness is literally "in" the object.

To situate the ontological position of this paper, I'll outline four broad categories of theories.

First, a theory may be "realist" or "non-realist" about consciousness. Non-realists assert that consciousness does not exist as a fundamental part of reality: it's an illusion, or a folk-psychological idea that does not correspond to any real thing; or a high-level emergent property of a complex physical system. The fact that we directly experience conscious sensations in everyday life negates all forms of non-realism about consciousness. Non-realism about consciousness persists only as an ideological position that flies in the face of empirical data.

Second, a theory may be "physicalist" or "nonphysicalist." Nonphysicalists maintain that the conscious mind, although undoubtedly part of reality, is not part of the physical universe. This minority position attracts a lot of naïve *ad hominem* criticism from physicalists who say that nonphysicalism entails a supernatural soul, or that it is a magical *deus ex machina*. Such comments stem from a misapprehension. Nonphysicalism simply asserts that the conscious mind is part of the natural world that does not have the usual characteristics of physical things, such as mass, energy, or location.

Third, a theory may be "local" or "nonlocal." A "local" theory of consciousness

asserts that a conscious mind is distributed over a volume of space (namely, the brain tissue, in the case of humans). Although it would be logically possible to propose a point-like location of the mind, there is little or no motivation for it, so I will not consider that possibility here. The motivation for assuming a local, distributed consciousness is, of course, that the activity of the brain is distributed over a volume of space, and the brain's activity interacts with the mind, so it is a short, albeit naïve, step to suppose that, say, conscious visual experiences are literally inside the visual cortex, auditory experiences are literally inside the parietal cortex, and so on. In contrast, a "nonlocal" theory asserts that spatial location is not a property that can be attributed to consciousness. Instead, the conscious mind resides in a domain of the natural world that is not included in the spatiotemporal manifold that physics defines, and yet still interacts with the physical brain. This is counter-intuitive but, as I shall argue below, the logic of consciousness forces us down this avenue.

Finally, a theoretical position on the embodiment of conscious minds requires an understanding of the nature of the physical world. Theories can be realist or non-realist about the physical world. Non-realism on this point is deeply contentious, but I shall argue that the logic of consciousness compels us to adopt it.

Although the aim of this and the accompanying paper (Lloyd, n.d.) is to address the embodiment of consciousness, we need some preparatory work on the nature of the relationship between the mental and physical worlds: that groundwork is covered in this paper. The argument for mental monism is semantic, while the argument for nonlocality stems from the logical structures of the mental and physical domains (the latter including Special Relativity). The arguments can be summarized as follows. (i) The discourse of the physical is topic neutral, as it is concerned only with the structure and dynamics of physical things, not with any putative intrinsic qualities, whereas the discourse of the mental is anchored in private ostensive definition. Therefore, the two discourses are disjoint, and no physical facts can entail any mental facts. Hence the conscious mind is nonphysical. (ii) The logical structure of consciousness is incompatible with the attribution of spatial location. Hence the conscious mind is nonlocal. (iii) Causal reference cannot be made to things that are not grounded in conscious observation. Therefore, reference cannot be established to a putative mind-independent physical world. Hence what we take to be the physical world in everyday life is a virtual construct grounded in consciousness.

These three strands lead to the conclusion of mental monism (that the conscious mind is real and nonphysical and the physical world a construct) plus the additional conclusion that the conscious mind is nonlocal. Those conclusions will form the premises for the "portal theory" in the accompanying paper.

2. Basic Theory

2.1. Consciousness Is Nonphysical

The notion that the conscious mind is nonphysical can be traced back to the sub-

stance dualism of Descartes (1641/1975), but it gained a wider visibility in the scientific community with the property dualism of Chalmers (1996). Dualism is the philosophical theory that consciousness forms an additional component of the natural world, besides the physical: substance dualism asserts that minds are nonphysical things, while property dualism says that consciousness is a nonphysical property that attaches in some way to physical things.

Arguments for nonphysical consciousness take the general form of saying that physics deals only with the structure and dynamics of the world, and cannot engage with the qualitative properties that are central to phenomenal experience. Chalmers' key argument invoked the "philosophical zombie", a replica of a human being that reproduced every detail of its structure and dynamics, but simply did not embody consciousness. As we can conceive of such a zombie, Chalmers argued, the phenomenal properties must be additional to all the physical properties. The main weakness of the argument is that it lacks a proof that we really can conceive of such a zombie, as opposed to merely imagining that we can conceive of it. A less obvious weakness is that it renders all discussion of consciousness, including the argument's own statement, empty. If zombie Chalmers can write a book on consciousness without either possessing consciousness or having an analytical definition of consciousness, then it is not using the word "consciousness" to refer to consciousness; and if real Chalmers' actual brain does what the zombie brain does, then Chalmers himself is also failing to refer to consciousness with the word "consciousness". The argument is self-defeating.

Another approach along the same lines is Jackson's (1982, 1986) thought-experiment of Black-and-White Mary. Here is a modified version that patches up some of the holes in the original argument. Mary is born with a brain defect that prevents her from seeing colors. Later in life, she studies neuroscience and ironically specializes in color vision, acquiring a profound knowledge of the visual cortex and access to all physical facts about the cortex. One day, a new surgical procedure is invented that can rewire her visual cortex and give her color vision. On the way to the operating theatre, she is stopped by a philosopher who asks, "Mary, you have a complete and correct theory of how the visual cortex works, and you have access to all physical facts about your own visual cortex. You are about to conduct an experiment on it, namely to get the surgical team to rewire it for color. What outcome will you observe upon completion of this experiment?" Of course she has no idea what she will observe. Yet, according to a Popperian notion of science (Popper, 1934/1959), if she has a correct theory, and complete data on the prior conditions and the experimental protocol, then she should be able to predict the outcome that will be observed. As she clearly cannot do so, either the physical theory or the set of physical facts is inadequate, so the phenomenal colors must be something additional to all the physical stuff. (Jackson's original hinges on whether Mary acquires new knowledge after she sees colors, which leaves the argument open to certain lines of attack. By focusing instead on the observation rather than knowledge, we close those particular vulnerabilities.) This argument serves well

as an "intuition pump" (Dennett, 2013) but fails as a proof, as it uses such ill-defined notions as "all physical facts", against which an emergentist might argue consciousness is a high-level emergent physical fact and therefore Mary does not possess all the physical facts.

Foster (1982, 1991, 2008) provides a more robust refutation of the physicality of consciousness. I will not summarize his intricate argument here, but the argument developed below captures the same intuition as Foster's argument, Chalmers' zombies, and Jackson's Mary: namely, that phenomenal consciousness eludes the concepts and language of physics.

2.1.1. Physics Is Topic Neutral

An account of something is "topic neutral" if it pertains only to extrinsic relations, leaving out the intrinsic qualities of the relata. The "extrinsic properties" of a physical thing are those that it possesses by virtue of its relations to other physical things. An obvious example is velocity: a body can travel at x ms^{-1} only in relation to some other body. Every other physical property—mass, energy, charge, field strength, and so on—is extrinsic in the same sense. A gravitational field is defined entirely by its effect on massive bodies: there is nothing else to the field over and above its extrinsic relation to other physical things. If you were to imagine a purely intrinsic physical property, having no extrinsic component, call it "hylasity", then it would *ipso facto* not interact with other physical things, since a disposition to act upon other physical things would, by definition, be an extrinsic relation. Pure hylasity would leave no physical trace and could never be measured. It would be indistinguishable from a fiction. That is why all the properties that physics actually handles are extrinsic relations, as it is only through extrinsic interactions that observation becomes a possibility, and physics constitutionally has nothing to say about things it cannot observe.

Physics is built from formulae stating the structure and dynamics of entities that are defined analytically through the structure and dynamics of smaller entities, and ultimately in undefined fundamentals (such as mass and charge). Physics formulae have no terms for the intrinsic quality, the "quiddity", of these entities, nor any theoretical apparatus for handling qualitative properties, nor any third-person experimental procedures for measuring them. Therefore, physics is wholly about extrinsics and is silent about any putative intrinsics, which is to say it is topic neutral. Consciousness, in contrast, is constituted by an assemblage of intrinsic qualities, and propositions about consciousness are manifestly not topic neutral. Several philosophers have made this point in consciousness studies, for example Rosenberg (2004).

Physicists mostly hold a pre-philosophical view of the "quiddity" of physical entities, and have a vague notion that they are studying a mind-independent reality. They insist that they are measuring "real" things in the laboratory, and that therefore physics cannot be just a closed, formal system. Take any physics textbook, though, and peruse its equations: you will find no term for quiddity, because physics has no need of it. Physics as it is practiced—as opposed to how it is fanta-

sized—is topic neutral.

Some philosophers have tried to insert wiggle room between the qualityless world of physics and the qualityful world of the mind. Thus, "panprotopsychism" says that physical things have "proto-phenomenal" properties that are not phenomenal but give rise to the phenomenal (Chalmers, 2013/2015). Similarly, "neutral monism" claims that basic things possess proto-physical and proto-phenomenal properties that yield physical and phenomenal properties, respectively. Any such proto-phenomenal properties are irrelevant to the present philosophical discussion. Either we have direct acquaintance with them (that is to say, they are included in the phenomenal content of our conscious minds), or not. In the latter case, all terms are analytical and ultimately resolve into undefined fundamental properties such as mass, charge, and whatever novel proto-phenomenal properties are conjured up. This paper's reasoning treats any topic-neutral thing in the same way, and would therefore also apply to proto-phenomenal things. For clarity, I shall consider only physical things below under the rubric of the topic neutral. In the unlikely event of anyone actually discovering proto-phenomenal things in the future, the same arguments would apply.

2.1.2. Consciousness Is Not Grounded in Physics

Several physicalist reference theories have been put forward over the 20th Century. First, there is eliminativism, which denies that "consciousness" refers to anything at all, as what we think is consciousness does not actually exist. There is nothing to be said about this: if a person claims to be permanently lacking in conscious awareness then no argument can disprove it, but the rest of us know we are conscious.

Second, there are theories that seek to ground consciousness in physics. For example: identity theories (identifying mental states with physical states), functionalist theories (identifying conscious states with the performance of certain functions), and supervenience theories (claiming mental entities exist in virtue of certain physical states). A large literature has grown up around such theories, but I will not survey them here. Such theories have in common the claim that all mental facts are entailed in some way by physical facts. My argument below addresses that common claim, rather the specific problems of individual theories.

The core idea of this argument can be traced to Berkeley's semantic argument (Berkeley, 1713, Section III: p. 222). Articulated in a modern form (Lloyd, 2006), the argument presented here follows Berkeley's central thinking on the matter, which is different from his so-called "Master Argument" (a phrase invented by Gallois (1974)). What is sometimes referred to as his "semantic argument" is the only argument in his books that works, and he attaches a central importance to it. I shall outline my Berkeleyan semantic argument here, and expand it below. As Pearce (2014) observes, this linguistic interpretation of Berkeley differs from the standard scholastic readings—the subjunctive and ideational interpretations—both of which are easily defeated. Pearce shows that the linguistic interpretation is both a better fit to Berkeley's writings, and philosophically more robust.

There is no canonical definition of what it is to be physical. A perusal of standard physics, current and past, reveals certain universal features of what is taken to be physical. All physical terms are defined analytically in terms of fundamentals, which are undefined except for their extrinsic relations to other physical things—whereas all mental terms are anchored by private ostensive definition. Therefore, physical terms and mental terms form disjoint sets, and so physical propositions and mental propositions are disjoint, hence no mental fact can be grounded in physical facts. I will unpack that terse summary below.

As we noted earlier, the terms of physical discourse are defined analytically in terms of fundamental unknowns. For example, an electron is an entity with a designated mass, electric charge, spin, and so on. But the question of what mass "really" is lies outside the discourse of physics. The equations of physics include mass as a term but the actual nature of it—what mass really is—cannot be expressed in the language of physics.

Consequently, all propositions in the discourse of physics are expressed wholly in these physical terms. From any set of propositions in that discourse, further propositions can be inferred by propositional logic and predicate logic, but all such derived propositions will also be expressed in terms from the lexicon of physics. That is, the discourse that expresses itself in physical terms is closed under logical inference.

To be sure, new terms are continually being devised within physics, but these are either defined analytically in existing terms, or denote new fundamental quantities, which are undefined except for their possession by fundamental entities and their extrinsic relations to the rest of physics For example, the quark flavors of upness, downness, and strangeness (discovered in 1964), charm (1970), and topness and bottomness (1973)—these were novel physical properties that were defined in terms of behavior observed in other, already defined properties. No physicist is ever going to introduce a property such as hylasity that has no observable effect on other physical things.

In contrast, the discourse of consciousness gives meaning to its basic terms by private ostensive definition; and it builds up further terms analytically from them. That is to say, one mentally attends to some element of experience, and assigns a designation to it. For example, "phenomenal redness" (or, simply, "redness") can be defined only by having an experience of that phenomenal content, and accepting that label for it. An individual with congenital color blindness who cannot see red can never know the meaning of the term "red." This is so, even if that individual knows about the physics of red light, and the physiology and psychology of its perception.

The discourse of the mental expresses all of its propositions wholly in these terms. From any set of propositions in that discourse, we can infer further propositions by logic, but all such derived propositions will also be articulated in expressions from the repertoire of mental terms.

Could we add consciousness to physics as a new fundamental property, like

mass and charge? Future generations might choose to expand the concept of "physics" in this way, but the outcome would no longer be "physics" as the term is used today. Terms that denote phenomenal experiences are not anchored by extrinsic relations to other physical things but by private ostensive definition. Therefore, they do not comply with the normal usage of "fundamental physical property".

Thus, we find ourselves possessed of two corpora of propositions and hence of facts: the disjoint sets of physical propositions and mental propositions, and hence the disjoint sets of physical facts and mental facts. Since the corpus of physical propositions is closed under rules of inference, no mental proposition can be entailed by any assemblage of physical propositions. *Ergo* mental facts are additional to the set of all physical facts. To put it bluntly, consciousness is nonphysical.

Certainly, one could posit a psychophysical bridging proposition. For example, that a person experiences pain if and only if a certain C-fiber in her brain fires. But any such psychophysical proposition lies outside the corpus of physical facts, and therefore cannot follow from physics. Hence, it cannot defeat the conclusion that consciousness is non-physical.

Strictly speaking, the reasoning is symmetrical: no facts about a mind-independent physical world can be entailed by any mental facts. Nonetheless, a "physical construct", which is essentially a convenient fiction, can be grounded in the mental. The key difference is that the physical *construct* is not topic neutral, unlike the putative physical *world* itself which *is* topic neutral. This is because the construct is phenomenally grounded: if, in everyday speech, I say I am sitting on a chair (i.e. in the construct), that fact is grounded in my visual and tactile experiences; whereas if a philosopher asserts that there is a mind-independent physical chair, then it is not so grounded, and consequently is topic neutral. A lot of confusion has arisen from mixing up those two.

2.2. Consciousness Is Nonlocal

Where is your mind? The obvious naïve answer is that your mind is in your head, along with all your thoughts and experiences. But if you stub your toe, surely the pain is in your toe? If you taste horseradish, surely the taste is on your tongue? The aroma of whisky is in your nose. The music from stereophonic headphones is in the middle of your head, but the sound of a telephone ringing seems to be in the telephone. If you have damage to your sciatic nerve, the pain does not sit in the lesion but runs down the length of your leg. An amputee can feel pain in a phantom limb that does not physically exist. So, your intuition for the location of your mind and its contents seems a bit iffy. What does neuroscience say? That the mind, if it is acknowledged as really existing at all, is supposed to be in the brain tissue, where sensory nerve fibers terminate and motor nerve fibers begin. At first glance this seems a plausible hypothesis, but closer examination will show it is untenable.

The question of locality must be addressed by property dualists, substance du-

alists, and mental monists. Property dualists, such as panpsychists, are committed to local consciousness: conscious experiences are literally inside the brain tissue or computer circuit that embodies consciousness. Substance dualists have to ask whether mental events have a one-to-one mapping with brain events, even if the mind is a different thing from the brain. And mental monists have to examine whether the physical construct is so formed that mental events have a one-to-one mapping with virtual physical brain events.

2.2.1. What Does It Mean to Be Somewhere?

Consider a doughnut. What precisely do you mean when you say it is in a particular location? For example, on a plate in front of you. Well, for starters, you can look at the plate and see it there. You can reach out and touch its sticky surface texture, and feel resistance to the pressure of your fingers. You can lick it and smell it, get some gustatory and olfactory sensations. You can put it on a weighing scale and determine its weight. If you're in a lab, you can stick electrodes in it and measure its electrical conductivity. In other words, there is a range of observations that you can make repeatedly on that doughnut-shaped volume of space, and you will get persistent observational data from it. These are what we call third-person observations. They are public, in the sense that anyone else can come and see the doughnut or weigh it, and so on. As a general principle, we say that a thing is in a given place if there is a full range of third-person observations that reveal the presence of the thing in that place.

Some things might lack some of these observations. A cloud of methane gas is not visible but you can smell it and stick chemical detectors in it. The principle is the same, a thing's being in a certain place means that third-person observations of all its properties in that place will reveal it.

Sometimes you have to make a distinction between a thing in one place, and a correlate of it somewhere else. For example, you can see a news broadcast on your television screen, but it does not mean that that the newscaster is inside your television set. A small flat, weightless, two-dimensional image of the newscaster does not tell you that she is in your television. You have to able to carry out the full range of third-person observations to make a valid claim of presence - mass, volume, breath, warmth, etc.

Now consider someone's conscious mind. You cannot see it, smell it, taste it, weigh it, measure its electrical conductivity or charge. In fact, there are no third-person observations at all that you can make on someone's conscious mind. So, to put it bluntly, a conscious mind is not "third-person discoverable." Therefore, it cannot mean anything to assert that a conscious mind is in a certain volume of space. In order to be able to say meaningfully that a conscious mind is in some place, we would have to be able to observe it there. And we can't. So it isn't.

Hold on! Surely neuroscientists can observe the workings of a conscious mind by sticking electrodes in the brain tissue and measuring the electrical pulses on neural axons. Or put the person in a PET (Positron Emission Tomography) scanner and conjure up a colorful picture of the neural activity. Doesn't that count as

observing the mind? I understand where this question is coming from, but it's wrong.

By analogy, consider a program running on a computer. Software and hardware engineers will agree that the program is "in there." An engineer can attach electrodes to the device and see the flow of current on an oscilloscope, can see the binary digits at specific points. Why is this different from a neuroscientist examining a brain with electrodes? Quite simply: the software is nothing over and above the electronic signaling. It is true that the software also has a symbolic representation—the code that the programmer writes—but the program as compiled into binary code is present precisely as binary states in the hardware. The engineer can make a complete observation of the program in its fullness because the program is nothing but its binary embodiment, in physical voltages.

The situation with the neuroscientist is crucially different. The neural activity that the neuroscientist can observed is not the same thing as the phenomenal experience. For example, suppose the subject is exploring the aforementioned doughnut. In the actual sense-data there is a colorful visual image, a taste, a delicate aroma, a tactile stickiness: none of these essential qualities are observed by the neuroscientist. Whatever parameters the neuroscientist observes, they will definitely not include a doughnut flavor or any other of those phenomenal qualities. What the neuroscientist observes is a correlate of the conscious experience.

All we know is that some conscious experiences are correlated with brain events. (It would be an unwarranted assumption to assert that all mental activities are so correlated.) Nevertheless, we cannot legitimately say that any mental experiences are literally located in the space of the cranium. This leaves wide open the depth and scale of that correlation. If the whole of the mind turns out to be correlated with the brain's activity, then that agrees with neuroscientists' intuition that the conscious mind is tightly coupled to brain activity, even if it is not literally in the brain. As we will eventually see, however, there are reasons to believe that much of the mind's doings are not correlated with neurons' firing.

2.2.2. Argument from Relativity I: Consciousness Is Not Spatiotemporal

Lockwood (1989) used Einstein's (1905) special theory of relativity to argue for localization of consciousness. Here I reject that claim, and in the next subsection make a counter-claim, that relativity makes it impossible for consciousness to be separable in space.

Russell (1927: p. 384), Weingard (1977), and Lockwood (1984a, 1984b, 1985, 1989: pp. 71-78) used relativity to argue that, as conscious events are in time (proposition R_T), so they must be in space (R_S) since relativity melds time and space inextricably in spacetime. The reasoning from R_T to R_S is sound, but premise R_T is not. Lockwood's defense of R_T is as follows. We continually witness that the mind perceives and causes physical events, but physical events are in physical time, therefore (Lockwood says) mental events must be in physical time, at least to the granularity of the physical stimulus and response. For example, if I hear a buzzer at T_0 and press a button at T_1, then my perception and volition must have

occurred between T_0 and T_1. Now, for any brain event B that correlates with a mental event, Lockwood (1989: p. 75) wrote: "If it were now possible to find pairs of physical events standing to B as cause and effect that were separated by ever smaller temporal intervals then it would be possible to define the spatial location of B with any required degree of precision."

The neural correlates of the acts of perception and volition must be localized in spacetime. But Lockwood's program of narrowing down arbitrarily the volume of spacetime within which the correlate of any mental event occurs is unfeasible because the speed of light is so high that to localize B even within the 140 mm width of a human brain, the interval between T_0 and T_1 would have to be less than 10^{-10} s but no discernible mental events occur at that scale (even if we allow Pelczar's (2017) claim that mental events can be as brief as 10^{-4} s). Hence Lockwood's program cannot localize the correlates better than the crude observation that they are somewhere in the head.

Nevertheless, we must address Lockwood's core assertion that since mental events can be circumscribed in an interval of time (even if no finer than half a second), they must be circumscribed in a volume of space (albeit a large one). If Lockwood can prove that my mind is at least somewhere on the Earth, then he will have proved his point that the conscious mind is localized. Lockwood's argument rests on the premise that a physical event can cause a mental one. Any intelligible notion of causation, however, entails temporality, hence assuming that a physical event causes a mental one already presumes that consciousness sits in physical time, which is what the argument is supposed to prove. Hence it is circular. That was a bit quick, so let us expand the argument.

Physical events cannot cause mental ones under mental monism (see Corollary 3 below). So, what is happening in the simple experiment where I press a button whenever I hear a buzzer? According to mental monism there is a mental event M_{Buzz} that is rendered in the physical world as ren (M_{Buzz}), the physical sound waves. But it is M_{Buzz} that acts on my mind, not ren (M_{Buzz}). M_{Buzz} yields the mental sensation of the buzzing noise, but its doing so is a cause-and-effect in the mental domain and it has no coordinate in physical time. My volition to press the button is again a mental event M_{Press}, which is rendered in the physical world as ren (M_{Press}). Thus, in mental monism, there are two decoupled temporal sequences: on the one hand, the mental event M_{Buzz} causes me to hear the sound, and my volition then performs M_{Press}; on the other hand, within the virtual physical world, ren (M_{Buzz}) occurs, and then ren (M_{Press}), which occur at specific physical times. Relativity theory pertains to the rendered events ren (M_{Buzz}) and ren (M_{Press}), which must indeed occur in spacetime; but the mental process from M_{Buzz} to M_{Press} does not supervene on that physical world, but rather subvenes on it. As ren (M_{Buzz}) cannot cause M_{Buzz}, the causal link that Lockwood wants to use to tie consciousness into spacetime does not exist, and his argument fails.

The theory of relativity explicitly concerns the relationship of physical observations in spacetime. If mental experiences were observable events in spacetime then

Lockwood would be right, but they are not third-person observable, so there can be no operational meaning in ascribing physical time or space coordinates. Relativity says nothing about physically unobservable things, and certainly does not require them to be embedded in spacetime. Gibbins (1985) tried to soften the blow by writing, "The temporal results of relativity will apply to mental events only indirectly via the space-time events with which they are associated," but being "associated" with an event in spacetime does not constitute actually being in spacetime. In fact, mental events are not in physical time any more than they are in physical space.

2.2.3. Argument from Relativity II: Physical versus Mental Simultaneity and Sequence

I argued above that mental events cannot be in spacetime under the theory of mental monism. In this subsection I present an argument for this conclusion that does not rely on mental monism. In fact, from relativity we can derive two anti-localization arguments, as follows.

(a) First argument: A mind can have distinct, absolutely simultaneous experiences but the brain cannot, because simultaneity is relative to the observer's frame of reference, hence mind cannot be embedded in brain. Russell (Russell, 1926: p. 130) briefly mentions this as a puzzle but concludes nothing from it; de Silva (1995, 1996) likewise; Lee (2007) concludes that mental simultaneity is an illusion; Pelczar (2017) concludes that consciousness cannot supervene on events in spacetime. These authors disregard the limited temporal acuity of a human mind, which cannot tell whether two mental events are precisely simultaneous. Hence the argument, although well-intentioned, does not work in practice.

(b) Second argument: A mind has experiences in an absolute sequence, but the brain cannot, because the sequence of spacelike-separated physical events is relative to the observer's inertial frame of reference, hence mind cannot be embedded in brain. Lee (2007) also mentions this. This argument (b) is stronger than (a) as the sequence of sufficiently separated mental events is discernible and unambiguous.

This argument as it stands, however, is still flawed because events in a brain of normal size are not spacelike separated. For example, the separation of two brain events at $(X_0, 0, 0, T_0)$ and $(X_1, 0, 0, T_1)$ is spacelike if $(X_1 - X_0)^2 > c^2 (T_1 - T_0)^2$ where c is the speed of light, about 3×108 ms^{-1}. For a brain diameter of X = 0.14 m, T cannot exceed about 10^{-10} s, but this is way below the ability of any human mind to discern. Lee (2007) did not recognize this as a problem, and Phillips (2009: p. 222) rightly rejected Lee's relativistic argument because of it; but Phillips wrongly supposed that there could be no such argument from the relativity of sequence. In fact, we can consider a thought-experiment in which the brain is expanded to make its parts spacelike separated.

Property dualist theories, such as panpsychism, assert that conscious experiences inhere in physical substance. These theories do not specify any limits on the distances between the bits of physical stuff. Therefore, a mind that inheres in a

particular brain must still continue to inhere therein if (as a thought experiment) we were to expand the brain tissue over a much larger volume. In the following, I argue that a sufficiently enlarged brain would (if property dualism were true) imply a temporal structure that contradicts what we know from everyday experience. It is inferred that property dualism cannot be true, and *a fortiori* cannot be true in brains of normal size.

As property dualism does not prescribe any maximal volume of space that must contain the physical components whose elemental minds combine to form a personal mind, let us consider a human brain expanded to the size of the Earth's orbit around the Sun, with the same architecture as a regular brain but the axons stretched over millions of miles. A neural pulse travelling at a hundred meters a second will take ninety-five years to traverse the brain. Local transmissions, between, say the visual and auditory cortex would take a few decades. Obviously, this is an outlandish thought-experiment but the logic of panpsychist or supervenience theories and relativity must still apply.

In such a big brain, neural events are spacelike separated: it would take light sixteen minutes to cross the diameter of the brain.

Considering the mind that is associated with this brain in accordance with, say, panpsychism, suppose that it experiences a bright flash and, several years later, a loud bang. We cannot doubt the mental sequence. In relativistic physics, however, there is no such thing as absolute sequence of spacelike-separated events.

Given two spacelike-separated brain events B_1 and B_2, in some frames of reference, B_1 will precede B_2; in some B_1 and B_2 will be simultaneous; and in some B_2 will precede B_1, all depending on the velocity of the observer. Yet, if the phenomenal contents M_1 and M_2 that are associated with events B_1 and B_2 are combined into a unified mind, then M_1 and M_2 will have a determinate temporal sequence. The fact that the observers would have to be travelling at very high velocities to change the brain events' relative sequence is not to the point. What matters is the brute physical fact that sequence is not absolute for spacelike-separated events, and therefore the combination of spatially distributed elemental minds into a unified mind is untenable.

2.2.4. The Hypothesis of Universal Mental Time

Relativity prohibits faster-than-light communication (Tolman, 1917), but the decoupling of mental time and physical time seems to open up the possibility of instant communication. We need to check this potential conflict.

(a) Physically instant communication is impossible. Relativity states that if two observers, A and B, travel in uniform motion relative to each other, then in relation to A's frame of reference, time will slow down in B, and *vice versa*. This is well established experimentally. Consequently, superluminal communication would lead to contradiction. Suppose at T_3, A sends an instantaneous signal to B. Because of time dilation, it arrives at B at $T_2 < T_3$. Upon receiving this signal, B sends an instant message to A, saying "Don't send the signal," which arrives at T_1. Because of time dilation, $T_1 < T_2 < T_3$. On receiving this message, A decides not to send the

first signal at T_3, and we have a contradiction.

(b) Mentally instant communication is permitted by mental monism. Since minds are not in space, communication between minds is not required to travel through any intervening gap. Therefore, mental monism, in principle, permits communication between two minds in an instant of mental time. Furthermore, mental monism, in principle, also permits disembodied minds. Therefore, even if we try to block instant communication between two embodied minds by saying it cannot "travel" faster than light between the respective brains, this could be circumvented by using a disembodied mind as an intermediary.

(c) Therefore, if mentally instant communication entails physically instant communication, then mental monism cannot be true.

What exactly do we mean by "instant communication" from A to B? In a physical system, it means that P_A originates a signal at time T and it arrives at P_B at the time T' that, by the Lorentz time dilation equation, exists in P_B when P_A is at T. In a mental system, it is not so well defined, as we have no units or instruments for measuring mental time. The most plausible model is one that has been in currency from James (1890) to Hameroff and Penrose (2014), namely that consciousness comprises a series of discrete moments of awareness. Whether we accept this discrete model, or one in which consciousness flows continuously in time, the "clock" that drives consciousness is the succession of experiences. For an embodied mind that is interacting with its environment, that "clock" will in turn be correlated with the succession of neural impulses entering the sensory cortex, or arising from internal brain activity. Therefore, the physical correlate of elapsed mental time will be the elapsed physical time within the local frame of reference of the avatar (be it a brain or an artificial device), which is referred to as "proper time." If, at time T, mind M_A sends a mentally instantaneous signal to M_B, then it will arrive at its destination at the same local time T (not the Lorentz dilated time $T' < T$). Whereas physically instantaneous communication goes backward in time, mentally instantaneous communication does not. Hence, mentally instantaneous communication, as permitted by mental monism, does not entail a relativistic paradox.

To illustrate this, we can consider Einstein's thought-experiment of two lamps at opposite ends of a high-velocity train. One observer, Passenger, sits in the middle of the train; and another, Stationmaster, in a station. As the train passes the station, each lamp makes one flash. If Passenger receives both light signals at the same time, she concludes that—in her reference frame—the lamps flashed simultaneously. Meanwhile, Stationmaster sees the rear lamp flash before the front lamp, but both signals reach Passenger together. (Photons from the rear lamp must travel farther. In the time it takes the photons to travel from the rear lamp to Passenger, she will have moved forward slightly with the train.) If a third observer, Aeronaut, is travelling faster in the same direction (in an airplane, say) alongside the train, then she will observe the opposite sequence: in her reference frame, the front lamp signals first.

So far, this is standard relativity theory in the physical domain. Now, let us con-

sider the mental domain, where mental events are not in physical spacetime, and therefore not embedded in frames of reference. We will suppose, for the sake of illustration, a Berkeleyan model in which each physical object has a mental archetype, which acts like a kind of miniature mind. According to this model, a lamp's flashing is a physical rendering of a state change in the corresponding archetype. Given that the sequence of physical flashes depends on each observer's frame of reference, we naturally want to ask: In what sequence do the archetypes actuate these flashes? As the archetype's state change is not in physical time, there is no fact of the matter as to which mental state changed physically first; but we can still ask which occurred first in mental time. Imagine (for the sake of argument) that observers can "telepathically" communicate with archetypes. Now the observers can interrogate the archetypes of the lamps and discover which archetype triggered a signal first. Suppose, for example, that the rear lamp's archetype triggered its signal first. Now, the Aeronaut will be surprised to discover that, although front lamp flashes first in physical time (in her reference frame), the rear lamp's archetype-mind triggered a signal first. Thus, she has foreknowledge of a future physical event, which is spooky but not paradoxical. Next, imagine (again, for the sake of argument) that an observer can "telekinetically" control the archetype of a lamp. Let us vary the experimental protocol: as soon as Stationmaster discovers (telepathically) that the rear lamp's archetype has actuated its flash, she decides whether or not to veto (telekinetically) the front lamp's flash. This potentially creates a problem for the other two observers. For, Passenger should see both flashes at the same time, and Aeronaut should already have seen the front flash first, which Stationmaster has just cancelled telekinetically.

This paradox arises only if telepathy is physically instantaneous; but as we have seen, it is instantaneous only mentally, not physically. To be sure, in Aeronaut's reference frame, the time at which the front lamp is scheduled to flash can precede the time when Stationmaster decides whether to cancel the front flash. Aeronaut can telepathically ask Stationmaster if he has done so, and even ask the front lamp if it has flashed, and therefore can know whether the front lamp will flash, and do so at a time physically before the lamp would have flashed, but the light signal from the front lamp (travelling, obviously, at the speed of light) will not reach Aeronaut until after that time. So if the front lamp is vetoed, Aeronaut will not see the light, and there is no paradox, and no violation of causality. Admittedly, in Aeronaut's frame of reference, it seems that Stationmaster's mind retrocausally went back in time to cancel the front lamp's flash, but in Stationmaster's own frame of reference, there is no retrocausation.

We see that the mere conceivability of instant communication ("telepathy") between remote minds is enough to impose an absolute ordering on all mental events in the universe. This has the weird (but not paradoxical) result that events in two remote physical avatars could occur in any order (depending on the inertial frame of reference of the observer) but mental events in their corresponding minds must occur in a single, definite order. We are, however, left with the surprising conclusion that all conscious minds in the universe are running on the

same mental "clock time."

2.2.5. Local Consciousness Is Untenable

I have argued that assigning spatial location to phenomenal experiences leads to a contradiction with relativity in the (admittedly impractical) thought-experiment of an expanded brain. To avoid this problem with panpsychism, some writers have advocated "cosmopsychism", associating a universal mind with the state of the whole physical universe. For example, Mathews (2011), Jaskolla and Buck (2012), Shani (2015), Nagasawa and Wager (2016). At first, this seems like an extrapolation of panpsychism, but jettisoning any structural mirroring between the physical construct and the mental world loses the core intuition of panpsychism. Cosmopsychism is a variant of mental monism in all but name. It is not, however, equivalent to the most general form of mental monism as the latter allows for disembodied mental entities.

Berkeleyan mental monism seems more plausible, and offers more explanatory power, than cosmopsychism, but it remains to be seen which theory is right.

2.3. The Physical World Is a Construct

The argument for nonphysical consciousness (§2.1) delivered the first limb of the argument for mental monism. The second limb stems from the fact that we have direct experience of mental content but not of mind-independent physical systems. This enables us to establish reference to mental entities but prevents us from making reference to physical things. Any assertion that a mind-independent world exists is therefore an incoherent attempt to refer to what is unreferenceable, or what Berkeley termed an "unknown somewhat" (Berkeley, 1710, Section §80, 1713: p. 482). Therefore, the whole of pure physical discourse has the character of a formal text that refers to nothing outside itself, that is to say, it is a fiction, albeit a convenient one. I will expand this argument below, but first I want to motivate the ideas of "shadow causality" rendered by "the metamind".

2.3.1. The "Metamind" and Shadow Causality'

As soon as we say that consciousness is nonphysical, we are confronted by the question of how a conscious mind could have causal interactions with a physical embodiment (brain or machine). For a dualist, the options are: to deny causal interaction, which is the epiphenomenalism that we rejected earlier; or to say it is an inexplicable brute fact. Neither is appealing, but fortunately mental monism has another option.

For a mental monist, causality between the mental and physical domains is not a fundamental problem, because the physical world is a construct wholly grounded in consciousness. So, the underlying causal nexus lies within the conscious realm; and what we naively thought was a cause-and-effect in the physical domain is just a "shadow causality."

The situation is closely analogous to that of a virtual reality game: within the virtual gameworld, you might draw a bow and shoot an arrow; but this is initiated

in the normal world by pressing a button on a controller, which sends an electronic signal to the computer, which renders the moving image of the arrow on the screen. The actual causal chain is shadowed by a virtual causal chain on the screen. Likewise, if I push my mug across my desk, then: (a) the real causal chain comprises a thought from my conscious mind, which is conveyed to the mental infrastructure of the universe, which then renders in my visual consciousness the image of the moving mug; while (b) a shadow causal chain comprises a nerve signal from my brain to my fingers, producing a mechanical force causing the mug to move.

Mental monism must invoke a mental infrastructure to explain the regularities of the observed world. Shankara called it "Brahman", Berkeley called it "God", and Hoffman calls it a "Conscious Agent." I will use the neutral term "metamind" for the mechanisms within the domain of consciousness that carry actual causality, and subtend the shadow causality of the physical world.

The metamind is an explanatory posit. It is concrete entity of the same kind as our personal minds, but operating on a vastly greater scale. It controls all natural phenomena that are not produced by the volition of personal minds. Without the metamind as an organizing mechanism, mental monism reduces to phenomenalism. The actual existence of the metamind, and its internal architecture, are empirical hypotheses that could, in principle, be tested if we could one day devise a means of direct inter-mind communication.

2.3.2. Sensory Constructs

The second limb of the argument also requires the idea of a "sensory construct." To motivate this, however, I shall first touch briefly on the social construct as an illustration, as it is more familiar (e.g. Searle, 1995). Consider, for example, chess. The coordinates of the board map onto the squares of, say, a flat wooden block; the pieces—the King, Queen, Knights, and so on—map onto moveable pieces of wood; the "move" functions map onto particular movements of the pieces across the board, and predicates such as "is a valid move" and "is in checkmate" map onto particular movements and configurations.

Any usable construct needs both a formal system of entities and relations between them (for example, the board and pieces of chess), and a binding (for example, an agreement that this piece of wood "is" a White Knight, and that one a Black Queen, and so on). To be sure, we may study a formal system without a binding. We might want to study chess strategies *in abstracto*. Likewise, in pure mathematics, we study axiomatic systems without semantically binding them to a model. To use a formal system as a construct, however, we have to bind its terms.

As we lack commonly agreed terminology for some of these issues, I will propose some notation. I shall define a construct as a system of predicate logic C that includes a set of constant terms T_C, functions F_C, and predicates P_C. A semantic binding is a mapping of constants T_C into a set of substrate elements P, functions F_C into relations between elements of P, and predicates P_C to facts about the substrate.

It is obvious that the entities of a construct *per se* have no independent exist-

ence: they have a purely notional existence that is exhausted by their defining propositions.

The constructs of interest to us in this paper are "sensory constructs", and in particular the subset of "physical constructs", rather than social ones. Sensory constructs include the world depicted in a dream, or in a virtual-reality computer simulation, or by our ordinary waking experiences. The last of these is of special significance in our lives and I will label it the "physical construct." A "phenomenal binding" maps the entities of a physical construct into the mental world.

We may distinguish two kinds of phenomenal binding, which I will call "representing" and "embodying." To motivate this distinction, consider two examples.

- When I look at my laptop, I have visual images corresponding to the laptop, varying in shape as I look from different angles. This association I am calling the "representing binding": a set of sensory images is bound to an object in the physical construct and serves to represent the object. When I dream during sleep, I behold images that do not have that binding. If I dream of my laptop, I might have the same images as when I look at the real thing, but they are not bound to the physical laptop, but rather to a "dream laptop" in the dream construct, an ephemeral and unstable simulacrum of the physical construct.
- Second, one special object has a dual relationship to the contents of my mind. Namely, my brain. If I open up my skull, and peer at the grey matter through a mirror, I have a series of visual images that have a representing binding to the brain. My sensory impressions represent the brain in the same way that other impressions represent my laptop. In addition, the contents of my mind have a more intimate connection with my neural activity: I shall call this the "embodying binding." My physical body is my mind's "avatar" in the physical construct. If I had no embodying binding, then I could have no avatar, and I would be just a disembodied observer.

Although a construct is virtual, rather than actual, in many cases we may talk as if it were real. We can do this if, and to the extent that, the construct is grounded in epistemically accessible facts. Take, for example, the constructed world of Sherlock Holmes. A reading group who are studying the novels can talk as if 221B Baker Street exists, and the address is thus grounded. Holmes' shoe size, on the other hand, is not stated, and it would make no sense to try to refer to it. As for the physical construct, we can talk as if it were real because it is grounded in phenomenal reality. There is a fact of the matter whether I am sitting on this physical chair, just because we construe this as expressing experiential facts. When physicists depart from phenomenal observation, as when they talk about unobservable parallel universes, it no longer makes sense to talk as if they were real.

Formally, then: an embodying binding maps constant terms of the physical construct into the set of phenomenal contents of the mind; and maps the functions into mental operations; and the predicates into facts of those phenomenal contents. The intention is that this binding defines a relationship between the mind and its embodiment. In contrast, a representing binding maps constant terms in

the physical construct into the power set of the set of phenomenal contents (that is, it maps a constant term to multiple instances of sets of phenomenal content). The intention is that these terms correspond to observable bodies in the physical domain, and each of the target sets of phenomenal content corresponds to one perception of that body.

In the practice of physical science, it makes no difference whether we use the physical construct or the physical universe. The additional baggage that physicalism attaches to physics—namely the claim that physics concerns a mind-independent reality—is thus otiose: the question of its existence does not admit of a falsifiable answer and therefore lies outside the precinct of science.

Let us say that a construct is "grounded" if it has a binding to a substrate that is epistemically accessible. The physical construct is grounded because it has bindings to the mental world, which is known by direct acquaintance.

Thus, we are led to an internally consistent explanatory framework for the mind-body problem: (a) minds are the only reality, and (b) what we take to be a physical universe is actually a construct, which is an ontologically and epistemically grounded virtual world that enables us to model our experiences.

2.3.3. A Mind-Independent World Is Not Referenceable

With the concept of a "construct", we can now articulate the second limb of Berkeley's semantic argument. Pearce (2014, 2017) presents similar reasoning, but using the less specific terminology of "quasi-entities" for the contents of what I call the "physical construct".

The argument hinges on the act of referring. We can grasp this concept from straightforward cases of reference in computer science. A computer program has an address of a memory location, and uses the address as a reference by storing data into, and retrieving data from, that location. Thus, the act of reference is an act of reaching out to something by a determinate means. In like manner, I can refer to parts of my sensory field. For example, I can refer to the blue strip above the text I am writing in my visual image of my computer screen. I can also refer to elements of a construct, provided that it has a grounded binding. For example, I can refer to the chess pieces in a game I am playing. Although the King is an abstraction with no independent existence, I can refer to it, insofar as my reference is construed via the binding onto, say, the particular wooden pieces on a wooden board in front of me. If I assert that my King has taken my opponent's Queen, then my notional reference is to the abstract entities in the game construct of chess: but the substantive reference—the "cash value" of this statement—is to be found in the physical binding: I mean that my King piece of wood has moved into the little square where my opponent's Queen piece of wood previously stood, and the latter piece has been taken off the board. The effective meaning of the statement is the state of affairs on the physical board. (In turn, the reference to the physical pieces and board resolve into references to my perceptions of them.)

Likewise, I can refer to elements in a physical construct. If I say I am sitting on

a chair, then the notional reference is to my physical body and the physical chair (in the construct), but the substantive reference is construed as being to my phenomenal content (via the semantic binding). It means that I can see and feel my body on the chair: the statement that I am sitting a chair is capable of being "cashed out" in phenomenal experience, even if it is impracticable to enumerate all the individual experiences. An objection is that this conflates epistemology with ontology. In fact, the physical construct is "grounded ontologically and not just epistemically" in the phenomenal world (Robinson, 2009, Section §10.2.3). This echoes Berkeley's strategy to "collapse the truth conditions for a claim into the evidence for the claim" (Pearce, 2014, Section §9.3.1).

For illustration, Lloyd (2003) and Chalmers (2005) both argued that, in the Wachowskis' (1999) film *The Matrix* (as in traditional brain-in-a-vat thought-experiments), statements that notionally refer to an external physical world are effectively referring to a simulated world implemented in a computational rather than physical substrate. My present argument extends that line of thinking to show that notional physical references are effectively referring to the construct, and hence substantively referring to the phenomenal world that supports the construct.

It is possible to make, and understand, references to things within the physical construct just because the phenomenal semantic binding gives us a determinate means of making the references operational, of giving them meaning or "cash value." In contrast, an attempt to refer to a supposed mind-independent physical universe is a dangling reference. The mind-independent physical world is, by definition, not part of our direct experience. We are left with no determinate means of operationalizing a reference to it.

In all practical matters, both in everyday life and in the physical sciences, when people refer to physical things, they are actually referring to the physical construct. The only people who pretend to be referring to a mind-independent physical universe are philosophers, who attempt—and fail—to refer to something for which there is no mechanism of reference.

We have an almost unconquerable intuition that the physical domain to which we refer—namely the construct—somehow possesses a mind-independent reality, because we conflate the physical construct with whatever external reality is driving our experiences of the natural world. That is, however, a conflation that we must resist.

Pearce (2014) uses the terminology of "quasi-referring" to physical "quasi-entities", whereas here I prefer to say that we refer to entities in the physical construct but with the proviso that this is an indirect reference via a semantic grounding; but we agree that there can be no reference to a mind-independent domain. This terminology fits better with the linguistic practices around artificial virtual reality and lets us import the conceptual framework of virtuality into Berkeleyan philosophy.

Chalmers (2018, personal communication) offered the following objection. If

physical terms are topic-neutral, why can't they just "pick out" whatever is "playing" those roles, even if they are parts of a mind-independent world? But "picking out" is not something that words do by themselves. A word is deemed to "pick out" something insofar as a human uses it to pick something out. Chalmers seems to want "picking out" to be an autonomous activity with no human intervention in the loop. To say that there might be unobservable somethings that exist independently of the mind and that our words "pick out" those somethings without our instrumentally instructing the words to do so is to impute a kind of magic to words. As if the "hylous bulb" could pick out some unobservable and unknowable thing without our knowing anything about it.

A semantic binding, or in Chalmers' terms, a topic-neutral term's picking out a thing that plays that role, does not give us a mechanism for referencing an unknowable reality. We cannot assign a semantic binding to something that we cannot, in principle, refer to. As nobody can refer to any putative mind-independent substrate, it follows that a topic-neutral formal system cannot pick out entities in a mind-independent substrate, either.

Denial of the existence of the physical world, as opposed to mere agnosticism about its existence, is justified because any attempt to say that the physical world *might* exist turns out to be vacuous. Whatever might exist beyond our personal experiences, it cannot be a "physical universe" because that term does not define any intrinsic qualities. Suppose that there were to exist an undiscoverable noumenon of unknown nature that happened to be structurally isomorphic to the physical world: could we say that that was what physicalists have been referring to all along? No, because any such reference has not been part of the intention of the term. To give a trivial illustration of this principle: suppose that one day, a voyager on the High Seas were to discover a land calling itself Brobdingnag, populated by giants precisely as in Jonathan Swift's yarn. Would that mean that Swift was referring to that island all along, even though he did not know it? No, because that was never Swift's intended reference.

What the "physical universe" denotes is notional because it is constructed within a formal system without grounding in observation. Therefore, whatever exists beyond our experiences, it is not what we have been calling the physical universe. Hence, the physical universe cannot be said to exist.

2.3.4. Fundamental Reality Consists of Conscious Minds

Bringing these strands of argument together, we find: first, the domain of consciousness is not reducible to physics; second, the physical domain to which we ordinarily refer is a construct grounded in our phenomenal content; third, any attempt to refer to a mind-independent physical world is incoherent. We must conclude that reality comprises only conscious minds, and that what we take to be the physical world is a construct within this mental world.

Corollary 1: Most experiences are not under our volitional control, and it is implausible to suppose that their regularities are due to chance. We can rule out the physical construct as the cause, as it is a derivative of the mental world and

therefore cannot be the source of the observed natural order. Also, we can rule out the mind-independent physical world as it is an incoherent notion. So the only option left on the table is that the natural order is driven by something of the nature of a mind, which I have referred to earlier as the "metamind."

Corollary 2: The decombination problem: how do our private personal minds stand in relation to the mental universe? Chalmers (2018) used the term "cognitive fragmentation" for this partitioning of the mental universe into personal minds. The Advaita Vedanta anatomizes the individual as a series of sheathes surrounding the Atman, which is non-different from the Brahman. This is picturesque but has no obvious explanatory power. Likewise is Shani's (2015) metaphor of the personal mind as "a 'vortex' surging from the oceanic background" of cosmic consciousness—a metaphor that brings a lot of baggage from fluid dynamics without offering any account of the structure and dynamics of the connection between personal consciousness and the metamind.

Kastrup & Kelly (2018) drew an analogy from psychiatry: just as an individual may dissociate into multiple personalities (in Dissociated Identity Disorder, DID), so the metamind is supposed to dissociate into the personal minds of people (and, I assume, animals and disembodied beings). This is vivid but explanatorily unrewarding. Insofar as dissociation involves the partitioning of a larger system of experiences into private sub-systems, each accompanied by a subjective awareness, personal minds are indeed dissociations from the metamind. In human psychiatry, however, this phenomenon is driven by a malfunction of the mind, often produced by psychological trauma. How is a dissociation of the metamind produced and maintained? The absence of explanatory power in this analogy was evident in Kastrup & Kelly (2018), "[I] f something analogous to DID happens at a universal level, the one universal consciousness could, as a result, give rise to many alters with private inner lives like yours and ours;" and, as clinical dissociation is correlated with changes in brain function, "… We posit that this appearance is life itself: metabolizing organisms are simply what universal-level dissociative processes look like". This says nothing over and above the bare fact that organisms have private streams of consciousness.

Instead of these metaphors, I suggest an extension of Lloyd's (1999) reductionist picture of mental monism:

- "In set-theoretic terms, the mental universe is defined as the union of all existing minds. So, it contains the conscious minds of all human beings and all animals in the world we see around us. It will also contain any disembodied minds, if they exist. It also contains streams of mental activity that govern the complete panoply of what we think of as natural phenomena … It is a key point that the mental universe is a union of minds, and not a set of minds. For instance, if A = $\{a_1, a_2, a_3, \ldots\}$ and B = $\{b_1, b_2, b_3, \ldots\}$ are minds comprising experientiae $a_1, a_2, a_3, \ldots$ and $b_1, b_2, b_3, \ldots$ then the union is U = $\{a_1, a_2, a_3, \ldots, b_1, b_2, b_3, \ldots\}$ whereas the set of minds is {A, B}". (§6.4),
- "An 'ordinary mind' is a subset of the metamind, closed under ordinary oper-

> ations of mental access. By 'operation' I mean any of the actions that a mind can carry out on its contents such as perceiving, thinking, recalling memories, imagining things, focusing, paying attention, and so on. By being 'closed' I mean that any operation carried out on any of the contents yields some new contents that are still within that mind. So, whenever you recollect a memory, or focus on some bodily sensation, you remain within your own mind and do not slip into someone else's mind". (§6.5)

Within mental monism, all mechanisms must reduce to conscious volition and qualia. There is no non-mental substrate that can serve to explain vortices or dissociations. *A fortiori*, the individuation of personal minds must be accounted for in those terms, namely volition and qualia. (By the way, I write here of individuation of minds, not of subjects. The agent of the mind cannot be individuated as it is featureless; but the contents of the mind obviously are featured. Hence the mind as a whole is individuated by virtue of the contents being individuated.) Instead of metaphors from fluid dynamics and psychiatry, a computing analogy seems more apt as it addresses the same logical problems that face the metamind. In a multi-user computer, the operating system partitions memory into private areas, and each user can access only her own section of memory. Processes in different user areas communicate only through defined channels.

Corollary 3: According to mental monism, the brain cannot cause anything to happen in the conscious mind, because the brain has no real existence. This is often cited as a knock-down objection, as brain events of several kinds—trauma, drugs, electrodes—appear to affect the mind. This objection, however, rests on a misunderstanding of the theory.

Mental monism has a universal set of minds $U = \{E_0, E_1, E_2, \ldots\}$ where E_0 is the metamind and $E_1, \ldots$ are whatever other minds exist, including people, animals, inanimate objects, and disembodied minds. Causation operates within and between these minds. A subset comprises minds that are rendered as observable objects, $U_{obs} \subset U$, with an archetype E_i rendered as a physical entity P_j = ren (E_i). That is to say, the archetype of any physical object, animate or inanimate, is a component of the metamind that controls the rendering of the object in personal minds. There is a close analogy to the object-oriented architecture of artificial virtual-reality software. What appears within the manifest world as causation between physical entities is actually a rendering of the real causation in mental world. It is the "shadow causation." In rough terms, the scheme is as follows. Suppose E_m is my mind, ren (E_m) is my brain, E_{LSD} is the archetype-mind of a tab of LSD and ren (E_{LSD}) is the tab of LSD itself. Ingesting the LSD allows E_{LSD} to act causally on E_m, producing hallucinations. A third-person observer will see ren (E_{LSD}) entering ren (E_m) and consequent changes occur in ren (E_m), namely different brain activity and reports of hallucinations. See **Figure 1**.

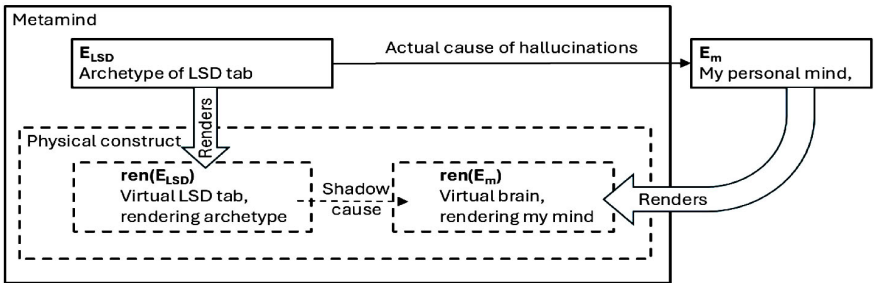

Figure 1. Schematic showing actual versus shadow causation.

Corollary 4: The brain is the mind's avatar in the physical construct, and it needs two interfaces. On the one hand, it has the sensory and motor organs plus the brain tissue that carries out pre-conscious input processing and post-conscious output processing. On the other hand, it has the physical correlate of consciousness, which acts as a portal to the conscious mind. Thus, the brain is an interface between virtual transducers and actuators on the one hand and the correlate of consciousness on the other; and this interface must operate within the laws that rule the physical construct, hence it has to be a physical object with specific characteristics.

3. Conclusion

Reality comprises only conscious minds, and the physical world is a construct grounded in the regularities of conscious experience, these regularities being driven by a background conscious mind, referred to as the "metamind". The conscious minds of humans and other animals are embodied in avatars inside the physical construct. As the physical world in general, and the avatars in particular, obey the laws of physics, a conscious mind is able to interact with its avatar only through nondeterministic physical processes. The structures that execute this interaction are referred to as "portals," which are discussed in the accompanying paper (Lloyd, n.d.).

Acknowledgements

This paper was improved by comments by Bill Adams, David Chalmers, John Gregg, and David Field on earlier drafts. The anonymous referee's comments on the final draft were much appreciated.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Berkeley, G. (1710). *The Principles of Human Knowledge* (Volume I). George Pepyat.

Berkeley, G. (1713). *Three Dialogues between Hylas and Philonous.* Henry Clements.

Chalmers, D. (1996). *The Conscious Mind.* Oxford University Press.

Chalmers, D. (2013/2015). Panpsychism and Panprotopsychism. In T. A. Alter, & Y. Nagasawa (Eds), *Consciousness in the Physical: World Perspectives on Russellian Monism* (pp. 1-32). Oxford University Press. http://consc.net/papers/panpsychism.pdf

Chalmers, D. (2018). Idealism and the Mind-Body Problem 1. In W. E. Seager (Ed.), *The Routledge Handbook of Panpsychism* (pp. 353-373). Routledge. https://doi.org/10.4324/9781315717708-28

Chalmers, D. J. (2005). The Matrix as Metaphysics. In C. Grau (Ed.), *Philosophers Explore the Matrix* (pp. 132-176). Oxford University Press. https://doi.org/10.1093/oso/9780195181067.003.0009

de Silva, F. (1995). Consciousness and Special Relativity? (Letter). *IEEE Engineering in Medicine and Biology, 14,* 358.

de Silva, F. (1996). Consciousness and Special Relativity? *IEEE Engineering in Medicine and Biology, 15,* 21-26.

Dennett, D. (2013). *Intuition Pumps and Other Tools for Thinking.* Allen Lane.

Descartes, R. (1641/1975). *Meditations on First Philosophy, in Which the Existence of God and the Immortality of the Soul Are Demonstrated* (J. Veitch, Trans.). Dent.

Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Annalen der Physik, 322,* 891-921. https://doi.org/10.1002/andp.19053221004

Foster, J. (1982). *The Case for Idealism.* Routledge & Kegan Paul.

Foster, J. (1991). *The Immaterial Self.* Routledge.

Foster, J. (2008). *A World for US: The Case for Phenomenalistic Idealism.* Oxford University Press.

Gallois, A. (1974). Berkeley's Master Argument. *The Philosophical Review, 83,* 55-69. https://doi.org/10.2307/2183873

Gibbins, P. F. (1985). Are Mental Events in Space-Time? *Analysis, 45,* 145-147. https://doi.org/10.1093/analys/45.3.145

Hameroff, S., & Penrose, R. (2014). Consciousness in the Universe: A Review of the 'Orch OR' Theory. *Physics of Life Reviews, 11,* 39-78. https://doi.org/10.1016/j.plrev.2013.08.002

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly, 32,* 127-136. https://doi.org/10.2307/2960077

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy, 83,* 291-295. https://doi.org/10.2307/2026143

James, W. (1890). *Principles of Pyschology.* Henry Holt & Co.

Jaskolla, L. J., & Buck, A. J. (2012). Does Panexperiential Holism Solve the Combi-Nation Problem? *Journal of Consciousness Studies, 19,* 190-199.

Kastrup, B., & Kelly, E. F. (2018). *Could Multiple Personality Disorder Explain Life, the Universe and Everything?* Scientific American Blog. https://blogs.scientificamerican.com/observations/could-multiple-personality-disorder-explain-life-the-universe-and-everything/

Lee, G. (2007). Consciousness in a Space-Time World. *Philosophical Perspectives, 21,* 341-374. https://doi.org/10.1111/j.1520-8583.2007.00130.x

Lloyd, P. B. (1999). *Consciousness and Berkeley's Metaphysics.* Self-Published.

Lloyd, P. B. (2003). *Exegesis of the Matrix.* Self-Published.

Lloyd, P. B. (2006). Mental Monism Considered as a Solution to the Mind-Body Problem. In A. Batthyany, & A. Elitzur (Eds), *Mind and its Place in the World* (pp. 101-144). de

Gruyter. https://doi.org/10.1515/9783110325683.101

Lloyd, P. B. (n.d.). Embodying Consciousness in Brains and Machines II: Portals. *Open Journal of Philosophy*.

Lockwood, M. (1984a). Einstein and the Identity Theory. *Analysis, 44,* 22-25. https://doi.org/10.1093/analys/44.1.22

Lockwood, M. (1984b). Reply to Gordon. *Analysis, 44,* 127-128. https://doi.org/10.1093/analys/44.3.127

Lockwood, M. (1985). Einstein, Gibbins and the Unity of Time. *Analysis, 45,* 148-150. https://doi.org/10.1093/analys/45.3.148

Lockwood, M. (1989). *Mind, Brain, and the Quantum: The Compound 'I'.* Black-Well.

Mathews, F. (2011). Panpsychism as Paradigm. In M. Blamauer (Ed.), *The Mental as Fundamental* (pp. 141-156). de Gruyter. https://doi.org/10.1515/9783110319859.141

Nagasawa, Y., & Wager, K. (2016). Panpsychism and Priority Cosmopsychism. In G. Brüntrup & L. Jaskolla (Eds), *Panpsychism* (pp. 113-129). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199359943.003.0005

Pearce, K. L. (2014). *Language and the Structure of Berkeley's World.* University of Southern California.

Pearce, K. L. (2017). *Language and the Structure of Berkeley's World.* Oxford University Press.

Pelczar, M. (2017). *Sensorama: A Phenomenalist Analysis of Spacetime and Its Contents.* Oxford University Press.

Phillips, I. B. (2009). *Experience and Time.* Ph.D. Thesis, University College London. http://discovery.ucl.ac.uk/16139/1/16139.pdf

Popper, K. (1959). *The Logic of Scientific Discovery.* Hutchinson & Co.

Robinson, H. (2009). Idealism. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds), *The Oxford Handbook of Philosophy of Mind* (pp. 189-205). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199262618.003.0011

Rosenberg, G. H. (2004). *A Place for Consciousness.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168143.001.0001

Russell, B. (1926). The World of Physics and the World of Sense. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy* (Chapter 4: pp. 101-128). George Allen & Unwin.

Russell, B. (1927). *The Analysis of Matter.* Kegan Paul.

Searle, J. (1995). *The Construction of Social Reality.* Simon and Schuster.

Shani, I. (2015). Cosmopsychism: A Holistic Approach to the Metaphysics of Experience. *Philosophical Papers, 44,* 389-437. https://doi.org/10.1080/05568641.2015.1106709

Tolman, R. C. (1917). *The Theory of the Relativity of Motion.* University of California Press.

Wachowski, L., & Wachowski, A. (1999). *The Matrix* [Video Recording]. Warner Brothers.

Weingard, R. (1977). Relativity and the Spatiality of Mental Events. *Philosophical Studies, 31,* 279-284. https://doi.org/10.1007/bf01855233