# A quadrilemma for theories of consciousness

Christian List*

Discussion paper, this version 18 October 2023

**Abstract:** In this discussion paper, I argue that no theory of consciousness can simultaneously respect four initially plausible metaphysical claims – namely, "first-person realism", "non-solipsism", "non-fragmentation", and "one world" – but that any three of the four claims are mutually consistent. So, theories of consciousness face a "quadrilemma". Since it will be hard to achieve a consensus on which of the four claims to retain and which to give up, we arrive at a landscape of competing theories, all of which have pros and cons. I will briefly indicate which kinds of theories correspond to the four horns of the quadrilemma.

## 1. Introduction

It is widely felt that the study of consciousness has reached an impasse. The field is deeply divided along several dimensions: between those who seek to give a physicalist and/or materialist account of consciousness and those who embrace dualist or other non-physicalist views; between those who think that – with or without physicalism – we should seek to give a "third-personal" and "objective" account of consciousness and those who think that this is infeasible or even misguided; between those who think that consciousness is somehow fundamental and those who think it is merely derived; between those who follow panpsychists in thinking that consciousness, at least in some form, is quite ubiquitous in the world and those who think it is rather special; and so on. David Chalmers (2018) has coined the term "the meta-problem of consciousness" to refer to the problem of explaining why it seems (to many of us) to be so hard to explain consciousness and why there is so little agreement on both substantive and methodological questions concerning its explanation.

My aim in this paper is to draw attention to one perhaps under-appreciated aspect of the difficulty of explaining consciousness. I will argue that any attempt to explain how consciousness fits into the world faces a "quadrilemma":

> There are four at first sight plausible claims that we might expect any satisfactory metaphysical and/or scientific theory of consciousness to be consistent with, but those four claims are mutually inconsistent. Any theory can retain at most three of them at once and must give up at least one.

* This paper is a sequel to, and draws on, List (2023a, 2023b). It is based on material first presented as part of the online Wendy Huang Lectures on "What's wrong with physicalism" at the National Taiwan University, June 2023. There I presented the quadrilemma as a trilemma, taking as given the first of the four claims ("first-person realism"), which I had already defended in the run-up to introducing the conflict between the other three claims. I am grateful to the participants for helpful comments, especially the discussants at these lectures, Lok-Chi Chan, Tony Cheng, Ye Feng, Shao-Pu Kang, Plato Tse, and Wenjun Zhang. I have also benefitted from discussions with Jonathan Birch, David Chalmers, Caspar Hare, and Anna Mahtani and with audiences to whom I presented my previous work on consciousness, as well as from the anonymous referee reports on List (2023a).

Since different people are likely to disagree about which of the four claims to retain and which to give up, we arrive at a landscape of competing theories, all of which have something going for them, but all of which also leave some participants to the debate unsatisfied.

**2. The quadrilemma**

I will first state the four claims and then explain why each of them is at least initially plausible and why they are mutually inconsistent.

> **First-person realism:** For any conscious subject, there are first-personal facts.
>
> **Non-solipsism:** More than one conscious subject is real.
>
> **Non-fragmentation:** The totality of facts that hold in any given world are compossible.
>
> **One world:** Reality consists of one world, not of many.

Let me begin with first-person realism (an idea that also occurs in Fine 2005 and Merlo 2016, as discussed later). A widely recognized feature of phenomenal consciousness is its first-person nature. My conscious experiences are first-person experiences. Consciousness is not merely something that is happening out there in the world impersonally, but *I* am conscious. I have experiences, perceptions, feelings, sensations, and so on. The first-person nature of consciousness is one of the things on which there is some common ground between many analytic philosophers of consciousness and phenomenologists in the tradition of Edmund Husserl and others. On the analytic side, for example, David Chalmers (2004, p. 1111) writes:

> "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."

And Thomas Nagel, cited by analytic philosophers and phenomenologists alike, speaks of conscious states as having "essential subjectivity" (1965, p. 354) and as being "essentially connected with a single point of view" (1974, p. 437), and he emphasizes as central

> "the fact that *I* (or any self), and not just that body, am the subject of those states" (1965, p. 354, emphasis in the original).

Relatedly, Lynne Rudder Baker (2013, pp. xiii–xiv) notes that

> "[N]aturalism takes the world to be impersonal; what exists are all individuals and all their properties, but none of these requires appeal to anything expressible in the first person."

But she observes that such a worldview leaves out any subject's first-person perspective (ibid.):

> "How, then, is there a place for the putative fact that some particular person is me? […] How could a centerless world accommodate me? […]
>
> [T]he answer to the questions in the last paragraph, on naturalistic views, is that there are no irreducibly first-person facts. I shall argue otherwise: If we take first-person facts to entail properties expressible only in the first person ('first-person properties'), then […] there are irreducible first-person facts."

Similarly, Dan Zahavi (2017, p. 194) observes:

> "[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."

These ideas lend at least some initial support to the claim that, for any conscious subject, there are first-person facts, such as, in my case, the fact that I am currently in a particular experiential state. Later I will discuss the opposing view, which denies the existence of first-person facts or even the notion that there is a third-person/first-person distinction at the level of facts at all.

Next, consider non-solipsism. This should be quite uncontroversial. Apart from a few adherents to solipsism, most people are likely to believe that they are not the only conscious subject (for discussion, see, e.g., Avramides 2020). It is very reasonable to expect a good theory of consciousness to vindicate this idea. Consciousness occurs not just in myself but in many subjects, at least in all awake and non-comatose people and plausibly also in many non-human animals, including but not restricted to the great apes. A theory that asserts that I am the only conscious subject would be extremely counterintuitive.

Third, let me move on to non-fragmentation. This is arguably another uncontroversial thesis, in fact so uncontroversial that it is seldom spelt out explicitly. If we think of a "world" – either the actual world or some other possible world – as being "populated" by a large body of facts (plausibly, the world is constituted by the totality of facts that hold in it), then a basic necessary condition for the possibility of the world in question is that all those facts are compossible, i.e., it is possible for them to be simultaneously instantiated, namely at that world. A world consisting of mutually incompatible ("non-compossible") facts would be incoherent and thus not a possible world, let alone a candidate for being the actual world. At most so-called "impossible worlds", sometimes discussed in metaphysics, could include non-compossible facts, but such worlds are not possible ones. (On impossible worlds, see, e.g., Berto and Jago 2019.)

Finally, let me turn to one world. A central tenet of a standard scientific but also philosophical worldview is that reality consists of a single world – the actual world – which is shared by all

of us and of which science aspires to give us an objective picture. It is very unusual and non-standard for a scientific or philosophical worldview to postulate that there are many distinct worlds all of which are equally real even if only one is "actual" for any subject. (We find such an idea at most in some special interpretations of quantum mechanics such as many-worlds interpretations in the tradition of Everett, as discussed by Wallace 2014, or QBism, as discussed by Fuchs 2010 and Mermin 2019. Philosophical arguments against one world can be found in Vacariu 2005 and Gabriel 2015. Lewis 1986 defended a form of modal realism.)

I do not deny that one could object to some or even all of the four claims – indeed, I will argue that at least one claim must ultimately be dropped – but I suggest that the four claims each have some *initial* plausibility at least as baseline theses for a metaphysical investigation of consciousness. Needless to say, there may be other plausible claims that one might add to this list, but as I will now explain, even those first four claims cannot be simultaneously true.

The argument is relatively straightforward. Suppose non-solipsism is true. Then more than one conscious subject is real. For each of them, according to first-person realism, there are first-person facts. However, the totality of first-person facts for different conscious subjects are not compossible *qua first-person facts*. To see this, suppose I am in conscious state X, and you are in conscious state Y, where X and Y are the complete conscious states that we are each in, respectively. Thus, "I am in conscious state X" is a first-person fact for me, and "I am in conscious state Y" is a first-person fact for you. Moreover, the two conscious states, X and Y, *qua complete token subjective states that we are each in*, are mutually exclusive. After all, we are distinct subjects, with a different perspective on the world. An implication is that the conjunction of the first-person sentences "I am in conscious state X" and "I am in conscious state Y" is necessarily false (because X and Y are mutually exclusive). And so, the corresponding first-person facts are not compossible *qua first-person facts*: they cannot be co-instantiated from any single point of view. One of these facts is instantiated from where I stand, the other is instantiated from where you stand.[1]

Of course, there are corresponding third-person facts of the form "Christian is in conscious state X" and "Christian's interlocutor is in conscious state Y", and these are perfectly capable of being co-instantiated; they are entirely compossible. But if we accept first-person realism,

[1] Others who have noted the non-compossibility of different subjects' first-person facts include Fine (2005), who observes that if we accept different subjects' first-person facts as real, we might end up with a "fragmented" picture of reality, and Merlo (2016), who notes that recognizing the equal reality of such facts would (without some other theoretical move) lead to "an overall incoherent totality of facts" (p. 324). The argument can be formalized by representing first-person facts by means of first-personally centered propositions (List 2023a). Formally, a first-personally centered proposition is a set of first-personally centered worlds. A first-personally centered world (more on this below) is an ordered pair consisting of an ordinary, third-personal world and a first-person perspective on it. The totality of first-personally centered propositions that are true from where I stand is not consistent with the totality of first-personally centered propositions that are true from where you stand. Given our different perspectives, the intersection of all of these first-personally centered propositions (yours and mine) is empty: no first-personally centered world can satisfy all of them together.

we must not confuse the first-person fact that *I* am in conscious state X with the third-person fact that Christian is in conscious state X. The latter fact, here expressed in third-person language, holds as much for you as it does for me. What is distinctive about a first-person fact is that it is not invariant under all shifts in the subjective perspective, and this non-invariance is a core ingredient of its "essential subjectivity", to use Nagel's term again. The first-person and third-person facts are not the same precisely because only the former but not the latter has this "essential subjectivity".

Now, if – as just assumed – two or more conscious subjects are real, each with first-person facts, and those first-person facts are not compossible, it follows that

- *either* the claim of non-fragmentation is false, and the world subsumes those non-compossible facts;
- *or* the claim of one world is false, and there are many worlds, namely one "subjective" world for each conscious subject, rather than just a single "objective" world.

Thus, if we accept first-person realism and non-solipsism, we must give up either non-fragmentation or one world. If we take the first route (dropping non-fragmentation), we can retain the claim that there is a single world, but this world will be internally fragmented and thereby incoherent: not all the facts populating it can be instantiated together. If we take the second route (dropping one world), we can retain the traditional idea that the totality of facts making up any world are co-instantiated in that world, so that "worlds" are always internally coherent, but we must embrace the view that different subjects are associated with different "subjective worlds".

In sum, first-person realism, non-solipsism, non-fragmentation, and one world are mutually inconsistent. We cannot accept all four claims together.

## 3. Escape routes from the quadrilemma

I will now show that, while the four claims are mutually inconsistent, any three of them can be simultaneously true. I will illustrate this by showing that different existing metaphysical theories of consciousness differ in which of the claims they retain, and which they give up.

### *3.1. Giving up first-person realism*

In the analytic philosophy of consciousness, the most common strategy to avoid the quadrilemma is to reject the idea that there are genuinely first-personal facts and to insist that the first-person/third-person distinction is not a distinction at the level of facts but only a distinction at the level of language or cognitive representation.[2] According to this view,

[2] I have discussed this objection to first-person realism in more detail in List (2023b). The objection is influenced by an approach to the semantics of indexicals (Kaplan 1989), according to which indexical sentences (of which a first-person sentence is a special case) express non-indexical propositions once we fix the context of utterance.

there can be different modes of presentation of certain facts, such as first-personal and third-personal modes of presentation, and these correspond to different ways of linguistically describing the facts that are being represented, but the facts themselves are always the same. So, the fact that I am in a particular conscious state is just the fact that Christian is in that state. I might have access to a special mode of representing this fact, which you do not have access to, but there is no further fact here. There is no first-person fact that *I* am in the conscious state in question as distinct from the third-person fact that Christian is in that state.

If we accept the idea that facts are always impersonal or aperspectival – not endowed with any perspective – then there is no longer any problem in embracing the view that the totality of facts about different subjects' conscious states can be co-instantiated in a single, objective, and non-fragmented world. Most standard theories in the analytic philosophy of consciousness are committed to a version of this view, whether explicitly or implicitly.

Physicalist views – whether reductive or non-reductive – obviously fall into this camp, but so do dualist views – whether of the traditional Cartesian sort or of the updated naturalistic variety defended by Chalmers (1996) – and arguably also the various recently influential Russellian, neutral, or double-aspect monist views. (For a good overview of the theoretical landscape, see Chalmers 2010.) Although these theories differ in many respects, one often-overlooked commonality among all of them is that they assume that there is a single objective and non-fragmented world, in which certain properties are instantiated (some of which may be fundamental while others may be derived), and they all try to give us an "inventory" of those properties. The differences between the theories lie in the details of this inventory. The theories give us different answers to questions such as the following: Are there only physical properties or also phenomenal ones? Do phenomenal properties supervene on physical ones or not? Is there more than one kind of fundamental property? However, they all support non-solipsism, non-fragmentation, and one world, and, in consequence, they must reject first-person realism.

Of course, all those mainstream theories, except so-called illusionist theories (a special subset among the physicalist ones, e.g., Frankish 2016), will insist that they are realist with respect to consciousness. Nevertheless, by denying first-person realism, they are committed to a form of anti-realism about genuinely first-personal facts. Whether one considers this a feature or a bug of those theories depends on one's stand towards first-person realism.

*3.2. Giving up non-solipsism*

A relatively uncommon but coherent strategy to avoid the quadrilemma is to give up non-solipsism and to accept that there is just one first-personally conscious subject, namely myself. One well-developed theory along these lines is Caspar Hare's "egocentric presentism" (2007, 2009). This is a fairly radical theory according to which I, as a conscious subject, live in my own "subject world", defined as

> "a world in which there are functionally sentient creatures, the experiences of one and only one of which have the monadic property of being-present" (Hare 2007, p. 366).

In my subject world, there are other sentient creatures in a purely functionalist sense – that is to say, they display the functions of cognition and awareness – but their conscious experiences are not present.

Egocentric presentism is an instance of what Benj Hellie (2013) calls an "inegalitarian" (or I prefer to say: "asymmetrical") theory of consciousness. It draws a structural distinction between my own conscious experiences, which are first-personally present to me, and the conscious experience of others, if any, which are first-personally inaccessible to me and "absent" from where I stand.

According to Hare's "egocentric presentism", it may be true, in my subject world, that I am not the only one in pain, and that other sentient beings can be said to be in pain too, where this is understood in some third-personal and functionalist sense. However, Hare writes:

> "For an egocentric presentist, the situations are not symmetrical. It's not that my pain is present to me and his present to him. Mine is present and his is absent" (2007, p. 372).

This should illustrate why egocentric presentism has a solipsistic character. Indeed, Hare writes:

> "an egocentric presentist believes that only one subject world exists. There are no other subject worlds" (2009, p. 41).

At most, Hare seems to suggest, we may *hypothetically imagine* the subject worlds of others, for instance when we think about what things would be like from another person's perspective, but those other subject worlds are sorts of fictions.

Evidently, Hare's theory has no difficulty supporting first-person realism, non-fragmentation, and one world, insofar as it postulates that there is a single non-fragmented world, namely my own subject world, which moreover encompasses all my first-personal facts. But the cost is a form of solipsism.

Many of us will find this hard to swallow. Of course, strictly speaking, none of us have any "hard" evidence that others have conscious first-personal experiences too. The hypothesis that others are zombies is empirically unfalsifiable. However, symmetry considerations – in the scientific sense of symmetry – make it implausible to think that I am the only first-personally conscious being, not to mention how morally troubling and/or lonely such a solipsistic scenario would feel (for discussion, see again Avramides 2020).

Hare himself recognizes the peculiar solipsistic character of his theory. Commenting about how things were before he was born, he writes:

> "Isn't it amazing and weird that for millions of years, generation after generation of sentient creatures came into being and died, and all the while there was this absence [i.e., no conscious experiences were ever first-personally present], and then one creature, CJH [Caspar Hare], unexceptional in all physical and psychological respects, came into being, and POW! Suddenly there were present experiences!" (2009, p. xv)

Still – and to his credit, from the perspective of philosophical coherence – he bites the bullet and embraces the noted implications.

*3.3. Giving up non-fragmentation*

A third strategy to avoid the quadrilemma is to give up the claim that any world – whether actual or possible – consists only of compossible facts. The result would be a theory according to which a world can be internally fragmented. A "world" can then be such that only some proper subsets of the facts populating that world – "fragments" of the world – can be jointly instantiated. Any such fragment would be an internally coherent ("compossible") collection of facts, but there would be no coherence ("compossibility") across different fragments.

Kit Fine (2005) describes – without endorsement – a theory along these lines, which he simply calls "first-personal realism". He writes:

> "The first-personal realist believes that there are distinctively first-personal facts. Reality is not exhausted by the 'objective' or impersonal facts but also includes facts that reflect a first-person point of view" (p. 311).

Fine's theory explicitly accommodates first-personal facts and distinguishes them from third-personal ones, thereby supporting the first of my four claims above, which I have labelled "first-person realism". However, a difficulty with the theory, which Fine recognizes, is that if reality includes such first-personal facts, and it includes them for both you and me (which he finds more plausible than the solipsist alternative), then reality must somehow be fragmented. At least under the widely held assumption that "reality" and "the world" are more or less synonymous, in the sense that reality consists of only one world, the theory described by Fine upholds one world while giving up non-fragmentation. This is broadly how Fine presents the theory (or at least the version of it that he conditionally recommends, *if* one is inclined to accept realism about first-person facts): he describes it as "fragmentalist".

If we go with this framing of the theory, however, there is a significant theoretical cost. Non-fragmentation is a key tenet of the standard understanding of what a world is, both in metaphysics and in logic, and postulating fragmented worlds would require a significant revision of the way we think about worlds in philosophy, logic, and scientific modelling.

*3.4. Giving up one world*

A fourth route out of the quadrilemma is to give up the claim of one world. One theory that does so is the "many-worlds theory of consciousness", which could also be called the "many-centered-worlds theory of consciousness", presented in List (2023a). (Earlier works that have given up the assumption of one world and defended metaphysical or epistemological theories without that assumption include Vacariu 2005, Honderich 2014, and Gabriel 2015.) As I tentatively prefer the many-worlds response to the quadrilemma to the others (though it is still counterintuitive), I will explain it in a bit more detail, without fully defending it.

To introduce the basic idea, it is helpful to recall Hare's "egocentric presentism". One can think of that theory as combining

(a) the thesis that each conscious subject has their own subject world

with (b) fictionalism about the subject worlds of others.

The combination of (a) and (b) is, of course, consistent with the idea that there is a single non-fragmented world, namely my own subject world. But as noted, the price to pay for this is the theory's solipsistic character. We can avoid this solipsism and still retain non-fragmentation if we accept a version of (a) while replacing (b) with a form of realism, rather than fictionalism, about the subject worlds of others. This requires us to postulate many different "subjective worlds", one for each conscious subject. The "many-worlds theory of consciousness" does just this. It has two core features:

- It rejects the assumption that the first-person facts associated with different subjects' conscious experiences hold at one and the same world; instead, it associates different subjects with different "first-personally centered worlds", which coincide with respect to all third-personal facts but differ with respect to some first-personal facts.
- It embraces a special form of modal realism, according to which different subjects' first-personally centered worlds are all real, but only one of them is present for each subject.

As I will now explain, these ideas lend themselves to a neat formalization. The formal framework of "centered worlds" (a notion that goes back to W. V. Quine and David Lewis and is sometimes employed to capture indexical content) can be adapted and re-interpreted to represent first-personally centered worlds.

To sketch this formalization (drawing closely on the exposition in List 2023a), I begin with the notion of a "third-personal world". A "third-personal world" – call it $\omega$ – encompasses all third-person facts, i.e., all facts that hold from a third-person perspective and that are thereby invariant under shifts in subjective perspective. One could also think of those facts as impersonal facts and use the term "impersonal world" instead of "third-personal world". The

underlying fact-based definition of a world, in turn, goes back to Wittgenstein's dictum: "the world is everything that is the case". According to it, a "world" can be defined as the total collection of *facts* that hold in that world. A third-personal world, as I am defining it, subsumes all facts that hold third-personally or simply impersonally. Roughly speaking, this encompasses all facts that would feature in a complete, exhaustive description of that world by an omniscient but external observer – an observer taking what Thomas Nagel (1986) calls "the view from nowhere".

However, the collection of facts making up a third-personal world – i.e., the collection of all facts that hold third-personally or impersonally in that world – does not contain any subject's first-personal facts, such as the fact that *I* am in a particular experiential state right now. What perspective one occupies in relation to the given third-personal world is left open by it. Note that, even from complete third-personal information about who experiences what, it would not follow who *I* am and what perspective *I* occupy inside that world. These questions are not settled – but are left underdetermined – by the totality of all third-personal facts.[3]

To place a subject such as myself in the world, we must introduce something in addition to the third-personal world $\omega$, namely a "locus of subjectivity" inside it. I call this $\pi$. It encodes a subject's first-person perspective on $\omega$. A "first-personally centered world" is then defined as an ordered pair $<\omega,\pi>$ consisting of a third-personal world $\omega$ and a perspective $\pi$.

Formally, this definition is an instance of the standard definition of a "centered world", as previously defined by Quine and Lewis: an ordered pair consisting of a standard, "uncentered" world and some "location" or "center" in it (Liao 2012). However, "centers" are traditionally understood as something quite narrow: they are usually taken to be mere spatio-temporal coordinates or pointers to an individual in the world, like the dot indicating one's location on a map. The many-worlds theory of consciousness requires us to understand "centers" as something richer, namely as encoding a subject's entire first-personal perspective on a world, in as much detail as needed for the pair $<\omega,\pi>$ to encode the totality of facts – third-personal and first-personal – about the world $\omega$ from the subjective perspective given by $\pi$.

A subject's first-personally centered world thus encompasses everything that is the case for that subject, which includes everything that is the case subject-invariantly and also everything that is the case for that subject alone. Among other things, this encompasses the totality of first-personal experiences of that subject, as first-personally presented to them. Generally, first-person facts hold only at first-personally centered worlds, not at third-personal ones. This is in line with Lynne Rudder Baker's above-quoted observation that a "centerless world" wouldn't accommodate a subject's first-person perspective; only a suitably centered world does.

[3] As Benj Hellie (2013) notes, even the totality of facts about who experiences what leave open the question of which of these experiences are *mine* and why. He calls the unanswered question the "vertiginous question".

The notion of a "first-personally centered world" resembles Hare's notion of a "subject world", although its definition is more abstract and not dependent on Hare's specific views about "presence". Moreover, a "third-personal world" corresponds to an equivalence class of first-personally centered worlds that are equivalent with respect to all third-personal facts but may differ with respect to the locus of subjectivity. Facts are objective if they are invariant under all shifts in locus of subjectivity, and subjective if they vary with some such shifts.

Unlike Hare's egocentric presentism, the many-worlds theory does not assert that only my own first-personally centered world is real, while those of others are fictional. Instead, it adopts a "modal realist" commitment to the reality of others' first-personally centered worlds. As noted, it asserts that there are many "parallel" first-personally centered worlds, all of which are real, but only one of which is present for each subject.

According to this theory, then, your first-personally centered world is as real as mine. Yet, my first-personally centered world is present for me, and yours is present for you. This picture is structurally similar to David Lewis's realism about possible worlds (1986), though applied to first-personally centered worlds, instead of third-personal or impersonal ones, and with worlds defined in the (Wittgenstein-inspired) fact-based way explained earlier.

It should be clear that the many-worlds theory supports first-person realism, non-solipsism, and non-fragmentation, while giving up the one-world picture of reality and replacing it with a picture in which there are as many first-personally centered worlds as there are conscious subjects. These differ not in their "reality", but just in their "presence" or "accessibility" for each subject.

**4. Concluding remarks**

My aim has been to discuss an under-appreciated quadrilemma for theories of consciousness. Although I have noted my tentative preference for the fourth route out of the quadrilemma over the first three, I have not sought to defend this route here. Indeed, I think that there are serious arguments for each of the four routes, and my aim has been merely to show that the quadrilemma offers a map of some of the metaphysical disagreements at issue.

In fact, the quadrilemma itself can be used to derive formally valid arguments in support of each of the competing routes. This is because, from any triple of the four claims, one can infer the negation of the fourth. For instance, those who reject first-person realism can argue for their view by accepting non-solipsism, non-fragmentation, and one world as premises. Proponents of a solipsist view along the lines of Hare's egocentric presentism can use first-person realism, non-fragmentation, and one world as premises. Fragmentalists of the kind described by Kit Fine can use first-person realism, non-solipsism, and one world as premises. And finally, proponents of a many-worlds theory can use first-person realism, non-solipsism, and non-fragmentation as premises. All four arguments are formally valid, in the sense that their premises entail their conclusions, but which argument, if any, we consider sound

depends on where we stand in the debate on which of the four claims to uphold and which to give up.

Others have equally engaged with the challenge of reconciling a number of initially plausible but ultimately conflicting claims that we might like a metaphysical theory of consciousness to support. In his discussion of first-personal realism, for instance, Kit Fine (2005) contrasts and compares several different versions of such a theory before conditionally recommending broadly the "fragmentalist" version summarized above. He discusses, for instance, a contrast between "standard" and "non-standard" versions of such a theory and between "fragmentalist" and "relativist" ones. While the distinctions associated with my four claims do not map exactly onto Fine's, my general investigation still very much echoes his.

Similarly, Benj Hellie (2013) is well aware of the challenges raised by the quest for a coherent theory of consciousness which takes the asymmetry between a subject's own conscious experiences and those of others seriously and does not leave certain explanatory gaps open, such as Hellie's "vertiginous question" of why *I* am having my conscious experiences and not those of anyone else – a question that remains unanswered by the totality of all third-personal facts about the world.

Thirdly, Giovanni Merlo's discussion of "subjectivism about the mental" (2016) raises many related issues. Merlo's subjectivism asserts that "one's own mental states are metaphysically privileged vis-à-vis the mental states of others, even if only subjectively so" (p. 311) and entails a form of realism (albeit a "subjectivist" one) about first-person facts. Furthermore, Merlo seems (implicitly, at the very least) aware of the fact that this kind of first-person realism cannot be consistently combined with the claims I have described as "non-solipsism", "non-fragmentation", and "one world", so that at least one of these claims must be dropped. I am less clear which of these claims Merlo would drop.

On one interpretation, Merlo leans towards a less radical form of "solipsism" that is nuanced enough to assuage some of solipsism's implausibility. Quoting Fine's (2005, p. 313) remark that "[i]t seems quite bizarre to suppose that, from among all the individuals that there are, the subjective world-order is somehow oriented towards me as opposed to anyone else", Merlo (2016, p. 324) asks:

> "Doesn't [subjectivism about the mental] deserve the same 'incredulous stare' with which I look at other far-fetched and outlandish philosophical theses?"

In answer, he writes:

> "the kind of inegalitarianism implied by SVM [Subjectivism about the Mental] is not so far-fetched and outlandish as a superficial understanding of the view might suggest. If SVM is true, reality is, indeed, oriented towards a single point of view. But remember that, given Subjectivism, reality is not objectively the way it is, so which point of view

> gets to be privileged is, itself, a subjective matter. The claim is not that my point of view is firstpersonal from every point of view, but only that it is firstpersonal – that the way things have always appeared to me to be (this individual being special […]) is also the way things are. SVM, then, is inegalitarian, but in a subtler – and, I think less incredible – way than would justify me to dismiss it out of hand." (2016, p. 324)

My inclination is to read this as affirming a subtle way of relaxing non-solipsism, though Merlo's position appears to be also compatible with the retention of non-solipsism and the relaxation of either non-fragmentation or (perhaps) one world. In a footnote, he adds the following qualification:

> "there might be ways to reconcile the thesis that the totality of facts is oriented towards one point of view with the idea that, most fundamentally, all points of view are metaphysically on a par. One option would be to adopt a conception on which the totality of what is most fundamentally the case extends beyond the totality of facts […] Alternatively, one could take all points of view to be on a par vis-à-vis truth simpliciter by treating them as different 'fragments' of an overall incoherent totality of facts […] My own preference goes to the first strategy – the second runs the risk of undermining the sense in which I am special vis-à-vis all other subjects" (ibid.).

The present discussion, I hope, illustrates that the tensions captured by the identified quadrilemma have already manifested themselves in earlier contributions to the debate about the metaphysics of consciousness and especially in debates about the status of the first-person perspective.

Finally, I want to note that the quadrilemma is relevant to the question of what a scientific explanation of consciousness could look like. It will make a difference to the structure of such an explanation how we answer the following questions:

(1) Does the *explanandum* include some first-personal facts for every conscious mind?
(2) Are there other conscious minds?
(3) Is the world internally coherent as opposed to internally fragmented?
(4) Is there one "objective" world as opposed to several "subjective" ones?

Crucially, giving a "yes" answer to all four questions appears to be incoherent. Depending on which of the four questions we answer in the affirmative and which in the negative, we are likely to arrive at different approaches to accommodating consciousness within a scientific worldview. Evidently, then, the identified quadrilemma matters, and it challenges us to come up with a coherent combination of answers to questions (1) to (4).

**References**


Avramides, A. (2020). "Other Minds." In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Winter 2020 Ed.). <https://plato.stanford.edu/archives/win2020/entries/other-minds/>.

Baker, L. R. (2013). *Naturalism and the First-Person Perspective.* Oxford: Oxford University Press.

Berto, F., and M. Jago (2019). *Impossible Worlds.* 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*, 3rd ed., 1111–1120. Cambridge, MA: MIT Press.

Chalmers, D. (2010). *The Character of Consciousness*. New York: Oxford University Press.

Chalmers, D. (2018). "The Meta-Problem of Consciousness." *Journal of Consciousness Studies* 25(9–10): 6–61.

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

Frankish, K. (2016). "Illusionism as a Theory of Consciousness." *Journal of Consciousness Studies* 23 (11–12): 11–39.

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

Gabriel, M. (2015). *Why the World Does Not Exist*. Cambridge: Polity.

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.

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

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.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Liao, S. (2012). "What Are Centered Worlds?" *The Philosophical Quarterly* 62(247): 294–316.

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

List, C. (2023b). "The first-personal argument against physicalism." Working paper, <https://philarchive.org/archive/LISTFA>

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.

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.

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

Wallace, D. (2014). *The Emergent Multiverse*. 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.