Aporia vol. 34 no. 2–2024

An Anti-physicalist Argument From the Unity of Consciousness

Suyu Wang

In the Sixth Meditation, René Descartes famously argues that the mind and the body must be two distinct kinds of substances because the mind is indivisible, yet the body is divisible (AT VII 86-87: CSM II 59). Though Descartes's indivisibility argument presumably fails to establish his mind-body dualism, it nevertheless suggests that there is something special about the mind that physicalist theories cannot easily accommodate. This special feature can be characterized by the notion of the (synchronic) phenomenal unity of consciousness, according to which all the conscious experiences of a subject form a unified whole at any given moment. In this paper, I will argue that the phenomenal unity of consciousness has a mereologically inverted part-whole structure (in the sense that the unity is metaphysically prior to its parts, namely the individual conscious experiences), whereas no physical aggregate does. From this I will conclude that we should reject the popular physicalist thesis that everything in the world is an aggregate of physical entities. The rest of my paper will proceed as follows: In Section I, I will discuss my formulation of physicalism, the unity of consciousness, and mereological inversion to set up the discussion. In Section II, I will present my anti-physicalist argument based on the unity of consciousness and defend

Suyu Wang is a junior at the University of Wisconsin–Madison, majoring in mathematics and philosophy with a minor in political science. His philosophical interests lie in the philosophy of mind, philosophy of science, epistemology, and political philosophy. After graduation, he plans to pursue a PhD in philosophy, with the goal of contributing to academic philosophy research.

each premise. In Section III, I will respond to several potential objections against my argument. In Section IV, I will make some concluding remarks.

Section I: Physicalism, Unity of Consciousness, and Mereological Inversion

Physicalism is a popular and intuitive solution to the mind-body problem in contemporary philosophy of mind. In a slogan, physicalism is the thesis that everything in the world is physical. While it admits a variety of distinct characterizations, the version of physicalism which I will primarily discuss in this paper is the following:

> **Ontological Physicalism:** All things that exist in the world are necessarily aggregates of fundamental physical particles in accordance with physical laws.[1]

Here "physical" means "non-mental." In particular, fundamental physical particles are assumed to be the basic entities that serve as the building blocks for other larger entities and that do not have mental essences or irreducible mental properties. Contemporary physics does not clearly suggest that the basic entities of reality have any mental essence or property, so our current physics is compatible with ontological physicalism. However, if it turns out that the basic entities posited by the future, complete theory of physics necessarily have mental essences or irreducible mental properties, then ontological physicalism automatically fails.[2] I focus on ontological physicalism because it is accepted by many contemporary physicalists and implicitly assumed by many paradigmatic physicalist solutions to the mind-body problem, such as the psycho-neural identity theory and empirical functionalism.[3]

The phenomenon that I find incompatible with ontological physicalism is the unity of consciousness in human beings. Consider the following scenario: When I drafted this paper, I stared at my laptop's screen and

[1] Some physicalists only defend a weaker version of physicalism: Instead of arguing that all things (which may include, for example, Platonic universals or numbers) are physical, they merely claim that all mental substances, properties, events, and/or facts are fundamentally physical. I will set aside this detail given our focus on the mind-body problem.

[2] I do not define "physical particles" and "physical laws" as whatever the most advanced physical theory assumes them to be in order to avoid the so-called Hempel's dilemma (1980). Similar attempts to define physicalism can be found in Papineau (2002) and Goff (2017).

[3] Typically, the psycho-neural identity theorist claims that mental states are identical to certain physical-biological structures (Smart 1959), and the empirical functionalist submits that mental states are functional states realized by physical-biological structures (Putnam, 1967). It is natural for such physicalists to hold that any entity one might find in the actual world is composed of minuscule fundamental physical particles, as suggested by our physical science.

listened to the music in my headphones. I felt a slight headache from staying up late the night before, so I drank some coffee, which left a lingering bitter taste in my mouth. Now, for all such conscious experiences I have, I am not having each of them separately.[4] Instead, I am experiencing a unified whole of all experiences. There are different ways of understanding this notion of a "unified whole." In this paper, I understand it as the phenomenal unity of conscious experiences, or phenomenal unity in short. More precisely, I will follow Nagel (1974) and make the following definitions: A mental state is said to be phenomenally conscious if there is something that it is like to be in that state, and certain conscious experiences are said to be phenomenally unified if there is something that it is like to feel the unified whole of experiences in addition to each individual experience. Thus, in the scenario above, my headache is a phenomenally conscious experience because there is something that it is like for me to be in the state of headache (in contrast, there is presumably nothing that it is like to have a belief or thought); moreover, my pain experience, auditory experience, visual experience, and gustatory experience are phenomenally unified so long as the phenomenal aspects of those experiences in my mind are not exhausted by the phenomenal aspects of all individual experiences.

Intuitively, there is a close relation between phenomenal unity and consciousness. Tim Bayne (2010) characterizes this relation with the Unity Thesis: Necessarily, all conscious experiences of a subject are phenomenally unified at a given moment. The Unity Thesis appears plausible on the introspective grounds that, for example, when I see and touch fire and feel pain at some moment, I can conceive of my visual experience and pain experience only conjointly. This subject-dependent formulation, however, appears incompatible with empirical cases like split-brain patients, who seem to exhibit two phenomenally disunified streams of consciousness in a single subject due to the biological disconnection of the left and right hemispheres of the brain. In contrast, Farid Masrour (2020) advocates the following subject-independent account:

> **Pure Extrinsic Unity**: Necessarily, all conscious experiences are extrinsically unified, where "an experience, E, is extrinsically unified iff there is some experience, F, such that F is numerically distinct from E, is unified with E, and is neither a part of E nor contains E as a part" (219).

This account avoids the problem of split brains by not alluding to the notion of subjects in its description of unified conscious experiences. Hereafter, by phenomenal unity, I am referring to Masrour's Pure Extrinsic Unity.

[4] I assume that we have an available individuation scheme of conscious experiences.

The tension between phenomenal unity and ontological physicalism is best explained in terms of mereological inversion. For most ordinary composite objects, it is clear that some of their parts could exist independently of their wholes.[5] In the "Transcendental Aesthetic" of the *Critique of Pure Reason*, however, Kant argues that the parts of one's determinate representation of space could not exist independently of the whole because whenever one's mind represents a determinate or bounded spatial region, that representation exists only as part of the representation of a single overarching space (A25/B40). In contemporary terminology, we might say that one's determinate representation of space is mereologically inverted. More precisely, I suggest the following definition of mereological inversion:

> **Mereological Inversion**: An object is mereologically inverted if and only if none of its parts could exist prior to or metaphysically independently from its whole.

Section II: An Anti-physicalist Argument from Phenomenal Unity

In this section, I will argue that ontological physicalism should be rejected because it fails to accommodate the mereologically inverted nature of our phenomenal unity. My argument can be laid out as follows:

1. Necessarily, all conscious experiences are phenomenally unified.
2. All conscious experiences are necessarily phenomenally unified only if phenomenal unity is mereologically inverted; that is, only if individual conscious experiences could not exist metaphysically independently from their phenomenal unity.
3. Therefore, phenomenal unity is mereologically inverted. [From 1 and 2]
4. No aggregate of fundamental physical particles is mereologically inverted.
5. Anything that is mereologically inverted is not necessarily identical to or composed of something that is not mereologically inverted.

[5] I assume that mereological nihilism is false.

6. Therefore, phenomenal unity is not necessarily identical to or composed of aggregates of fundamental physical particles. [From 3, 4, and 5]
7. If ontological physicalism is true, then phenomenal unity is necessarily identical to or composed of aggregates of fundamental physical particles.
8. Therefore, ontological physicalism is false. [From 6 and 7]

I will now defend each premise. Premise 1 is primarily supported by the introspective evidence I provided briefly in Section I. Premise 7 follows directly from the definition of ontological physicalism.

Given that there exists a necessary phenomenal unity of all conscious experiences, I claim that this unity is mereologically inverted (so Premise 2 is true). To begin with, I argue that we can adopt the following criterion for determining mereological inversion:

> **Criterion for Mereological Inversion**: For a whole *W* and some of its parts *P*, the part-whole relation between *P* and *W* is mereologically inverted if whenever I conceive of *P* (to the best of my knowledge and with well-functioning cognitive faculties), I must thereby conceive of *W* and conceive of *P* as part of *W*.

The validity of this criterion ultimately boils down to the idea that there is an intimate connection between conceivability and possibility. If someone with normal rational capacities and complete knowledge of relevant concepts cannot conceive of *P* as existing without *W*, then we have a good reason to believe it is impossible that *P* could exist without *W*, because the metaphysical impossibility of *P* without *W* seems to be the best explanation for the inconceivability of *P* without *W*. For example, the simplest and most intuitive explanation for the fact that I cannot conceive of any determinate region of space without embedding it in the whole overarching space seems to be that it is simply impossible that a determinate region of space could exist independently from the whole space.

Having provided some justification for this criterion, I argue that phenomenal unity satisfies it, which entails that phenomenal unity is mereologically inverted and hence warrants Premise 2. Suppose that all experiences are necessarily phenomenally unified. Then the case for phenomenal unity and the case for Kant's spatial representations seem completely analogous: Given that there is always a phenomenal unity of individual experiences, it is no less difficult to conceive of any individual experience without the "background" phenomenal unity than to conceive of a determinate representation of space without the "background"

overarching space. Indeed, it is not implausible to say that it is by the nature of the mechanisms of our mind that, when conceiving of conscious experiences, we conceive of them against a unified phenomenal whole. If we conceive individual experiences necessarily as parts of their phenomenal unity, it follows from our criterion that we should accept that phenomenal unity is mereologically inverted.

On the other hand, aggregates of fundamental physical particles are not mereologically inverted. The legs of a chair do not depend for their existence on being part of the chair. Similarly, the fundamental physical particles that compose some neuron in a human brain do not depend for their existence on being part of the neuron. One might retort that the chair legs are not chair legs anymore if the chair is destroyed, and in this sense, their status as chair legs does depend for their existence on being part of the chair. I think, however, that when the chair is destroyed, its chair legs as aggregates of physical particles do not cease to exist, and we may still properly call them chair legs, even though they no longer belong to any complete chair. Hence, no aggregate of fundamental physical particles is mereologically inverted, so Premise 4 is true.

Premise 5 (a mereologically inverted thing cannot be necessarily identical to or composed by non-mereologically inverted things) follows quickly from the definition of mereological inversion: Let *A* be a mereologically inverted object and *B* be an object that is not mereologically inverted. Assume for contradiction that *A* is necessarily identical to *B*. In the language of possible worlds, since *B* is not mereologically inverted, there is a possible world *W* in which the parts of *B* exist, but the whole of *B* does not. The existence of these parts of *B* implies that some parts of *A* would exist in *W*. Because *A* is mereologically inverted, the whole of *A* must exist. Given that *A* and *B* are identical in all possible worlds, the whole of *B* exists, too. Contradiction. Thus, *A* is not necessarily identical to *B*. Likewise, assume for contradiction that *A* is entirely and necessarily composed by (finitely many) objects $B_1, \ldots, B_n$ where there exists at least one integer $k \in \{1, \ldots, n\}$ such that B_k is not mereologically inverted. Then, there exists some possible world W' in which the whole of B_k does not exist, yet all B_k's parts exist, and the wholes and parts of B_j exist for all $j \in \{1, \ldots, n\} \setminus \{k\}$. Due to the non-existence of the whole of B_k, the whole of the object composed by $B_1, \ldots, B_n$ should not obtain in W'. But the existence of the parts of all B_i for $i \in \{1, \ldots, n\}$ implies that some parts of *A* exist in W', so the whole of *A* exists in W' by its mereological inversion, or the whole of the object composed by $B_1, \ldots, B_n$ exists in W', a contradiction. Thus, *A* is not necessarily composed by $B_1, \ldots, B_n$. I conclude that *A* cannot be necessarily identical to *B* or composed by $B_1, \ldots, B_n$ if *B* is not mereologically inverted or if $B_1, \ldots, B_n$ are not all mereologically inverted. In other words,

things that are mereologically inverted are not necessarily identical to or composed by things that are not, so Premise 5 holds true.

To summarize, Premises 1–2 entail that phenomenal unity is mereologically inverted, and Premises 4 states that aggregates of physical particles are not. Because things that are mereologically inverted are not necessarily identical to or composed of non-mereologically inverted things (Premise 5), phenomenal unity is not necessarily identical to or composed of any aggregate of physical particles. However, ontological physicalism entails that all existent objects are necessarily aggregates of physical particles (Premise 7), so I conclude that ontological physicalism is untenable.

Section III: Objections and Replies

In this section, I will consider some objections against my argument. The most immediate one perhaps comes from functionalism. A functionalist who is committed to ontological physicalism may deny Premise 4 (no aggregate of fundamental physical particles is mereologically inverted) by arguing that the functional whole of my mind is mereologically inverted. For the sake of simplicity, suppose that my mental property M is realized by some physical property *P* of my brain. On one common functionalist conception, I am in M in virtue of the fact that *P* obtains as a physical realizer of M, that is, *P* plays a particular role in an appropriate causal network constituted by my sensory inputs, mental states and processes, and behavioral outputs. Consequently, the existence of *P* is metaphysically dependent upon the existence of the whole causal network; if the causal network within which *P* is situated changed, *P* would presumably not be part of the causal network that gives rise to M[6] and, therefore, not obtain as a physical realizer of M. The functionalist might thus conclude that *P* is a mereologically inverted aggregate of fundamental physical particles.

In response, I argue that a functionalist theory like this can succeed in denying Premise 4 only insofar as it abandons its commitment to ontological physicalism. The problem for the functionalist is that the physical realizer *P* by itself does not seem to be mereologically inverted. Given that *P* is nothing more than the aggregate of certain physical substrates with physical properties, *P* does not seem to constitute a

[6] More rigorously, a mental property M can be defined as a secondary property specifying a causal network between sensory inputs, behavior outputs, and other mental properties such that a subject is in M at time *t* just in case it has physical properties that instantiate all the roles in the causal network at *t*.

counterexample against Premise 4 because each part of *P* could clearly exist independently of the whole (for instance, if *P* is composed of neurons X_1 and X_2, then intuitively X_1 could exist without being part of *P*).

Hence, the functionalist who intends to deny Premise 4 will have to claim that the causal network is not just the aggregate of all its physical properties of substrates. Indeed, this claim is more consistent with the functionalist spirit: when the functionalist says that *P* depends for its existence on the whole causal network, it is the existence of *P as a realizer of* M (that is, as a physical property that plays a certain causal role in the network) that depends on the whole network. Nevertheless, if the functionalist really thinks so, then the conclusion they should draw is not that some physical composites are mereologically inverted. Rather, they should conclude that the functional whole of one's mind, which is *not* just an aggregate of fundamental physical particles, is mereologically inverted. Now, it becomes clear that the functionalist accounts for our phenomenal unity precisely by introducing a mereologically inverted element (i.e., the functional whole) into their theory. Although the functionalist is able to accommodate the mereologically inverted nature of phenomenal unity now, they succeed in doing so only at the cost of implicitly denying ontological physicalism, because they are forced to acknowledge that the functional wholes are not aggregates of fundamental physical particles. Therefore, I conclude that many popular formulations of functionalism are not really physicalist, in the sense that they do not capture or necessitate the physicalist intuition that all objects are fundamentally physical objects.

Instead of appealing to functionalism, one might think a system of subatomic particles, such as electrons and neutrons, is mereologically inverted because being a subatomic particle is arguably just a matter of fulfilling all the relevant physical laws. This insight is often traced back to Bertrand Russell, who maintains that (contemporary) physics can only reveal the structural properties of the world and remains silent on the intrinsic nature or categorical basis that grounds those structural properties (1959, 7). Structural properties include, though are possibly not limited to, relational properties and dispositional properties. It seems that physical laws only tell us how physical entities change, not what these physical entities are. For instance, physics tells us what happens when a high-energy electron strikes a proton, but it does not reveal what instantiates all the dynamical properties of an electron. Thus, for something to be an electron is for it to instantiate the appropriate properties and obey the relevant laws specified by physics—that is, for it to play the "electron role." Assuming this picture of theoretical physics, one may argue that a system of subatomic particles is mereologically inverted: every subatomic particle in the system exists only in virtue of standing in the right causal relations to others and, therefore,

exists only in virtue of being part of the system. Just as the functionalist may insist that the physical realizer of a mental state is mereologically inverted because it is what plays the appropriate causal role specified by the mental state, one may insist that the particles are mereologically inverted because they are what play the appropriate causal role specified by the physical laws governing the system.

In reply, I can grant that micro-level systems of subatomic particles are mereologically inverted yet nevertheless maintain that macro-level systems, such as conscious human beings, should not be considered mereologically inverted through the lens of contemporary physics. Currently, our best physical theories do not handle objects of all sizes uniformly: particle-level systems are governed by quantum physics, whereas ordinary and very large objects behave in accordance with Newtonian mechanics and relativity theory. Admittedly, quantum physics attempts to understand quantum phenomena by studying the whole system of particles directly rather than by analyzing individual particles and then inferring the properties of the system, and nomic terms in quantum physics such as electrons and positrons can be defined functionally (as explained in the previous paragraph). If we adopt a naturalistic attitude toward ontology (namely adopting scientific theories as the most reliable guide to ontology), we should indeed accept that a micro-level particle system is mereologically inverted.

The same argument does not apply to macro-level systems though because many medium-sized objects, including most conscious creatures, are neither functionally definable nor necessarily treated as whole systems in Newtonian mechanics or relatively theory.[7] Thus, we may replace Premise 4 with Premise 4*:

> 4*. No macro-level aggregate of fundamental physical particles is mereologically inverted.

Here, a macro-level physical aggregate is defined as a physical entity governed by the laws of Newtonian mechanics or relativity theory, and a micro-level physical aggregate is defined as a physical entity governed by

[7] Note that some of the properties of macro-level conscious subjects, such as mass and electric charge, can be defined functionally via scientific laws. However, the consciousness of macro-level subjects seems to be an intrinsic property of theirs that cannot be defined in the same way.

quantum physics.[8] Since phenomenal unity is not mereologically inverted (Premise 3), and something that is mereologically inverted is not necessarily identical to or composed of things that are not (Premise 5), it follows that:

> 6*. Therefore, phenomenal unity is not necessarily identical to or composed of macro-level aggregates of fundamental physical particles.

Ontological physicalism entails that phenomenal unity should be some physical aggregate, whereas Premise 6* states that phenomenal unity cannot be a macro-level aggregate of fundamental physical particles. Consequently, phenomenal unity can only be a micro-level aggregate if ontological physicalism is true. In other words, ontological physicalism requires that phenomenal unity be explained in terms of the laws of quantum physics. However, it seems deeply confusing (if not incoherent) that the phenomenal unity of a human being should be understood as a micro-level system. Given that phenomenal unity is a property that arises in a macro-level biological system, any attempt to analyze it within a micro-level system seems to miss the mark. I conclude that it is difficult if not impossible for ontological physicalism to accommodate phenomenal unity.

Section IV: Conclusion

To recap, I propose an argument against ontological physicalism, the view that everything in the world is made up of fundamental physical particles. I argue that the necessary phenomenal unity of consciousness implies that the phenomenal unity is mereologically inverted, whereas composites of physical particles are not mereologically inverted. I conclude that conscious experiences cannot be identical to or composed purely of physical entities, thus disproving ontological physicalism.

The functionalist may object that the functional whole of my mind is a mereologically inverted physical entity. On closer inspection, though, it turns out that functionalism may succeed in undermining my argument only at the cost of implicitly rejecting ontological physicalism. Contemporary proponents of Russell's view on theoretical physics may claim that systems

[8] Such a way of defining "macro-level" and "micro-level" leaves open the conceptual possibility that a physical aggregate can be both macro-level and micro-level. Nevertheless, quantum physics and Newtonian mechanics are intended to characterize physical systems with very different natures, so it is safe to assume that no physical aggregate is explained by both quantum physics and Newtonian mechanics. Hence, every physical aggregate that can be explained by our physical theories is either macro-level or macro-level but not both.

of subatomic particles are themselves mereologically inverted due to the nature of quantum laws and the methodology of quantum physics. In reply, I distinguish micro-level systems from macro-level systems and insist that even if the former are *de facto* mereologically inverted, the latter are not, and phenomenal unity can be intelligibly understood only in terms of the latter. Ontological physicalism is still under serious threat.

Our intuition that everything is an aggregate of physical particles is undeniably compelling, yet it is surprisingly challenging to formulate and defend it once we bring in a broader range of philosophical and biological evidence. Certainly, the physicalist might deny altogether that physicalism requires everything to be composed of fundamental physical particles or insist that future physics express the structure of the world in a radically different way. Nonetheless, it would be unclear what makes physicalism a substantive thesis rather than mere optimism in the success of physical sciences.

Works Cited

Bayne, Tim. *The Unity of Consciousness*. Oxford UP, 2010.

Descartes, René. *The Philosophical Writings of Descartes*. Cambridge University Press, 1984.

Goff, Philip. *Consciousness and Fundamental Reality*. Oxford UP, 2017.

Hempel, Carl G. "Comments on Goodman's 'Ways of Worldmaking.'" *Synthese*, vol. 45, no. 2, 1980, pp. 193–99. JSTOR.

Kant, Immanuel. *Critique of Pure Reason*. Edited by Paul Guyer and Allen W. Wood, 1st ed., Cambridge University Press, 1998.

Masrour, Farid. "The Phenomenal Unity of Consciousness." *The Oxford Handbook of the Philosophy of Consciousness*, by Farid Masrour, edited by Uriah Kriegel, Oxford University Press, 2020, pp. 207–29.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, vol. 83, no. 4, Oct. 1974, p. 435.

Papineau, David. *Thinking about Consciousness*. Clarendon Press, 2002.

Putnam, H. "The Nature of Mental States." *Art, Mind, and Religion*, edited by W. H. Capitan and D. D. Merrill, UP of Pittsburgh, 1967.

Russell, Bertrand, and Nicholas Griffin. *My Philosophical Development*. 1st ed., Routledge, 2022.

Smart, J. J. C. "Sensations and Brain Processes." *The Philosophical Review*, vol. 68, no. 2, Apr. 1959, p. 141.