The first-personal argument against physicalism

Christian List*

Originally posted in July 2023 / revised in February 2024 / citable as a "Discussion Paper, LMU Munich"

Abstract: The aim of this paper is to discuss a seemingly straightforward argument against physicalism which, despite being implicit in much of the philosophical debate about consciousness, has not received the attention it deserves (compared to other, better-known "epistemic", "modal", and "conceivability" arguments). This is the argument from the non-supervenience of the first-personal (and indexical) facts on the third-personal (and non-indexical) ones. This non-supervenience, together with the assumption that the physical facts (as conventionally understood) are third-personal, entails that some facts – namely, first-personal, phenomenal ones – do not supervene on the physical facts. Unlike other arguments against physicalism, the first-personal argument, if successful, refutes not only physicalism but also other third-personal metaphysical pictures, including standard versions of dualism.

1. Introduction

Physicalism, which is at its core the thesis that everything supervenes on the physical, continues to be a leading view concerning the metaphysics of the mind.[1] In the most recent survey of professional philosophers conducted by David Bourget and David Chalmers, a majority of respondents (51.9%) described themselves as physicalists, while fewer than a third (32.1%) described themselves as non-physicalists; the rest held other views.[2] The motivating idea behind physicalism is that our world is at bottom physical and that all higher-level phenomena – from chemical and biological to psychological and social ones – are products of the complex organization and interplay of underlying physical properties.

From a physicalist perspective, we human beings are ultimately physical systems, albeit highly complex ones, and all the properties that are instantiated in us – including mental properties, from having beliefs and desires to being conscious – are either physical properties themselves or supervene on physical properties. We don't need to postulate anything non-physical, especially nothing that isn't the product of underlying physical properties, to account for the mind.

The challenge for any physicalist is to explain how exactly mental phenomena, especially consciousness, fit into a physicalist worldview. Physicalists tend to fall into one of three camps. First, there are a few who simply deny the reality of many or most of the mental properties that we conventionally postulate. The neurophilosophers Patricia and Paul Churchland may be examples in this camp. But most physicalists accept that mental properties are real and either

* Versions of this paper were presented at a Research Seminar in Decision and Action Theory at LMU Munich, April 2023, a Workshop on the "Objects of Credence" at the London School of Economics, May 2023, as part of the Wendy Huang Lectures on "What's wrong with physicalism" at the National Taiwan University, June 2023, at a Workshop on "Physicalism, Consciousness, and Agency" at LMU Munich, June 2023, and at a workshop on "Consciousness and Indexicality" at the University of Valencia, November 2023. I thank the participants for helpful comments and suggestions, especially the six discussants at the series of Wendy Huang Lectures, Lok-Chi Chan, Tony Cheng, Ye Feng, Shao-Pu Kang, Plato Tse, and Wenjun Zhang.

[1] For an overview, see Stoljar (2021). On the debate about physicalism, see also Gillett and Loewer (2001).

[2] See Bourget and Chalmers (2023).

offer a sophisticated reduction of those properties to physical ones – Jaegwon Kim is such a philosopher – or try to explain how, despite not being reducible to physical properties, mental properties still supervene on, or are somehow grounded in, physical properties. The last position goes back to Donald Davidson, Jerry Fodor, and Hilary Putnam. The three camps correspond broadly to "eliminativist", "reductive", and "non-reductive" physicalist views.[3]

Non-reductive versions of physicalism, in particular, are popular. On the one hand, they retain physicalism's core idea that everything in the world, including the mind, supervenes on the physical – the "physicalist supervenience thesis" – and this is often considered a central tenet of any scientific worldview.[4] On the other hand, they recognize that there is something special about mental properties. Despite supervening on physical properties, mental properties are high-level properties that are descriptively or explanatorily irreducible to physical properties. The attractive idea, here, is that there can be "supervenience without reduction".[5]

I will grant, for present purposes, that the claim that the mental supervenes on the physical is defensible in relation to many of the mental properties that feature in our explanations of human behaviour in psychology and the social sciences, such as being in a certain intentional state, having certain beliefs and desires, preferring one action to another, and so on. However, when it comes to consciousness, I will argue, it is difficult to uphold the physicalist supervenience thesis.

Specifically, I will put the spotlight on an argument against physicalism which, despite being implicit in much of the philosophical debate about consciousness, including in some of Thomas Nagel's classic works, has not received the attention it deserves.[6] It asserts, roughly speaking, that conscious experience is first-personal, the physical is third-personal, and the first-personal does not supervene on the third-personal. I will call it the "first-personal argument" against physicalism. I will suggest that, for those who accept a realist view about conscious experience, this argument is hard to reject. It is distinct from the better-known "epistemic", "modal", and "conceivability" arguments against physicalism, the debate about which seems to have reached a bit of a stalemate. To make progress, it is therefore useful to bring in another argument. Furthermore, unlike the better-known arguments that non-physicalists often present, the first-personal argument, if successful, refutes not only physicalism but also other third-personal metaphysical pictures, including – perhaps surprisingly – standard versions of dualism.

The paper is structured as follows. In Section 2, I will define "physicalism" as I understand it here, and in Section 3, I will briefly introduce the challenge of making sense of consciousness. In Section 4, I will review the three better-known arguments against physicalism and indicate why, despite having advanced the debate, they remain somewhat inconclusive. In Section 5, I will discuss the "first-personal argument". I will conclude, in Section 6, by asking which non-physicalist views about the mind might avoid the challenge raised by that argument.

---

[3] See, respectively, P.M. Churchland (1981), P.S. Churchland (1986), Kim (2005), Davidson (2001), Fodor (1974), and Putnam (1975).
[4] See, e.g., Papineau (2001).
[5] For formal analyses of this idea, see, e.g., Butterfield (2011), Leuenberger (2018), Dewar (2019), and List (2019).
[6] See Nagel (1965, 1974, 1986).

2. Physicalism

I will take *physicalism*, in its core form, to be the thesis that all facts that hold in our actual world (even those that we do not conventionally regard as "physical" facts) supervene upon the physical facts.[7] *Supervenience*, in turn, is defined as follows: one set of facts (say, "the A-facts") *supervenes* on another (say, "the B-facts") if the second set of facts (the B-facts) necessitates the first (the A-facts), i.e., if it is metaphysically impossible for the A-facts to be any different without some of the B-facts being different too.[8]

David Lewis illustrates the idea of supervenience with the help of a simple example.[9] Consider an image printed on a piece of paper – say, the image of a triangle – that is made up of little dots. Perhaps the image was printed by an old dot-matrix printer. Clearly, the positioning of the dots fixes the image. It would be impossible for the image to be any different – for instance, to be an image of a circle rather than a triangle – without a difference in the underlying pattern of dots. The shapes displayed in the image supervene on the configuration of the dots.

Physicalism asserts that the relationship between the physical facts and all other facts, including mental ones, is like the relationship between the dots and the shapes. Once the physical facts are in place as they are in our physical universe, all other facts, including mental facts, are a necessary byproduct. This version of physicalism is also called "supervenience physicalism".

Supervenience physicalism could be defined equivalently (but more longwindedly) as the thesis that all facts that hold in our actual world *either* are physical facts themselves *or* supervene upon the physical facts. But since any fact that *is* physical trivially supervenes on a physical fact (namely itself), we can shorten the definition by dropping the first disjunct.

Supervenience physicalism is distinct from four other theses that are sometimes called "physicalist", none of which is logically entailed by it.[10] One is *strong physicalism*, the thesis that all facts *are* physical facts. This is more demanding than supervenience physicalism, insofar as supervenience physicalism permits (but of course doesn't require) the existence of facts that we call "non-physical", as long as these supervene upon – are necessitated by – physical facts. A second thesis from which supervenience physicalism is distinct is *grounding physicalism*, the thesis that all facts are *grounded in* physical facts. This is more demanding, insofar as grounding is a stronger relation of metaphysical dependence than supervenience.

---

[7] For an excellent overview of different definitions of physicalism, on which I draw, see Stoljar (2021). To avoid complications, the quantification over "all facts" can be understood to be restricted to all "positive" facts, i.e., the kinds of facts that the sciences (including the sciences of the mind) are concerned with for descriptive and explanatory purposes, and to exclude "normative" or "evaluative" facts, i.e., the facts studied in moral philosophy or value theory. For present purposes, a commitment to physicalism need not settle one's metaethical views. I do not want to presuppose, for example, that any physicalist must also be a moral naturalist.

[8] This notion of supervenience is also called "metaphysical supervenience", to distinguish it from "nomological supervenience" (Chalmers 1996), where necessitation is defined less demandingly as nomological rather than metaphysical necessitation ("it is nomologically impossible for the A-facts to be any different without some of the B-facts being different"). Whenever I refer to supervenience in this paper, I mean metaphysical supervenience.

[9] See Lewis (1986, p. 14).

[10] Again, I refer readers to Stoljar (2021) for an overview on which I draw.

Grounding, on standard accounts, implies supervenience (when A is grounded in B, then A also supervenes upon B), but supervenience does not imply grounding. Supervenience is a (mere) modal necessitation relation, while grounding carries further implications: when one set of facts is grounded in another, this means that the second set of facts metaphysically explains the first or is somehow more fundamental.[11] (There are different ways of spelling this out.) A third distinct thesis is *token physicalism*, the thesis that all particulars in the world are physical particulars. What this entails, and how it relates to supervenience physicalism, depends on what qualifies as "a particular". If the set of particulars includes all facts, then token physicalism entails supervenience physicalism and even strong physicalism, but it goes beyond those theses if there are also other entities among the particulars, distinct from facts. A supervenience physicalist could be agnostic about whether all *entities* should be characterized as "physical", as long as all *facts* about them supervene on physical facts. A final thesis from which supervenience physicalism is distinct is *type physicalism*, the thesis that all "types" (properties) are physical. The demandingness of this depends on our precise understanding of types, but type physicalism is conventionally regarded as more demanding than token physicalism.

These other physicalist theses differ not only in their demandingness but also in what kinds of physicalists they will appeal to. Strong physicalism and type physicalism are likely to appeal only to reductive physicalists (and to eliminativists), but not to non-reductive physicalists, as these strong versions of physicalism struggle to make sense of the above-mentioned notion of "supervenience without reduction". They do not leave room for facts or properties that are irreducible to physical facts or physical properties while still supervening upon physical facts or physical properties. Token physicalism, by contrast, may be more acceptable to non-reductive physicalists, provided it comes with a sufficiently restrictive account of what "the particulars" are, though some non-reductive physicalists may accept only supervenience physicalism. Grounding physicalism, finally, may appeal to those physicalists who seek to offer a grounding-based metaphysical theory, but as noted, it is more demanding than supervenience physicalism, and there is no consensus on how best to understand grounding.[12]

For the purposes of this paper, I define physicalism as supervenience physicalism because the physicalist supervenience thesis can be viewed as the "greatest common denominator" of all standard versions of physicalism. It is implied by practically all of them. Indeed, the supervenience thesis is something on which non-reductive, reductive, and even eliminative physicalists can agree. Moreover, for a critique of physicalism, it is best to use a basic and undemanding definition of physicalism. *If* one can argue against physicalism in a relatively undemanding form, *then* one will also have argued against any more demanding version of it.

Before moving on, I must note that, strictly speaking, my definition of supervenience physicalism is only a definition scheme. To render the physicalist supervenience thesis fully well-defined, we need to provide some criterion as to which facts or properties count as "physical". More restrictive or more permissive criteria give rise to more or less demanding conceptions of supervenience physicalism. One criterion, of which there are different versions,

[11] Note, for instance, that grounding is an irreflexive relation, while supervenience is a reflexive one.

[12] On the notion of grounding, see, e.g., Schaffer (2009) and Rosen (2010).

would say that a property is physical if and only if it is referred to by our (either current or future) best physical theories. Another criterion would say that a property is physical if and only if it is an intrinsic property of a paradigmatically physical entity.[13] Each of these criteria has advantages and disadvantages and raises some follow-up questions. Although I lean towards a version of the former criterion, it turns out that, for my analysis, I don't need to commit myself to any particular criterion of the "physical". I will later only need to rely on some relatively widely accepted judgments about what kinds of facts would *not* count as physical in any conventional sense.

3. Consciousness

What is consciousness?[14] A defining feature of any conscious subject, including us human beings, is that we not only function in a manner that can be described from an external, third-person perspective, in the same way in which we describe any ordinary physical process. But we also experience the world from a first-person perspective. There is something it is like to be a conscious subject, for that subject, as Nagel famously characterized it.[15]

The possession of subjective experiences distinguishes us from many, perhaps most, other entities in the world, from tables and refrigerators to rocks and flowers. These presumably lack a first-personal, subjective perspective on the world.[16] A complete third-personal description of those entities by some external observer exhausts everything that can be said about them.

David Chalmers describes the challenge of explaining consciousness as follows:

> "The task of a science of consciousness ... is to systematically integrate two key classes of data into a scientific framework: *third-person data*, or data about behavior and brain processes, and *first-person data*, or data about subjective experience."[17]

As Chalmers argues, these two kinds of data are fundamentally different from one another. The third-person data can be described in third-person language and studied using ordinary scientific methods, such as those of psychology, cognitive science, and neuroscience. Such data include data about an agent's wakefulness and sleep, their cognitive attention and computational capacities, their reasoning behaviour and its observable manifestations, and associated patterns of neural activity. We can formulate and empirically test various hypotheses about those phenomena in broadly the same way in which we study other phenomena in the sciences, from DNA to ecosystems and financial markets. Everything can be expressed in the

---

[13] On the present distinction between two such conceptions of the physical, the "theory-physical" and the "object-physical", see Stoljar (2001).

[14] This section draws closely on List (2023, Section 2), which, in turn, draws especially on Chalmers's (1996, 2004) framing of the issue.

[15] See Nagel (1974).

[16] Note that panpsychists think that consciousness is much more ubiquitous in the world than conventionally assumed, but this makes no difference to the arguments in this paper.

[17] See Chalmers (2004, p. 1111).

ordinary third-person language of science. Chalmers contrasts those relatively "easy" explanatory problems with the "hard" problem of making sense of the first-person data. Those data are about what a subject experiences from a first-person perspective: what it is like to be in a particular conscious state, as one undergoes the experience.[18]

As Nagel already wrote in his classic paper "What is it like to be a bat?",

> "[Subjective experience] is not captured by any of the familiar ... reductive analyses of the mental, for all of them are logically compatible with its absence."[19]

In line with Nagel's point, Chalmers notes that it is hard to explain why we have first-person experiences at all. In particular, it is hard to explain in physical terms (or more generally, in third-personal terms) why we aren't "zombies", hypothetical entities that are physically indistinguishable from us, display the same neural and behavioural responses to the world, and yet lack any first-person experiences. Any third-personal description of a zombie would look exactly like a third-personal description of you and me, and yet, by hypothesis, the zombie would have no first-personal standpoint, no inner stream of experiences, while you and I do. A focus on third-person data alone would not seem to capture the difference between a zombie and a human being like you and me. The challenge is to explain why, in our case, there are first-personal experiences that go along with various third-personally describable processes in our brains and bodies.[20]

Another way to characterize this challenge, suggested by David Levine, is to say that there is an "explanatory gap" between what a third-personal scientific explanation of the world can account for and what it would take to make sense of first-person experience.[21] When the *explanandum* – the thing to be explained – is first-personal, it is not clear that a purely third-personal *explanans* – the sort of explanation ordinary science might offer – could be satisfactory. I will return to this point when I discuss the "first-personal argument" below.

Still, physicalists will insist that the existence of such an explanatory gap need not entail a failure of supervenience. An explanatory gap is merely epistemic, i.e., a gap in our understanding of the world, while a failure of supervenience would be ontic, i.e., a gap in the world itself, for instance along the lines of René Descartes' suggestion that "the physical" and "the mental" are metaphysically distinct from one another. Physicalists suggest that, although we may not be able fully to *describe* and *explain* consciousness using physical concepts and categories alone, this does not imply that consciousness *fails to supervene* on physical properties. From a metaphysical perspective, consciousness may still be a necessary byproduct of underlying physical processes.

---

[18] See Chalmers (1996, 2004).

[19] See Nagel (1974, p. 436).

[20] See Chalmers (1996). Chalmers focuses primarily on the physical/phenomenal contrast in his exposition of the zombie scenario (i.e., zombies are physically identical to conscious humans but lack phenomenal experiences), but I find the framing in terms of the third-person/first-person contrast more congenial for present purposes.

[21] See Levine (1983). Note, however, that Levine himself frames his analysis of the explanatory gap in terms of the material/phenomenal contrast, not in terms of the third-person/first-person one.

4. Three classic kinds of arguments against physicalism

Non-physicalists in the analytic philosophy of mind have offered at least three kinds of arguments against the claim that the facts about subjective experience supervene on physical facts: "epistemic", "modal", and "conceivability" arguments. I will now briefly review them, to provide necessary context for the subsequent "first-personal" argument. In a debate replete with subtleties, it is useful to go over this material, even if it is familiar to many readers. Inevitably, I cannot do full justice to the rich philosophical debate on each of these classic arguments here.

4.1. Epistemic arguments

These arguments go like this:

> **Premise 1:** Some facts about subjective experience ("what it is like"-facts) are not knowable, even in principle, from knowledge of physical facts alone.
>
> **Premise 2:** Any fact that supervenes on physical facts is knowable, at least in principle, from knowledge of physical facts.
>
> **Conclusion:** Some facts about subjective experience do not supervene on physical facts. (So, physicalism is false.)

Note that Premise 1 and, by implication, the conclusion are cautiously worded, referring merely to "some facts" about subjective experience, rather than "most" or "all" such facts. But even with this cautious wording, the argument's conclusion suffices to imply that physicalism is false. For physicalism to be true, *all* facts about subjective experience must supervene on physical facts.

The most famous argument of the present kind is Frank Jackson's "knowledge argument".[22] Jackson asks us to imagine Mary, a brilliant neuroscientist who has spent her entire life locked in a black and white room. She has been interacting with the rest of the world via black-and-white computer interfaces and – ironically – has been studying the science of colour perception. She has acquired complete physical knowledge of this subject and, by hypothesis, knows everything there is to know, from a physical perspective, about how humans perceive colours. Jackson argues that, despite her extraordinary physical knowledge about colour perception, Mary does not know what it is like to see colour. In support of this claim, he cites the intuition that when Mary finally leaves the black-and-white room and sees colour for the first time – say, the bright red of a ripe tomato – she learns something new. This suggests that the fact which she learns – the "what is it like to see red"-fact – is not physical and does not supervene on physical facts. Otherwise, she should have known, or been able to infer, this fact all along. Many people share the intuition that there is a gap in Mary's knowledge even if she is omniscient with respect to the physical facts, and the argument then suggests that that there could not be such an epistemic gap unless there is also an ontic gap. So, physicalism is false.

---

[22] See Jackson (1982). For a helpful overview and critical discussion, see Nida-Rümelin and O Conaill (2019). Jackson's original formulation of the argument targets what I have here called "strong physicalism" rather than supervenience physicalism, so the present formulation can be seen as an amended version of Jackson's argument.

Although the epistemic argument, as I have stated it, is formally valid (its premises entail its conclusion) and the intuition on which it rests is powerful, the argument is vulnerable to several well-known objections.[23] In relation to Premise 1, one could deny that we have reliable intuitions about what we would be able to know if – like Mary in Jackson's thought experiment – we had *complete* physical knowledge about the world (or about some significant part of the world). This epistemic predicament would be so dramatically superior to our normal one that we cannot reliably predict what we would or would not know in such a situation.

Further, in relation to Premise 1, one could argue that, even if it seems plausible that Mary learns something new when she sees colour for the first time, she does not really learn any new fact. One possibility is that she merely learns a new mode of presentation of an old fact that she had known all along. She had previously known that fact – what it is like to see red – under a physical or neuroscientific mode of presentation, while she now comes to know it under a phenomenological one. In effect, she is now able to describe that fact using different concepts and categories. This does not, however, alter the fact itself. Alternatively, what Mary learns might be just a new skill or a new instance of "knowledge how", such as the know-how of recognizing a particular colour, not a new instance of factual knowledge ("knowledge that").[24]

Although more could be said about these objections to Premise 1, let me now turn to objections to Premise 2, the claim that any fact that supervenes on physical facts must be knowable, at least in principle, from knowledge of physical facts. Here, the main strategy is to argue that *supervenience does not imply learnability*, i.e., that even if A supervenes on B, one need not be able to derive knowledge of A from knowledge of B. There are at least two possible reasons for this. One is complexity. Supervenience relations, such as between certain low-level physical facts and certain high-level psychological facts, may be so complex that it is not humanly feasible – given reasonable computational constraints – to "read off" the latter from the former, even though the latter are necessitated by the former. This would then undercut Premise 2.

Another reason as to why supervenience need not imply learnability is inspired by Saul Kripke's work.[25] Kripke famously argued (in the different philosophical context of his book *Naming and Necessity*) that there are truths which, despite being metaphysically necessary, are not knowable *a priori*. One example of such an *a posteriori* necessity is "water is H_2O". Its truth is metaphysically necessary, given the necessity of the identity relation, and yet there is no way one could know this *a priori*, without empirical information. If – and this is of course a big "if" – the dependence of subjective experience on physical facts were an instance of a necessary truth that is knowable only *a posteriori*, this would imply that supervenience does not generally imply learnability and thereby again undercut Premise 2.[26]

---

[23] For excellent overviews, I refer readers again to Nida-Rümelin and O Conaill (2019) and Ludlow, Nagasawa, and Stoljar (2004). My short summary of the objections to the epistemic argument can be viewed just as restatements of earlier ideas from the work presented or reviewed therein.

[24] See Nemirow (1980) and Lewis (1983).

[25] See Kripke (1980).

[26] It appears, however, that Kripke himself did not endorse this kind of objection to the epistemic argument. See Renero's (2023) discussion of Kripke's unpublished 1979 "Lectures on the Philosophy of Mind". Indeed, the

To be sure, the proponents of an epistemic argument against physicalism have responses to some of these objections, which, in turn, invite physicalist counter-responses, and so on. I am not claiming that the knowledge argument is unsound. All I am suggesting is that the issue is far from settled, and the debate remains somewhat inconclusive at this point.

4.2. Modal arguments

These are roughly as follows:[27]

> **Premise 1:** It is metaphysically possible for some facts about subjective experience to be different (or even to be completely absent) without any difference in the physical facts.
>
> **Premise 2:** If it is metaphysically possible for some facts about subjective experience to be different without any difference in the physical facts, then the former facts do not supervene on the latter.
>
> **Conclusion:** Some facts about subjective experience do not supervene on physical facts. (So, physicalism is false.)

The modal intuition underlying Premise 1 is widely held among non-physicalists. (Again, note the cautious wording "some facts".) Scholars from Descartes to Nagel and Chalmers endorsed versions of the claim that a disconnect between physical facts and facts about subjective experience is possible from a logical or metaphysical perspective. Recall, for instance, Nagel's above-quoted remark that "[subjective experience] is not captured by any of the familiar ... reductive analyses of the mental, *for all of them are logically compatible with its absence*" (emphasis added).[28] Furthermore, Premise 2 follows immediately from the definition of supervenience and should thus be uncontentious. Finally, the argument is formally valid.

The biggest problem with the argument, however, is *not* that it may be unsound (indeed, its first premise could well be true, and its second premise is certainly true), but rather that its key premise – Premise 1 – is not really any less demanding than its conclusion. Given the definition of supervenience, which Premise 2 in effect just reasserts, Premise 1 is logically equivalent to – in fact, just a restatement of – the argument's conclusion: to say that some facts about subjective experience do not supervene on physical facts (as the conclusion asserts) is just to say that it is metaphysically possible for some facts about subjective experience to be different without any difference in the physical facts (as Premise 1 asserts).

This means that no-one who doesn't already agree with the conclusion is likely to agree with Premise 1. The argument is therefore dialectically ineffective. For an argument to be dialectically effective, one might say, its premises should at least individually be easier to

---

typical Kripkean examples of *a posteriori* necessary truths ("water is H_2O", "Hesperus is Phosphorus") seem quite different from the necessary truths that would be entailed by the physicalist supervenience thesis.

[27] For discussion of such arguments, see, e.g., Chalmers (1996).

[28] See Nagel (1974, p. 436).

accept than the conclusion, and it should only be the conjunction of these individually more acceptable premises that entails the conclusion.

Perhaps in recognition of this point, those who think that a modal argument is on the right track tend to offer further reasons for accepting Premise 1. Indeed, the third kind of argument to be discussed can be viewed as an amended version of a modal argument – in effect, offering a "conceivability test" in support of Premise 1 of the modal argument.

4.3. Conceivability arguments

These can be summarized as follows:

> **Premise 1:** It is conceivable that there could be zombies, i.e., entities which are physically indistinguishable from conscious human beings but lack any subjective experiences.
>
> **Premise 2:** Conceivability implies metaphysical possibility.
>
> **Premise 3:** If zombies are metaphysically possible, then some facts about subjective experience (as in the human case) do not supervene on physical facts.
>
> **Conclusion:** Some facts about subjective experience do not supervene on physical facts. (So, physicalism is false.)

David Chalmers has prominently defended this kind of argument.[29] It clearly avoids the main problem of the above-stated modal argument, insofar as none of its premises individually already presupposes the conclusion. Furthermore, the premises each have some plausibility. If we can coherently come up with a philosophical thought experiment in which there are zombies, as Chalmers suggests, this should support Premise 1. Premise 2 also seems to have some appeal, as conceivability might be thought to be a good indicator of metaphysical possibility. And Premise 3 follows from the definition of supervenience, together with the definition of zombies. (Once more, note the cautious wording "some facts", which again suffices for the conclusion that physicalism is false.) This argument, too, is formally valid. And so, on the face of it, it gives us reason to reject physicalism.

However, the argument has some widely recognized weaknesses. Firstly, some physicalists will not find zombies conceivable and will therefore challenge Premise 1. Daniel Dennett is an example of a philosopher who finds the notion incoherent.[30] He writes:

> "[T]his conceivability is *only* apparent; some misguided philosophers *think* they can conceive of a zombie, but they are badly mistaken."[31]

---

[29] Again, see Chalmers (1996).
[30] See, e.g., Dennett (2005).
[31] Ibid., p. 15.

Secondly and perhaps even more importantly, even if we grant the conceivability of zombies, the claim that anything that is conceivable is also metaphysically possible, which is Premise 2, is controversial.[32] If conceivability is some kind of epistemic or doxastic modality while metaphysical possibility is an alethic modality, it is not clear why the former should imply the latter. Conceivability may be more permissive than metaphysical possibility. That is, more things could be conceivable than are metaphysically possible.

For example, if there are necessary truths that are only *a posteriori*, such as "water is H_2O", as already mentioned, there could be some metaphysically necessary truths whose falsehood is nonetheless conceivable. Also, conceivability could be understood as the *lack of transparent metaphysical impossibility*. So, anything that isn't transparently metaphysically impossible – i.e., it is either metaphysically possible or metaphysically impossible but not in a way that is easy to recognize – could count as conceivable. The falsity of Goldbach's conjecture may be an example. This is the statement that every even number above 2 (4, 6, 8, 10 etc.) is the sum of two prime numbers (e.g., 4 = 2+2, 6 = 3+3, 8 = 5+3, 10 = 5+5 etc.). While many people suspect that it is true (and necessarily true, if it is a theorem of arithmetic), its truth has not yet been proven, and so we do not – strictly speaking – know that its negation is false; it is a well-known open problem in mathematics. We may therefore find the falsity of Goldbach's conjecture conceivable, even though it may turn out to be metaphysically impossible, relative to the standard axioms of arithmetic.

These considerations suggest that, despite their initial plausibility, Premises 1 and 2 are somewhat debatable. Only Premise 3 seems unproblematic. As in the case of the knowledge argument, then, the soundness of the present kind of conceivability argument has not yet been established.

My short discussion of the epistemic, modal, and conceivability arguments has left out much detail, and I have not covered all the strategies their proponents and opponents could use to defend their views. It is beyond question is that the arguments are formally valid – in each case, the conclusion follows from the premises – and that they have each advanced the debate about consciousness. Nonetheless, the status of their premises is far from settled, and the debate has reached a bit of an "impasse". It is therefore useful to consider a distinct argument.

5. A fourth kind of argument

Interestingly, the three classic kinds of arguments I have reviewed do not put much emphasis on the first-person nature of subjective experience. The epistemic arguments emphasize the idea that knowledge of the physical facts alone does not guarantee knowledge of the facts about subjective experience. The modal arguments rest on the idea that the facts about subjective experience are modally distinct from the physical facts. And the conceivability arguments emphasize the idea that hypothetical entities that differ from actual human beings in experiential respects but not in physical ones seem conceivable. The first-person nature of

[32] See, e.g., Gendler and Hawthorne (2002) and Hill (2016).

subjective experience, however, is not at the core of any of those arguments, at least not explicitly, even if it may implicitly underpin the case for some of their premises.

I want to suggest that there is a fourth kind of argument against physicalism that appeals directly to the first-person nature of subjective experience. The argument is implicit in much of the debate on consciousness, yet it does not seem to have received the attention it deserves.[33] Nagel articulates the basic idea when he writes:

> "If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view."[34]

And in an even earlier paper, he writes:

> "The feeling that physicalism leaves out of account the essential subjectivity of psychological states is the feeling that nowhere in the description of the state of a human body could there be room for a physical equivalent of the fact that *I* (or any self), and not just that body, am the subject of those states."[35]

My aim is to offer a simple and straightforward formulation of what I take to be the "first-personal argument" against physicalism and to put the spotlight on it. It goes like this:

> **Premise 1:** Some facts about subjective experience are first-personal and indexical.
>
> **Premise 2:** Physical facts (as conventionally understood) are third-personal and non-indexical.
>
> **Premise 3:** First-personal and indexical facts do not supervene on third-personal and non-indexical facts.
>
> **Conclusion:** Some facts about subjective experience do not supervene on physical facts (as conventionally understood). (So, physicalism is false.)

---

[33] The insight that the irreducibility of consciousness to physical properties stems at least in part from its first-person nature is implicit in many of the modern classics in the analytic philosophy of consciousness, from Nagel (1965, 1974, 1986) to Chalmers (1996). It can also be found in the less well-known work of Geoffrey Madell (2003), who argues that "[m]aterialism … is undermined by its failure to understand the first-person perspective in a number of ways" (p. 139, typo corrected), and Lynne Rudder Baker's work (2007, 1998), who argues that "[t]he first-person perspective is a challenge to naturalism", because "[n]aturalistic theories are relentlessly third-personal" (2007, p. 203). The insight is central to phenomenological approaches too. For a review, see Smith (2018). Another earlier work that tackles the relationship between the impersonal and uncentered perspective of the physical sciences and the first-personal and "situated" perspective of the conscious subject is Ismael (2007).

[34] See Nagel (1974, p. 437).

[35] See Nagel (1965, p. 354). The final section of that paper contains a precursor of the present argument.

Again, a cautious wording (referring to "some facts" rather than "most" or "all") suffices for the conclusion that physicalism is false. We can formulate subtly different variants of this argument, depending on whether we put the emphasis on the first-person nature of consciousness or on its indexicality (or both). In principle, we could replace every occurrence of "first-personal and indexical" with one of these characteristics alone and correspondingly replace every occurrence of "third-personal and non-indexical" with one characteristic alone. The effect of such a modification would depend on the precise relationship between first-personal and indexical facts. As I see it, every first-personal fact is indexical, while not every indexical fact is first-personal. For instance, the fact that I am experiencing the bright light of a beautiful sunny day today is both first-personal and indexical, while the fact that it is currently the year 2024 is indexical but not first-personal. Premise 1, therefore, becomes subtly logically weaker when the claim is merely that the relevant facts are indexical and no mention is made of their first-person nature.

Regardless of whether we state the argument in terms of the first-person/third-person contrast or in terms of the indexical/non-indexical contrast, the argument is formally valid. So, if we accept the premises, we must accept the conclusion. I now want to explain why I find it hard to challenge the argument's premises, even though I will indicate which pressure points physicalists could target if they sought to rebut the argument.

Note already that the first-personal argument is immune to some of the key objections that physicalists have raised in response to the classic arguments against physicalism. For instance, if physicalists can show that supervenience does not imply learnability, or that conceivability does not imply metaphysical possibility, they will have undermined the epistemic or conceivability arguments, including the "conceivability test" that is often used to back up the modal argument. By contrast, the first-personal argument remains unaffected by these moves.

Irrespective of whether readers will ultimately be convinced of the argument's soundness, my hope is that the discussion will shed further light on some important dividing lines between physicalists and non-physicalists and thereby help to advance the debate.

5.1. The first premise

Premise 1 asserts that some facts about subjective experience are first-personal and indexical. This can be viewed as a conjunction of two claims:

- There are facts about subjective experience.
- At least some of them are first-personal and indexical.

Let me explain why I find it hard to deny these claims. Firstly, it seems a firm data point of our own consciousness *that there are facts about subjective experience*. For instance, it is a fact that I currently have some subjective experiences. For instance, I see my computer in front of me and hear some outside noise. Up to this point, we are merely asserting that there are some such facts, so far leaving it open whether these have the property of *being first-personal*. The existence of some facts about subjective experience is perhaps the one insight in Descartes'

famous "Cogito, ergo sum" reasoning that is intuitively least contestable. Anyone who accepts realism about consciousness should agree that there are such "experiential facts", and anti-realism or illusionism about consciousness are not the easiest views to defend, though arguing against them is not my topic here. I am happy to concede that my argument against physicalism rests on the assumption that subjective experience is a real phenomenon.

Secondly, it also seems evident *that (at least) some of the facts about subjective experience are first-personal*, though I will discuss an important objection below. To begin with, the first-person nature of subjective experience is one of its defining features. As Nagel notes,

> "I have a type of internality which physical things lack; so in addition to the connection which all my mental states do admittedly have with my body, they are also mine – that is, they have a particular *self* as subject, rather than merely being attributes of an object."[36]

Indeed, a central explanatory challenge for a science of consciousness, as pointed out by Chalmers in the quote in Section 3, is the accommodation of first-person data, not just third-person data. Some of the *explananda* we are trying to explain are first-personal facts. And of course, these are indexical, insofar as they are not invariant under changes in their subject.

Conscious experiences are "subject-centered", in the sense that each of us experiences the world in a way that is experientially centered around a different subjective perspective.[37] I experience the world in a way that is experientially centered around my subjective perspective. You experience the world in a way that is experientially centered around your subjective perspective. Importantly, you need not think that there is such an *entity* or *substance* as "the self" to accept the claim that our experiences are perspectival and centered around a subjective perspective. The notion of "the self" *as an entity or substance* is controversial and, as many have argued (including Nagel), implausible.[38] What I mean by "subject-centered" is merely that conscious experiences involve a subjective perspective.

---

[36] See Nagel (1965, p. 353).

[37] Again, see Nagel (1965).

[38] Kit Fine (2005, p. 312) helpfully distinguishes between two notions of "the self": the "metaphysical self" and the "empirical self". He understands the metaphysical self as "the implicit subject of the egocentric facts", writing "it might be regarded as the locus of subjectivity, since it is relative to such a self that the egocentric facts will obtain". And he understands the empirical self as "the explicit subject of non-egocentric facts". It is the empirical self as an entity which many neuroscientists and philosophers find a dubious notion. The metaphysical self is not an entity in the world but the "locus of subjectivity" around which facts of subjective experience are centered. When I speak of the "subject-centeredness" of conscious experiences I am invoking only a metaphysical self in Fine's sense, a locus of subjectivity or a subjective perspective, not an empirical self as an entity. Similarly, Nagel (1965, p. 355) writes: "the quest for the self, for a substance which *is* me and whose possession of a psychological attribute will *be* its being mine, is a quest for something which could not exist. … [T]he self is not a substance, and … the special kind of possession which characterizes the relation between me and my psychological states cannot be represented as the possession of certain attributes by a subject, no matter what that subject may be. The subjectivity of the true psychological subject is of a different kind from that of the mere subject of attributes."

While this subject-centeredness or perspectival nature of conscious experiences should be uncontroversial, it is especially scholars in the phenomenological tradition who have emphasized it. Dan Zahavi, for instance, describes the first-person nature of subjective experience as follows:

> "[S]ubjectivity is a built-in feature of experiential life. Experiential episodes are neither unconscious, nor anonymous, rather they *necessarily come with first-personal givenness or perspectival ownership*. The what-it-is likeness of experience is essentially a what-it-is-like-for-me-ness."[39]

In other words, "what it is like"-facts are essentially "what it is like for me"-facts and thus first-personal.

Still, an important objection to Premise 1 is that facts themselves are never first-personal, but that the first-person/third-person distinction is only an epistemic or linguistic phenomenon. According to this objection, the first-person/third-person contrast corresponds to different *modes of presentation* with which we sometimes represent certain ordinary, entirely objective facts. When I say, for instance, "I am seeing a bright red tomato in front of me", the content of this first-person sentence is equivalent to that of the third-person sentence "Christian is seeing a bright red tomato in front of him". Both sentences express the same objective fact, and they merely do so with a different mode of presentation. No first-person/third-person distinction can be drawn at the level of facts. The distinction is only one at the level of how we represent certain facts, and thus it is epistemic or linguistic, not ontic.

This way of thinking is backed up by an influential approach to the semantics of indexicals, according to which indexical sentences such as "I am in Munich right now" still have ordinary, non-indexical propositions as their content, once the context of utterance is given.[40] The sentence "I am in Munich right now", uttered by me at this time, expresses the proposition that Christian is in Munich at this particular time. Language, this view tells us, is more fine-grained than reality itself, and there are linguistic distinctions, such as that between first-person and third-person expressions, that do not mirror any distinctions in reality.

However, it may be argued that this way of rejecting Premise 1 – i.e., denying that any facts (as opposed to our mere representations of them) could ever be first-personal or indexical – does not seem to work even for the weaker version of Premise 1 framed in terms of indexicality alone, let alone the stronger version that refers to the first-person nature of subjective experience in a richer sense beyond indexicality. Here, for instance, is David Chalmers noting that there are indexical facts as distinct from ordinary objective facts:

> "[E]ven if the indexical is not an objective fact about the world, it is a fact about the world as I find it, and it is the world as I find it that needs explanation."[41]

---

[39] See Zahavi (2017, p. 194, emphasis added). I thank Robert Prentner for drawing my attention to this quote.
[40] See especially Kaplan (1989). For an overview, see, e.g., Braun (2017).
[41] See, e.g., Chalmers (1996, p. 85).

He concedes that the idea of a "brute indexical" may seem "quite obscure" and hard to explain, but he stresses that indexical facts shouldn't be conflated with certain corresponding non-indexical facts, notwithstanding the conventional linguistic strategy of taking the meaning of the indexical sentence "I am in Munich right now", uttered by speaker *S* at time *t*, to be simply the ordinary, non-indexical proposition that *S* is in Munich at time *t*. Chalmers writes:

> "Of course, we can give a reductive explanation of why David Chalmers's utterance of 'I am David Chalmers' is true. But this nonindexical fact seems quite different from the indexical fact that I am David Chalmers."[42]

Similarly, Nagel observes that "the fact that I am Thomas Nagel", understood as an indexical fact, is distinctive:

> "This is not, of course, the fact ordinarily conveyed by those words, when they are used to inform someone else who the *speaker* is – for that could easily be expressed otherwise. It is rather the fact that I am the subject of these experiences; this body is my body; the subject or center of my world is this person, Thomas Nagel."[43]

Nagel argues that this indexical fact is not entailed by the facts that can be described in non-indexical terms ("without token-reflexives").[44] Similarly, Chalmers suggests that some such indexical facts may simply need to be recognized in our picture of reality:

> "The indexical fact may have to be taken as primitive. If so, then we have a failure of reductive explanation distinct from and analogous to the failure with consciousness."[45]

But Chalmers also thinks that the full-blown facts about subjective experience go beyond indexical facts, and he is thus on board with the view that the first-person nature of conscious experience goes beyond its mere indexicality. While he takes the failures of reductive explanation to be structurally analogous in the two cases, he says:

> "Still, the failure [of reductive explanation in the case of indexicality] is less worrying than that with consciousness, as the unexplained fact is so 'thin' by comparison to the facts about consciousness in all its glory. Admitting this primitive indexical fact would require far less revision of our materialist worldview than would admitting irreducible facts about conscious experience."[46]

Nonetheless, note that, for Chalmers, even admitting only "thin" indexical facts would already require a revision of our materialist worldview. Chalmers is of course a realist about consciousness and – consistently with what I am arguing here – willing to admit irreducible

---

[42] Ibid.
[43] See Nagel (1965, p. 355).
[44] Ibid., final section.
[45] See again Chalmers (1996, p. 85).
[46] Ibid.

facts about conscious experiences that go beyond indexical facts (even though Chalmers's view about the relationship between indexicality and consciousness differs somewhat from mine).[47]

In more recent work, Giovanni Merlo argues for realism about first-personal facts, which he calls "subjective facts", and which are presumably richer than mere indexical facts. Merlo notes that our ontological inventory of the world would be incomplete if we didn't recognize certain subjective facts:

> "if I were to write a book entitled 'The World As I Found It' or 'The World As It Really Is', Giovanni would have a role in that book that no other individual has. He would be (I blush to say) the *main character* of that book, the only and authentic *centre* of the world. That, of all individuals there are, Giovanni is the one having this role strikes me *as an undeniable and all-too-important fact*. To me, writing the book of the world without mentioning the fact that Giovanni is special would be writing an incomplete book."[48]

Clearly, this fact is a subjective or first-personal one. It doesn't hold from all perspectives, but still, from where Merlo stands, it holds simpliciter.[49]

The bottom line is that if we were to deny that there are first-personal facts, we would be committed to an ontological view that fails to do justice to "the world as we find it", and I am inclined to reiterate Chalmers's point that "it is the world as I find it that needs explanation".[50] For this reason, I accept the first premise of the first-personal argument.

5.2. The second premise

The argument's second premise is that physical facts, at least as conventionally understood, are third-personal and non-indexical. This claim should be fairly straightforward too. Although I have refrained from committing myself to any particular account of "the physical" here, it seems hard to deny that, on conventional understandings of "the physical", physical facts are third-personal and non-indexical. Such facts are not in any way subject-centered or perspectival.

Indeed, the de-subjectivization of our worldview (i.e., abstracting away from any subjective perspective) and a striving for greater objectivity have been key features of the history of science at least since the Enlightenment, including the history of physics. Think of the move from an anthropocentric and a geocentric worldview to a heliocentric one and later to a worldview that denies the existence of any privileged reference frame and aims to approximate

---

[47] For me, first-person facts, which hold when I am conscious, are particularly rich instances (perhaps special limiting cases) of indexical facts. To give a rough gloss, I am conscious, on my account, if and only if, from where I stand, some first-personal facts hold *simpliciter*. Crucially, this is *not* an account of what epistemic state I must be in to qualify as conscious (e.g., what I would need to believe or know to be conscious). Rather, it is an ontological account of what facts must hold for me to be conscious (namely, first-personal ones).

[48] See Merlo (2016, p. 319, emphasis added on "as an undeniable and all-too-important fact").

[49] For another recent discussion of "subjective facts about consciousness" (published after I completed the first version of this paper), see Lipman (2023).

[50] See once more Chalmers (1996, p. 85).

what Nagel has famously called "the view from nowhere".[51] If this is, or has been, the ambition of science, then it seems that, on a scientific understanding of "the physical", physical facts should be third-personal and non-indexical.

Furthermore, the idea that physical facts should be third-personal and non-indexical, and thereby invariant under shifts in any observer's perspective, is supported by an influential philosophical theory of scientific objectivity, namely Robert Nozick's theory of objectivity as invariance, according to which objective facts, which the sciences seek to identify, are those that are invariant under admissible shifts in reference frame.[52] On this picture, the quest for objectivity, including in the physical sciences, can be seen as the quest for an increasingly aperspectival and impersonal worldview.

Of course, fundamental physics is still in flux, and one may be able to find some revisionary approaches that take seriously the idea that there could be irreducible indexical or perspectival facts. Such facts might include tensed facts about what is past, present, and future, as opposed to tenseless facts about whether X precedes Y, X and Y are simultaneous, or X succeeds Y.[53] But tensed facts are recognized at most by some "niche" approaches in physics that are committed to an "A-theory" of time, i.e., a theory that takes tense to be real, not just temporal relations such as "before-after". Most standard physical accounts of time are firmly "B-theoretic", i.e., they offer a block-universe picture of the world which can be fully described in a tenseless way. According to such theories, time is simply one of several dimensions, along with the three spatial dimensions, and the theories do not postulate any indexical facts.[54] As Jenann Ismael describes the standard view, "[t]he world of physics is fixed and eternal", in sharp contrast to the "world of experience", which is "transient and changing".[55]

Another class of physical theories that potentially postulate some indexical or perspectival facts are some radically subjective or epistemic interpretations of quantum mechanics that recognize certain observer-dependent facts. QBism, whose etymological origin is "quantum Bayesianism", may be an example of a physical theory that puts the subject or the observer at the center and that might accept the notion of observer-dependent facts.[56] Similarly, relational quantum mechanics, as proposed by Carlo Rovelli, postulates an ontology in which observer-dependent facts are central. The idea is that certain facts about a quantum system's state hold only at system-observer pairs, not at systems simpliciter (with the special proviso that

[51] See Nagel (1986).
[52] See Nozick (2001).
[53] Note that, in relativistic physics, temporal relations are relative some reference frame, but facts about precedence, simultaneity, and succession relative to a particular frame are still tenseless.
[54] On time in physics, see, e.g., Callender (2017). For a recent argument to the effect that, contrary to conventional wisdom, our best physical theories do in fact support a form of presentism (according to which there are tensed facts) over (tenseless) eternalism, see Builes and Impagnatiello (forthcoming). Their thesis is that if the universe satisfies a certain kind of Markov property, this is better accounted for by presentism than by eternalism.
[55] See Ismael (2007, abstract).
[56] See, e.g., Fuchs (2010) and Mermin (2018).

"observers" are also taken to be physical systems).[57] Finally, in some many-worlds interpretations of quantum mechanics, the fact about which of several different branches of a multiverse we find ourselves in might be an indexical fact.[58] But again, such interpretations of quantum mechanics are hardly the mainstream that is presented in physics textbooks.

Furthermore, while one might find some revisionary approaches to physics that postulate indexical facts, as illustrated by the examples just given, approaches that recognize genuinely first-personal facts are even rarer. Again, I can only think of QBism as a possible example. Outside that special realm, I cannot think of genuinely first-personal facts in physics at all (though I would welcome learning about other examples). All this leads me to conclude that, at least on a conventional understanding of physics, physical facts are third-personal and non-indexical, as asserted by the second premise of the argument.

That said, some philosophers have proposed a broadened version of physicalism that recognizes certain "subjective" facts. This is the so-called "subjective physicalist view" defended by Tim Crane and Robert Howell.[59] According to it, there are two kinds of physical facts: "book-learning facts" and "non-book-learning facts". *Book-learning facts* are

> "facts the learning of which [does] not require you to have a certain kind of experience or occupy a certain position in the world".[60]

By contrast, *non-book-learning facts* are not like this. Learning them does "require you to have a certain kind of experience or occupy a certain position in the world".[61] The intuitive idea is that while book-learning facts can be communicated via a physics textbook, by conveying ordinary propositional information, non-book-learning facts cannot be learnt like this. You can come to know any such fact only if you undergo a certain experience or are appropriately positioned in the world.

The subjective physicalist view can thus categorize some of the facts about subjective experience, such as what it is like to see red, as non-book-learning facts, while insisting that they are still physical. This would accommodate Frank Jackson's original intuition that even after thoroughly studying the science of colour perception and learning all book-learning facts about this topic, one would still not know what it is like to perceive colour. And yet, the subjective physicalist can insist that this doesn't undermine the claim that those "what it is like"-facts are physical; rather, they just fall outside the "book-learning" category.

---

[57] See Rovelli (1996). Rovelli writes: "[Relational quantum mechanics] distinguishes relative facts from stable facts... Relative facts (or 'events') form the basis of the ontology; they are ubiquitous and do not require any special property of the physical systems involved in order to happen. Stable facts are facts stabilised by decoherence, in the sense that their relativity can be ignored by a large class of systems…" (Di Biagio and Rovelli 2022, p. 3). The relative facts can be viewed as observer-dependent, the stable ones as observer-independent.

[58] See, e.g., Wallace (2014).

[59] See Crane (2003) and Howell (2013).

[60] See Crane (2003, p. 78).

[61] Ibid.

Perhaps subjective physicalism, then, does give us an account of "the physical" under which physical facts need not be third-personal and non-indexical. If this were so, then subjective physicalism would have the right structure to respond to the first-personal argument. Indeed, Howell acknowledges that

> "[i]t might be thought that subjective physicalism bears some similarity to views that assimilate conscious knowledge to indexical knowledge."[62]

But he quickly adds:

> "Though subjective physicalism might seem to have a great deal in common with the indexical-knowledge views of consciousness, in the end the similarities are superficial. Indexical knowledge is no doubt closely linked with some sense of 'subjectivity,' but it isn't the sense [intended by subjective physicalism]."[63]

> "Subjective physicalism is a version of the view Chalmers dubs 'Type-B materialism,' that Frank Jackson calls 'a posteriori physicalism,' and I call epistemicism. These views hold that while the facts about conscious experience are necessitated by the physical facts, they cannot be inferred a priori from those facts."[64]

This clarification suggests that subjective physicalism, at least as understood by Crane and Howell, doesn't really abandon the third-personal and non-indexical picture of "the physical". It merely abandons the idea that all facts are book-learning facts or learnable from knowledge of book-learning facts alone. However, an even more important point for present purposes is that regardless of whether the postulated non-book-learning facts include some first-personal or indexical ones, subjective physicalism clearly goes beyond the conventional physicalist ontology. By accepting subjective physicalism, one will have conceded at least a core part of the philosophical case against physicalism in its conventional form.

5.3. The third premise

The third premise of the first-personal argument, which says that first-personal and indexical facts do not supervene on third-personal and non-indexical facts, should be the least controversial (at least under the assumption that there are such facts).

First-personal or indexical facts are clearly underdetermined by third-personal and non-indexical facts: the totality of third-personal and non-indexical facts is insufficient to fix the first-personal and indexical ones. David Lewis gives the imaginary example of two Gods who are omniscient with respect to all third-personal and non-indexical facts about the world. One of the two Gods, we may suppose, lives on mountain A, the other lives on mountain B. As far as ordinary propositional knowledge is concerned, they are not lacking anything. Yet, for each

---

[62] See Howell (2013, p. 160).
[63] Ibid.
[64] See Howell (2013, p. 161, comma added after "materialism").

of them, there is an indexical fact that he is unaware of: "neither one knows which of the two he is".[65] This indexical fact is left open (underdetermined) by the totality of non-indexical facts. Indeed, the total body of non-indexical facts is the same, irrespective of whether one occupies the perspective from mountain A or the perspective from mountain B. Non-indexical facts are completely non-perspectival. Indexical facts such as "I am the one on mountain A" therefore do not supervene even on the totality of non-indexical facts.

Nagel makes a similar point:

> "[C]onsider everything that can be said about the world without employing any token-reflexive [i.e., indexical] expressions. This will include the description of all its physical contents and their states, activities, and attributes. It will also include a description of all the persons in the world and their histories, memories, thoughts, sensations, perceptions, intentions, and so forth. I can thus describe without token-reflexives the entire world and everything that is happening in it – and this will include a description of Thomas Nagel and what he is thinking and feeling. But there seems to remain one thing which I cannot say in this fashion – namely, which of the various persons in the world *I* am. Even when everything that can be said in the specified manner has been said, and the world has in a sense been completely described, there seems to remain one fact which has not been expressed, and that is the fact that I am Thomas Nagel."[66]

Recall that he characterizes this fact not just as the fact that the speaker is Thomas Nagel, but as "the fact that I am the subject of these experiences; this body is my body; the subject or center of my world is this person, Thomas Nagel".[67]

These considerations indicate that the totality of third-personal and non-indexical facts about the world I inhabit leaves open the first-personal and indexical facts that hold from where I stand. These differ from the first-personal and indexical facts that hold from where you stand, even though the third-personal and non-indexical facts are the same. This underdetermination should be evident because the third-personal facts do not settle who I am or what perspective on the world I occupy. First-personal facts are more fine-grained than third-personal ones.

One way to formalize this point, which some readers may find helpful (though others may not find the suggested formalization congenial), is to note the following:

(i) A third-personal and non-indexical fact can be represented without information loss by an *uncentered proposition*, formally a set of possible worlds, namely, the set of those possible worlds at which the fact holds.

By contrast:

---

[65] See Lewis (1979, p. 520).
[66] See Nagel (1965, pp. 354–355).
[67] Ibid.

(ii) A first-personal or indexical fact cannot be represented without information loss by an uncentered proposition but would need to be represented by a *centered proposition*, formally a set of suitably interpreted centered worlds (pairs consisting of a world and a center in it), namely, the set of those centered worlds in which this first-personal or indexical fact holds.[68]

For instance, if we wish to represent the fact that I am Christian in a first-personal and indexical manner, and not just to reduce it to the trivial third-personal and non-indexical fact that Christian is Christian (which would clearly entail an information loss), we must represent it by a set of centered worlds that are centered around Christian as a subject. Centered worlds, which are formally world-center pairs, are more fine-grained than uncentered worlds. They can:

- coincide with respect to the uncentered-world component and thus with respect to all ordinary, non-indexical and third-personal facts that hold at them and yet
- differ with respect to the center and thus with respect to some of the indexical or first-personal facts.

For this reason, I consider the premise that first-personal and indexical facts do not supervene on third-personal and non-indexical facts to be unproblematic, especially once we have conceded that there are first-personal and indexical facts as distinct from third-personal and non-indexical ones.

Finally, note that the first-personal argument against physicalism would continue to go through if Premise 3 were reworded as the claim "*If* there are first-personal and indexical facts, then these do not supervene on third-personal and non-indexical facts", and this claim seems even harder to deny than Premise 3 in its original form.

In sum, I tentatively suggest that the first-personal argument against physicalism is not only valid, but also sound.

6. Beyond physicalism

All the arguments against physicalism, including the first-personal argument, suggest that the physicalist ontology is incomplete: its inventory of facts leaves out some facts, namely, some of the facts about subjective experience. To accommodate reality as it presents itself to us, this suggests, we must postulate a richer ontology, one that includes some "further facts", beyond conventional physical ones.[69] What kind of non-physicalist ontology could do the job?

David Chalmers has illustrated the challenge by reference to an analogy from the history of science. He notes that the theoretical move of accepting a richer ontology to account for some hitherto unexplained phenomenon is not unprecedented.[70] Specifically, he gives the example of electromagnetic fields invoked by James Clerk Maxwell and others in the 19th century to

[68] See also List (2023).
[69] On the case for further facts, see also Conitzer (2019).
[70] See Chalmers (1996).

explain electromagnetism. The previous ontology of the physical sciences, from Newtonian physics, was insufficient to incorporate the new phenomenon, and a new ontological ingredient, electromagnetic fields, had to be postulated as part of the ontological furniture of the world.

Schematically, let's write *P* for the classical physical ontology prior to the introduction of electromagnetism; and *P*+ for the revised ontology that adds electromagnetic fields. Then *P*+ is richer than *P*: it postulates more facts and properties, and some of the *P*+ facts do not supervene on the *P* facts. Nevertheless, there is nothing mysterious or unscientific about postulating the *P*+ ontology to account for electromagnetism. If it is the most parsimonious ontology that accommodates the phenomena to be explained, then we have good reasons to embrace it.

For Chalmers, the situation is similar in the case of consciousness. The conventional physicalist ontology is analogous to the *P* ontology in the historical example, whereas the ontology that is needed to accommodate subjective experience is analogous the *P*+ ontology. A science of consciousness must postulate such a richer ontology to accommodate first-person data.

Chalmers's strategy, then, is to suggest that we must recognize that there are not only physical properties in the world, but also phenomenal ones, where – importantly – the latter do not supervene on the former. There might still be some nomological constraints governing the relationship between physical and phenomenal properties – certain "psycho-physical" laws – but these do not hold as a matter of metaphysical necessity; they are contingent laws of nature. This kind of "naturalistic dualism" would explain why

(i) knowledge of phenomenal facts is not learnable from knowledge of the physical facts alone (there is no entailment relation between the two);
(ii) the facts about conscious experience could come modally apart from the physical facts (there is no relationship of metaphysical supervenience between the two); and
(iii) zombies are conceivable (a scenario in which the physical facts are present but the phenomenal ones absent is logically coherent and even metaphysically possible, if naturalistic dualism is correct).

So, Chalmers's proposal would seem to offer a structurally viable non-physicalist theory in response to the epistemic, modal, and conceivability arguments against physicalism. But would it also answer the challenge raised by the first-personal argument?

My impression is that it wouldn't. Note that, for Chalmers, what distinguishes the dualistic worldview from the physicalist one is simply that it postulates further properties that populate our world in addition to the physical properties. Indeed, the strategy of going beyond physicalism by postulating an amended inventory of fundamental properties is not unique to Chalmers's proposal but shared by several mainstream non-physicalist theories, including the recently influential "double-aspect", "Russellian", or "neutral" monist views. All these theories seem to have inherited one important structural feature from physicalism, which is sometimes overlooked. It is the assumption that there is a single, unified world, which is populated by certain properties, some of which are fundamental while others may be non-fundamental but

supervenient on the fundamental ones. The different theories merely disagree on what the relevant inventory of properties is and how they are related to one another and partitioned into fundamental and non-fundamental ones.

Crucially, when framed like this, each of the different theories still gives us a third-personal, non-indexical picture of the world: a picture of the world as seen by an Olympian observer looking at the world from the outside and asking which properties populate it. The theories are thus still formulated from what Nagel would call "the view from nowhere".[71]

If subjective experience is irreducibly first-personal and indexical, however, then it's not clear that any of these theories genuinely captures this first-personal and indexical character. In particular, given the structure just described, it's unclear how far these theories manage to go beyond physicalism when it comes to accommodating first-personal and indexical facts. When asked to give us the total inventory of facts making up the world – i.e., everything that is the case according to those theories – this inventory would still seem to be third-personal, non-indexical, and non-perspectival: an inventory of facts as seen from the "view from nowhere". The worry, then, is that a version of the first-personal argument's second premise, which asserted that physical facts are third-personal and non-indexical, might still be true for the facts as inventorized by those mainstream non-physicalist theories.

Nagel himself recognizes that the logic of the first-personal argument (or its precursor that we can find in his work) extends to "most other theories of mind, including dualism":

> "if we follow out this type of argument, it will provide us with equally strong reasons for rejecting any view which identifies the subject of psychological states with a substance and construes the states as attributes of that substance. A noncorporeal substance seems safe only because in retreating from the physical substance as a candidate for the self, we are so much occupied with finding a subject whose states are originally, and not just derivatively, mine … that we simply postulate such a subject without asking ourselves whether the same objections will not apply to it as well: whether indeed any substance can possibly meet the requirement that its states be *underivatively* mine."[72]

Another way to express this worry, also anticipated by Nagel, is to note that, like physicalism, the mainstream non-physicalist theories still seem to leave an important question open, which Benj Hellie has called "the vertiginous question". Calling himself "the Hellie-subject" and his interlocutor "the Chalmers-subject", Hellie asks:

> "The Hellie-subject: why is it me? Why is it the one whose pains are 'live', whose volitions are mine, about whom self-interested concern makes sense? . . . Granted that the Hellie-subject is acquainted with a certain class of phenomenal properties: if that subject is acquainted with right-arm pain, then I will feel right-arm pain ... But ... the Chalmers-subject is also acquainted with a certain class of phenomenal

[71] I here draw on List (2023).
[72] See Nagel (1965, p. 354).

> properties: if that subject is acquainted with left-arm pain, then Chalmers will feel left-arm pain and I might not. So facts about which subjects are acquainted with what cannot answer our question. Why should the acquaintance-relations of the Hellie-subject ... be the ones relevant to what I feel?"[73]

Hellie argues that only what he calls an "inegalitarian" theory can answer this question and capture the subjective and indexical nature of conscious experiences.[74] An *inegalitarian* – or I prefer to say: *asymmetrical* – theory is one that draws a structural distinction between my conscious experiences and yours. There is a sense in which, from where I stand, my subjective experiences are first-personally present and yours are not.[75] And from where you stand, it is the other way round. The first-personal facts are non-invariant under changes in perspective. Any philosophical theory that looks at the world solely from some Olympian third-personal perspective doesn't seem to capture this. Such a theory would be insufficiently "asymmetrical".

As noted, physicalist theories are not alone in having this third-personal, non-indexical, and thereby insufficiently "asymmetrical" structure. Several mainstream non-physicalist theories, from Chalmers's naturalistic dualism to standard versions of monism, still take such an Olympian third-personal (or impersonal) perspective on the world and are therefore what Hellie calls "egalitarian".

Now, Chalmers and others might raise the following objection:[76]

> **Objection:** We cannot reasonably expect to find an explanation as to *why* I am having Christian's experiences rather than someone else's. It is misguided to think that Hellie's vertiginous question could have an answer. There is simply a brute fact that I am Christian and have Christian's experiences and not someone else's. And so, the failure to answer the vertiginous question cannot count as undermining naturalistic dualism and the other above-mentioned non-physicalist theories.

I think one can concede that there may not exist an explanation as to why I am having my conscious experiences rather than anyone else's – and thus grant that the vertiginous question could not have a satisfactory answer – and yet insist that what matters is the recognition *that there is an irreducibly first-personal fact to the effect that I am having my conscious experiences*, even if that fact is a brute one and has no further explanation. Recall that this point was also emphasized by Merlo in the earlier quote. Arguably, the problem with the mainstream non-physicalist theories, as with physicalism, is that, by giving us an ultimately third-personal picture of the word, they cannot even accommodate this fact as a brute fact, irrespective of whether the fact could be further explained.

To answer the challenge raised by the first-personal argument, we must look for a metaphysical theory that recognizes genuinely first-personal and indexical facts. The sorts of theories that

---

[73] See Hellie (2013, 309–310).
[74] Ibid.
[75] My reference to "first-personal presence" is inspired by Hare's notion of presence (2007, 2009).
[76] For Chalmers's own response to Hellie, see Chalmers (2013).

might structurally fit the bill are radical subjectivist or phenomenological theories for which realism about first-personal facts is central. In the recent literature within the analytic philosophy of mind, theories that may have resources to accommodate such facts include the first-personal realism described by Kit Fine, Caspar Hare's egocentric presentism, the subjectivist view of Giovanni Merlo, Olla Solomyak's account of the metaphysics of perspectives, my own proposal of a "many-worlds theory of consciousness", and most recently Martin Lipman's account of "subjective facts about consciousness".[77] All these theories imply, or are at least compatible with, a certain form realism about first-personal facts and draw a structural distinction between my own first-personal perspective and the perspectives of others. Related ideas can be found in Benj Hellie's "inegalitarianism", Ted Honderich's account of "subjective physical worlds", Marcus Arvan's account of how subjects experience the world in a multiverse, and Gabriel Vacariu's proposal of epistemologically different worlds, when interpreted ontically.[78] This is not the place to discuss any of these (revisionary) theories, and it is obviously far from clear which, if any, of them will ultimately be defensible.

I conclude by noting that the first-personal argument raises a significant challenge not just for physicalism, as conventionally understood, but for any purely third-personal and non-indexical theory of the world.

References

Arvan, M. (2013). "A New Theory of Free Will." *The Philosophical Forum* 44(1): 1–48.

Bourget, D., and D. Chalmers (2003). "Philosophers on Philosophy: The 2020 PhilPapers Survey." *Philosophers' Imprint*, forthcoming.

Braun, D. (2017). "Indexicals." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). <https://plato.stanford.edu/archives/sum2017/entries/indexicals/>.

Builes, D., and M. O. Impagnatiello (forthcoming). "An Empirical Argument for Presentism." *Oxford Studies in Metaphysics*.

Butterfield, J. (2011). "Emergence, Reduction and Supervenience: A Varied Landscape." *Foundations of Physics* 41(6): 920–959.

Callender, C. (2017). *What makes time special?* Oxford: Oxford University Press.

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

Chalmers, D. (2004). "How Can We Construct a Science of Consciousness?" In M. S. Gazzaniga (ed.), *The Cognitive Neurosciences III*, third edition, 1111–1120. Cambridge, MA: MIT Press.

[77] See Fine (2005), Hare (2007, 2009), Merlo (2016), Solomyak (2020), List (2023), and Lipman (2023).
[78] See Vacariu (2005), Hellie (2013), Arvan (2013), and Honderich (2014).

Chalmers, D. (2013). “The Contents of Consciousness: Reply to Hellie, Peacocke and Siegel.” *Analysis* 73(2): 345–368.

Churchland, P. M. (1981). “Eliminative Materialism and the Propositional Attitudes.” *Journal of Philosophy* 78(2): 67–90.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Conitzer, V. (2019). “A Puzzle about Further Facts.” *Erkenntnis* 84(3): 727–739.

Crane, T. (2003). “Subjective facts.” In H. Lillehammer and G. Rodriguez-Pereyra (eds.), *Real metaphysics*, 68–83. London: Routledge.

Davidson, D. (2001). “Mental Events.” In *Essays on Actions and Events*, 2nd ed., 207–227. Oxford: Clarendon.

Dennett, D. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Dewar, N. (2019). “Supervenience, Reduction, and Translation.” *Philosophy of Science* 86(5): 942–954.

Di Biagio, A., and C. Rovelli (2022). “Relational Quantum Mechanics is About Facts, Not States: A Reply to Pienaar and Brukner.” *Foundations of Physics* 52(62): 1–21.

Fine, K. (2005). “Tense and Reality.” In *Modality and Tense: Philosophical Papers*, 261–320. Oxford: Oxford University Press.

Fodor, J. (1974). “Special Sciences (or: The Disunity of Science as a Working Hypothesis).” *Synthese* 28(2): 97–115.

Fuchs, C. (2010). “QBism, the Perimeter of Quantum Bayesianism.” Arxiv.org. <https://doi.org/10.48550/arXiv.1003.5209>.

Gendler, T., and Hawthorne, J. (2002). *Conceivability and Possibility*. Oxford: Clarendon.

Gillett, C., and B. Loewer (eds.) (2001). *Physicalism and Its Discontents*. Cambridge: Cambridge University.

Hare, C. (2007). “Self-Bias, time-bias, and the metaphysics of self and time.” *Journal of Philosophy* 104(7): 350–373.

Hare, C. (2009). *On Myself, and Other, Less Important Subjects*. Princeton: Princeton University Press.

Hellie, B. (2013). “Against egalitarianism.” *Analysis* 73(2): 304–320.

Hill, C. (2016). “Conceivability and Possibility.” In H. Cappelen, T. Szabó Gendler, and J. Hawthorne (eds.), *The Oxford Handbook of Philosophical Methodology*, 326–347. Oxford: Oxford University Press.

Honderich, T. (2014). *Actual Consciousness*. Oxford: Oxford University Press.

Howell, R. (2016). *Consciousness and the limits of objectivity: The case for subjective physicalism*. Oxford: Oxford University Press.

Ismael, J. T. (2007). *The Situated Self*. Oxford: Oxford University Press.

Jackson, F. (1982). “Epiphenomenal qualia.” *Philosophical Quarterly* 32(127): 127–136.

Kaplan, D. (1989). “Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals.” In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes From Kaplan*, 481–563. Oxford: Oxford University Press.

Kim J. (2005). *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Leuenberger, S. (2018). “Global Supervenience without Reducibility.” *Journal of Philosophy* 115(8): 389–422.

Levine, J. (1983). “Materialism and Qualia: The Explanatory Gap.” *Pacific Philosophical Quarterly* 64(4): 354–361.

Lewis, D. (1979). “Attitudes de dicto and de se.” *The Philosophical Review* 88(4): 513–543.

Lewis, D. (1983). “Postscript to ‘Mad Pain and Martian Pain’.” In *Philosophical Papers*, vol. 1, 130–132. Oxford: Oxford University Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Lipman, M. (2023). “Subjective Facts about Consciousness.” *Ergo: An Open Access Journal of Philosophy* 10: 19. <DOI:10.3998/ergo.4649>.

List, C. (2019). “Levels: Descriptive, explanatory, and ontological.” *Noûs* 53(4): 852– 883.

List, C. (2023). “The Many-Worlds Theory of Consciousness.” *Noûs* 57(2): 316–340.

Madell, G. (2003). “Materialism and the First Person.” *Royal Institute of Philosophy Supplements* 53: 123–139.

Merlo, G. (2016). “Subjectivism and the mental.” *Dialectica* 70(3): 311–342.

Mermin, N. D. (2019). “Making better sense of quantum mechanics.” Reports on Progress in Physics 82(1): 1–16.

Nagel, T. (1965). “Physicalism.” *The Philosophical Review* 74(3): 339–356.

Nagel, T. (1974). “What is it like to be a bat?” *The Philosophical Review* 83(4): 435–450.

Nagel, T. (1986). *The View From Nowhere*. New York: Oxford University Press.

Nemirow, L. (1980). “Review of Thomas Nagel, *Mortal Questions*.” *The Philosophical Review* 89(3): 473–477.

Nida-Rümelin, M. and D. O Conaill (2021). “Qualia: The Knowledge Argument.” In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). <https://plato.stanford.edu/archives/sum2021/entries/qualia-knowledge/>.

Nozick, R. (2001). *Invariances: The Structure of the Objective World*. Cambridge, MA: Harvard University Press.

Papineau, D. (2001). “The Rise of Physicalism.” In C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*, 3 – 36. Cambridge: Cambridge University.

Putnam, H. (1975). “Philosophy and Our Mental Life.” In *Philosophical Papers. Vol. 2, Mind, Language and Reality*, 291–303. Cambridge: Cambridge University Press.

Renero, A. (2023). “Kripke’s knowledge argument against materialism.” *Philosophical Perspectives* 37: 370–387.

Rosen, G. (2010). “Metaphysical Dependence: Grounding and Reduction.” In B. Hale and A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*, 109–135. Oxford: Oxford University Press.

Rovelli, C. (1996). “Relational quantum mechanics.” *International Journal Theoretical Physics* 35: 1637–1678.

Rudder Baker, L. (1998). “The First-Person Perspective: A Test for Naturalism.” *American Philosophical Quarterly* 35(4): 327–348.

Rudder Baker, L. (2007). “How Successful is Naturalism?“ In G. Gasser (ed.), *Publications of the Austrian Ludwig Wittgenstein Society*, 203-226. Frankfurt: Ontos-Verlag.

Schaffer, J. (2009). “On What Grounds What.” In D. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 347–383. Oxford: Oxford University Press.

Smith, D. W. (2018). “Phenomenology.” In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). <https://plato.stanford.edu/archives/sum2018/entries/phenomenology/>.

Solomyak, O. (2020). “Temporal ontology and the metaphysics of perspectives.” *Erkenntnis* 85(2): 431–453.

Stoljar, D. (2001). “Two Conceptions of the Physical.” *Philosophy and Phenomenological Research* 62(2): 253-281.

Stoljar, D. (2021). “Physicalism.” In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). <https://plato.stanford.edu/archives/sum2023/entries/physicalism/>.

Vacariu, G. (2005). “Mind, brain, and epistemologically different worlds.” *Synthese* 147(3): 515–548.

Wallace, D. (2014). *The Emergent Multiverse: Quantum Theory According To The Everett Interpretation*. Oxford: Oxford University Press.

Zahavi, D. (2017). “Thin, Thinner, Thinnest: Defining the Minimal Self.” In C. Durt, T. Fuchs, and C. Tewes (eds.), *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, 193–200. Cambridge, MA: MIT Press.