DOI: 10.1111/nous.12408

ARTICLE

Noûs

The many-worlds theory of consciousness

Christian List

LMU Munich

Correspondence
Christian List, Munich Center for Mathematical Philosophy, LMU Munich, Geschwister-Scholl-Platz 1, 80539 München, Germany.
Email: c.list@lmu.de

Abstract

This paper sketches a new and somewhat heterodox metaphysical theory of consciousness: the "many-worlds theory". It drops the assumption that all conscious subjects' experiences are features of one and the same world and instead associates different subjects with different "first-personally centred worlds". We can think of these as distinct "first-personal realizers" of a shared "third-personal world", where the latter is supervenient, in a sense to be explained. This is combined with a form of modal realism, according to which different subjects' first-personally centred worlds are all real, though only one of them is present for each subject. The theory offers a novel way of capturing the irreducibly subjective nature of conscious experience without lapsing into solipsism. The paper also looks at some scientific theories of consciousness, such as integrated information theory, through the proposed lens and reconsiders the hard problem of consciousness.

KEYWORDS

centred worlds, first-personal vs third-personal facts, hard problem, many worlds, modal realism, phenomenal consciousness, presentism

1 | INTRODUCTION

The aim of this paper is to sketch a somewhat heterodox metaphysical theory of consciousness: the "many-worlds theory". The theory seeks to give an account of how consciousness, in the sense of subjective experience, fits into the world and how it relates to other, non-subjective features. To capture the irreducibly subjective nature of conscious experience, the theory drops the common assumption that all conscious experiences – yours, mine, and those of everyone else – are features of one and the same world and postulates instead that different conscious subjects are associated with different "first-personally centred worlds". Roughly speaking, I am experientially located in my own first-personally centred world, in which my own conscious experiences are first-personally present. You are experientially located in your first-personally centred world, in which your conscious experiences are first-personally present. We can think of these as distinct and parallel "first-personal realizers" of a shared "third-personal world" – hence the reference to "many worlds". The third-personal world, on this picture, is supervenient in a sense to be explained, and "multiply realizable" at the first-personal level, where each realization corresponds to one conscious subject.

The many-worlds theory is metaphysical, insofar as it is concerned with the *metaphysical* question of what kind of ontology we must postulate in order to accommodate consciousness and to capture its relationship to the physical realm. The theory does not address the *scientific* question of what the physical and/or neural correlates of consciousness are. Different answers to that question are compatible with the metaphysical picture to be sketched.

The theory is heterodox, insofar as it differs from the standard theories of consciousness in the analytic philosophy of mind. Its closest precursors are some "presentist" and "subjectivist" theories. Caspar Hare and Ted Honderich, in particular, have introduced the notions of a "subject world" and a "subjective physical world", respectively, which – despite differences in the resulting philosophical picture – are structurally similar to the notion of a "first-personally centred world".[1] Other precursors are Markus Gabriel's and Gabriel Vacariu's critiques of the assumption that there is a single world. Gabriel suggests that the world as such – as a "container" of everything – does not exist, and Vacariu argues that the solution to the mind-body problem lies in rejecting the traditional postulate of "one world" and recognizing the existence of "epistemologically different worlds".[2] Finally, some of the ideas discussed

[1] Hare (2007, 2009) defines a "subject world" 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*" (2007, 366). Honderich (2014) writes, "being perceptually conscious now is the existence of a part or piece or stage of a sequence that is one *subjective physical world* among very many, as many as there are sets of perceivings of single perceivers" (192). Other related contributions are Fine's (2005) discussion of "first-personal realism", Hellie's (2013) "inegalitarian" approach to consciousness, and Merlo's (2016) "subjectivist view of the mental". For a discussion of what is at stake in the debate on presentism more generally (mainly in temporal ontology), see further Solomyak (2020).

[2] See Vacariu (2005) and Gabriel (2015). Vacariu develops his critique of the assumption of "one world" around "three elements: the subject, the observed object, and the conditions of observation (given by the internal and external tools of observation)" (2005, 515). He writes: "because of the conditions of observation, the mind and the brain belong to epistemologically different worlds" (ibid.). Gabriel argues that there is no such thing as "the world" as the complete and exhaustive "container" of everything. These works, which I became aware of only when the present work was already quite advanced, offer somewhat different philosophical pictures from the one sketched here and do not appear to build on, or to make connections with, Lewis's modal realism or the notion of centred worlds.

in this paper are similar in spirit to ideas from phenomenology in the tradition of Edmund Husserl.[3]

While some earlier "presentist" theories, such as Hare's "egocentric presentism", treat only a single subject's world as real and thereby invite a solipsistic interpretation, the many-worlds theory mitigates this problem by embracing a form of modal realism with respect to first-personally centred worlds. According to this modal realism, different subjects' first-personally centred worlds are all real, though only one of them is present for each subject. So, your conscious experiences are no less real than mine. It just so happens that their "locus" is a different first-personally centred world from mine, which is not present for me, and therefore I have no first-personal access to your conscious experiences. This is broadly analogous to David Lewis's classic modal realism with respect to uncentred worlds, according to which many possible worlds are real, though only one of them is actual.[4] Another earlier notion of "many worlds" with which the present picture could be compared appears in the Everett interpretation of quantum mechanics, though in quite a different form and context.[5]

The paper is structured as follows. In Section 2, I introduce the classic challenge of explaining how the first-person, subjective character of conscious experience fits into the world. In Section 3, I review how analytic philosophers of mind typically frame the issue and point out that implicit in the standard theories is a "one-world picture". In Section 4, I explain what I take to be the main shortcomings of that picture, most notably, its failure to do full justice to the subjective nature of conscious experience. In Sections 5, 6, and 7, I sketch the alternative "many-worlds picture" and explain how it arguably avoids those shortcomings. In Sections 8 and 9, I discuss how what I call the "first-personal" and "third-personal levels" are related to one another and look at some scientific theories of consciousness through the present lens. In Sections 10 and 11, I revisit the "hard problem" of consciousness and conclude. Overall, my aim is not to present a fully committed defence of the many-worlds theory, but rather to show that the theory merits further consideration.

2 | THE CHALLENGE

My point of departure is the familiar challenge identified by many classic works on consciousness.[6] As conscious subjects, we are not merely biological systems that function in certain ways and can be described from some external, third-person perspective, just as we describe any ordinary physical process. Rather, we experience the world from a first-person perspective. There is something it is like to *be* a conscious subject, *for* that subject, as Thomas Nagel puts it.[7] Many or perhaps even most other entities in the world – from chairs and smartphones to ecosystems – presumably lack such a first-personal "inner life". Everything that can be said about those systems can be said from the outside, from a third-person perspective.

[3] See, e.g., Gallagher and Zahavi (2019) and Zahavi (2017). Developing these connections is beyond this paper's scope.

[4] See Lewis (1986).

[5] For an overview and discussion, see, e.g., Wallace (2012).

[6] See, e.g., Nagel (1974), Jackson (1982), Levine (1983), Chalmers (1995, 1996), and Nida-Rümelin (1995).

[7] See Nagel (1974). While I find the third-versus-first-person distinction congenial for elucidating Nagel's characterization of consciousness, Nagel's paper itself contains only some passing references to the "first person".

How can we study consciousness? David Chalmers describes the explanatory challenge 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."[8]

The third-person and first-person data raise very different explanatory problems. As Chalmers points out, the third-person data can in principle be explained using ordinary scientific methodology, such as in psychology and neuroscience. The data here concern, for instance, a subject's wakefulness and sleep, observable attention, cognitive processing, and reasoning capacities, as well as associated patterns of neural activity. We can study all of this in essentially the same way in which we study other phenomena in science, by formulating and empirically testing hypotheses about the relevant phenomena, where those hypotheses are expressed in third-person language. Chalmers calls these explanatory problems the "easy problems of consciousness".

The first-person data are harder to explain. To quote Nagel again:

> "[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."[9]

The "hard problem of consciousness", as Chalmers calls it, is to explain why we have first-person experiences at all. Why is there something it is like to *be* us, *for* us? Why are we not "zombies": hypothetical systems that are behaviourally and neurally identical to us and indistinguishable from us by any external, third-person observer, but which have no first-person experiences? Insofar as science looks at the world from a third-person perspective, an ordinary scientific approach seems incapable of pinpointing the difference between such hypothetical zombies and us. Indeed, if the data we wish to explain are irreducibly first-personal, then any third-personal scientific approach cannot adequately *describe* our explanandum, let alone provide an explanation. At best, it can give us an account of the third-personally observable phenomena that are *correlated* with first-person experience; but that's not the same as explaining first-person experience itself.

These observations have led some, such as Joseph Levine, to suggest that there is an "explanatory gap" between what the most extensive third-person scientific explanation can deliver, and what we need to account for when we are trying to explain consciousness.[10] A common way of making this point is to say that consciousness involves not just physical and functional properties, but also phenomenal or "experiential" properties and to argue that ordinary science, from its third-person perspective, might account for the physical and functional properties, but not for the phenomenal or "experiential" ones.

[8] See Chalmers (2004, 1111). My framing of the challenge draws on Chalmers (1995, 1996).

[9] See Nagel (1974, 436).

[10] See Levine (1983).

3 | THE STANDARD PICTURE

The standard theories of consciousness in the analytic philosophy of mind can be interpreted as attempts to answer the question of how physical and functional properties of the kinds studied in the sciences are related to phenomenal or "experiential" ones. Different theories can be characterized in terms of their positions on whether there is any explanatory gap between the former and the latter, and, if so, whether this is due to some further metaphysical gap. In short, they ask: is there any sort of gap between physical and phenomenal properties, and if so, is this gap merely epistemic or also ontic?

In answer to this question, broadly speaking, physicalist theories deny that there is any metaphysical gap between physical and phenomenal properties. For them, all properties, including those we commonly describe as "phenomenal", supervene on, and are grounded in, physical properties. At most, these theories acknowledge some kind of explanatory gap, which they then try to account for. Dualist theories assert that there is both an explanatory and a metaphysical gap, and they take the world to be populated by both physical and phenomenal properties, which are distinct from each other and neither of which ground the other. Other more sophisticated monist theories suggest that there is a single class of fundamental properties subsuming both physical and phenomenal aspects of the world. I set the details aside and refer the reader to an appendix.

What I want to note here is that, despite their differences, the standard theories have one key presupposition in common. Implicit in all of them is what we may call a "one-world picture". They assume:

> **One world**: There is a single world – the actual world – which accommodates *both* all physical phenomena *and* all conscious experiences, i.e., the conscious experiences of all the conscious subjects in the world.

The different theories just disagree on which kinds of properties must be present in that world – which properties must "populate" it – to accommodate consciousness along with physical phenomena. Each theory designates certain properties as fundamental and asserts that everything else supervenes on (and is grounded in) those properties. For instance, according to physicalism, the world is fully specified by its physical properties, whereas according to dualism, it is fully specified only by its physical *and* phenomenal properties together. The theories thus give us different answers to the question of which properties one would have to bring into existence to generate the world as it is. Would it suffice to put in place the properties of one kind alone, say physical properties, to generate everything including consciousness, or would we require properties of more than one kind?

Notwithstanding the disagreement about this question, the basic assumption that there is a single world in which the relevant properties are to be found is not in dispute in the mainstream debate. It is easy to miss the significance of that assumption. The one-world picture may seem so natural and self-evident – and all the action may seem to lie in the debate about which properties populate that one world – that we may even fail to recognize it as an assumption in the first place.

However, it is questionable whether the one-world picture does justice to the nature of conscious experience. I will now look at its shortcomings. Even if my critical remarks, which are also in line with other scholars' criticisms of the mainstream theories, don't amount to knock-down arguments, they should motivate us to consider an alternative picture with an open mind.

4 | WHAT ARE THE SHORTCOMINGS OF THE STANDARD PICTURE?[11]

First of all, by taking all conscious subjects' experiences, like all physical properties, to be features of one and the same world, the one-world picture does not fully capture the inherently perspectival, first-person, and subjective character of conscious experience. It lacks the resources to tell us *in structural terms* what the source of the subjectivity of conscious experiences is. The one-world picture simply takes the world to be populated by a large number of properties, some of which we call "physical", others "phenomenal". When asked why some of those properties correspond to objective features of the world while others correspond to subjective features, the one-world picture can only give us a rather stipulative and ad hoc answer and assert that this is how it is.[12] In effect, the one-world picture is still a third-personal picture: a picture of the world as it would be seen from an Olympian perspective, the "view from nowhere", as Nagel calls it.[13] Even though this third-personally described world is said to be populated by both "physical" and "phenomenal" properties, it is not clear how we get anything genuinely first-personal and subjective out of it.[14]

Secondly, the one-world picture does not capture the centrality of the subject within any conscious experience. There is a sense in which each of us, as a conscious subject, finds him- or herself at the centre of his or her conscious experiences. The one-world picture does not really account for this "subject-centredness". Indeed, the very notion of "the subject" or "the self" has remained elusive in the analytic philosophy of mind. It is widely held that there isn't any good support for treating "the subject" or "the self" as an entity on a par with other more ordinary entities such as organisms, rocks, and armchairs. The ontological inventory of the world as supported by third-personal science does not include any such thing as "the subject" or "the self". Yet, the intuition that our conscious experiences are somehow "centred" around us as subjects remains powerful.

Thirdly, and relatedly, if I ask the (I think reasonable) question of why I am having *my* conscious experiences rather than those of someone else, the one-world picture can't give me an answer. It

[11] My critique of the mainstream, one-world theories of consciousness echoes some earlier critiques in the literature. As noted, critiques of a "one-world" ontology, both in relation to the mind-body problem and more generally, can be found in Vacariu (2005) and Gabriel (2015). Further, Hellie (2013) stresses the difference between each person's own consciousness that is present to that person and the consciousness of others that lacks that presence. One-world theories arguably do not adequately capture that difference. Hellie, however, doesn't invoke "many worlds" in response. And Merlo (2016) argues that the mainstream theories don't adequately account for the unity of consciousness, the contents of self-awareness, and experiential knowledge, and defends a subjectivist view according to which "some propositions are true simpliciter without being true from all points of view" (318). One way to avoid some of the criticism summarized here might be to combine the one-world picture with the (implausible) solipsistic thesis that there is only a single conscious subject in the world ("myself"), around which the world is centred. This would be a version of what I will later briefly describe as a "one-centred-world picture". I will not pursue such a picture here (cf. Hare, 2007, 2009) but criticize the one-world picture under the background assumption that there can be more than one conscious subject.

[12] Among "one-world" theories, the ones best placed to respond to this criticism may be those monist theories that assert that conscious experiences have to do with intrinsic or categorical properties, while physical and functional phenomena have to do with extrinsic or relational properties (see appendix). But arguably, these theories still misidentify the "locus" of conscious experiences, by taking phenomenal facts, like physical facts, to hold at the world *simpliciter*.

[13] See Nagel (1986).

[14] One might object that if the world is third-personally such that X feels hungry, then we can easily get something first-personal out of this: it will follow that "I feel hungry" holds for X. But the fact that "'I feel hungry' holds for X" is still third-personal. And the genuinely first-personal fact "*I* feel hungry" doesn't hold *at the world simpliciter* as understood by the one-word picture. Later I will say that "*I* feel hungry" holds in X's first-personally centred world.

can't even point to any indexical fact in response to that question, such as the fact that I am who I am, because no such indexical fact holds *at the world simpliciter*; it holds only *for me*. The world as such, as depicted by the one-world picture, is not endowed with any "centre" at which I, as the subject, am located. Benj Hellie makes a similar point. He notes that the picture of consciousness defended by David Chalmers and others leaves an important question open. He begins by calling himself the "Hellie-subject" and asks why he, Benj Hellie, has the conscious experiences of the "Hellie-subject" rather than those of someone else, such as the "Chalmers-subject":

> "[A] vertiginous question is right around the corner. 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 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?"[15]

Again, the problem is that even if the world is populated not just by physical properties but also by phenomenal ones, the one-world picture remains an essentially third-personal picture.

Fourthly, the one-world picture is not particularly well placed to account for what is often described as the unity of consciousness. Why do some phenomenal properties jointly constitute a unified experience, while others, such as those associated with different subjects, are unconnected to one another? As Giovanni Merlo puts the question, "what makes certain mental states … coalesce into a single mental life: what is the 'glue' that keeps together my beliefs and my hopes, my desires and my fears, my feelings and my experiences?"[16] By taking the world to be populated by a host of properties, including all the phenomenal properties corresponding to my experiences and all those corresponding to yours and to everyone else's, the one-world picture has a hard time explaining *in structural terms* what makes each person's conscious experiences seem unified to that person: belonging to a single subjective perspective. What explains the experiential "bundling together" of some phenomenal properties, but not others?[17]

Fifthly, the one-world picture does not satisfactorily capture the way in which others' conscious experiences are first-personally inscrutable to each of us. By "first-person inscrutability" of others' experiences, I mean that although we can construct a *third-personal theory* of another person's mind and/or empathize with them *through our own first-person experiences*, we can't step into another subject's first-person perspective itself. This is an instance of the familiar problem of "other minds". Any evidence we have about the conscious experiences of others is third-personal

[15] See Hellie (2013, 309–310).

[16] See Merlo (2016, 333).

[17] One might try to explain the unity of consciousness in a one-world framework by pointing out that a unified consciousness serves the useful role of creating a "global workspace" in which information from multiple sources is integrated and made available for use by different cognitive processes (Baars, 1988, 2003). But such a third-personal, *functional* explanation of the unity of consciousness arguably doesn't explain the first-person *experience* of unity.

and thereby very much indirect. My attribution of conscious experiences to others always involves an inferential leap of faith.[18] The one-world picture doesn't really explain why this is so.

Finally, the one-world picture does not fully clarify what is distinctive about phenomenal properties and why the hard problem of consciousness is hard.[19] The debate about the hard problem – and relatedly about whether zombies are conceivable and/or metaphysically possible – has reached an impasse, and there is still no agreed diagnosis of what it would take to make progress in that debate.

These shortcomings of the one-world picture, I think, motivate the exploration of an alternative picture, even if one is not yet convinced that one should abandon the one-world picture in the end.

5 | AN ALTERNATIVE PICTURE

My starting point, as noted, is the observation that if we take the "locus" of conscious experiences to be "the world simpliciter", in some third-personal sense, then we fail to capture the perspectival, first-person, and subjective character of consciousness. The core idea of the alternative picture is the following:

> **Many worlds**: The "locus" of each subject's conscious experiences is not the world as such, in a third-personal sense, but a subject-specific "first-personally centred world". The first-personally centred worlds of different subjects can be viewed as distinct first-personal realizers of a shared "third-personal world". If the third-personal world admits more than one conscious subject, then there can be many first-personally centred worlds, which are all real and "parallel" to each other. For each subject, one such world is "present".

To explain this idea, let me begin with the notion of a "third-personal world". This can be defined as the totality of all facts that hold at that world from a third-person perspective. We can think of a "third-personal fact", in turn, as a fact that would feature in a complete description of the relevant world from the perspective of an omniscient Olympian observer studying the world "objectively" – the "view from nowhere" in Nagel's terms. The present definition is a version of Wittgenstein's famous dictum "[t]he world is everything that is the case", which he further clarifies by adding "[t]he world is the totality of facts, not of things".[20] Amending Wittgenstein's wording, we might say: "the third-personal world is everything that is the case third-personally", where "something that is the case third-personally" is a "third-personal fact".

For example, a third-personal world includes all the facts about all the physical entities and properties in that world and their configurations relative to one another, as well as all the facts

[18] Indeed, we cannot have certainty about the presence of conscious experiences in others at all. One might say that skepticism about others' conscious experiences is no harder to reject than skepticism about the external world in general. However, the two forms of skepticism are distinct. One could consistently accept the existence of the external world while denying the existence of others' consciousness. Our epistemic access to others' conscious experiences is arguably more indirect than our epistemic access to the external world. Notably, there can never be a single unified perspective from which different subjects' experiences are simultaneously first-personally accessible. For related points, see Merlo's discussion of "experiential knowledge" (2016, section 5.3).

[19] Again, perhaps the qualifications from note 12 apply.

[20] See Wittgenstein (1922).

that supervene upon those "physical" facts, where facts that supervene on third-personal facts are presumably themselves third-personal. If there are biological organisms such as humans in the world, then the third-personal facts will include all facts about their brains, bodies, and behaviour – even all facts about their psychology and cognition, to the extent that these can be described from a third-person perspective – as well as all the facts about their environments. All those facts would feature in a complete description of the relevant world from an Olympian perspective.

At the same time, a third-personal world does not include any "first-personal facts". By a "first-personal fact", I mean a fact that holds for me, from my first-person perspective. Examples are the fact that *I* am in a particular experiential state, the fact that *I* am having Christian's conscious experiences, or the fact that some object is present to me in a particular way. In his discussion of "first-personal realism", Kit Fine recognizes such facts. 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".[21] First-personal facts are left indeterminate by "the world as such", when this is understood third-personally. Rather, they hold from a particular subjective perspective. Their "locus", I will say, is a first-personally centred world.

First-personal facts, such as the fact that I am seeing an illuminated computer screen in front of me, must not be confused with certain corresponding third-personal facts, such as the fact that Christian is seeing such a screen. How exactly a subject's first-person perspective is related to a given third-personal world – say, where *I* fit into that world, or indeed whether there are any first-personal facts at all – is left open by the third-personal world. These questions go beyond third-personal facts.

David Lewis makes a structurally analogous point, albeit in relation to a subtly different issue:

> "Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude [with third-personal content], they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is."[22]

Although Lewis's topic here is indexicality and *de se* belief, not first-person experience, his quote illustrates the point that some facts, such as how *I* fit into the world or what *I* experience, are left open by a third-personal world. Even if I knew the totality of third-personal facts that hold at the world as depicted by the one-world picture, but lacked any first-personal information, my third-personal knowledge by itself would not allow me to infer my own first-person perspective on the world. The first-personal facts are under-determined by the third-personal ones, just as the indexical facts are under-determined by the non-indexical ones.

[21] See Fine (2005, 311). He offers a taxonomy of different forms of realism in relation to tense and/or the first person, also drawing on Prior (1968).

[22] See Lewis (1979, 520). Lewis uses examples such as this one to argue that the contents of an agent's beliefs cannot generally be captured by ordinary propositions that the agent takes to be true (where a proposition is a set of possible worlds), but must be expressed by self-ascribed properties (where a property is a set of individuals). For instance, to believe that one is located on mountain A rather than on mountain B is to self-ascribe the property whose extension is the set of all individuals (actual and possible) on mountain A. This is essentially equivalent to taking the content of any belief to be a set of centred worlds, as discussed below. For a recent discussion, see Jackson and Stoljar (2020).

To place a subject inside the world, we need to specify something above and beyond the third-personal world, namely a first-person perspective on it. I call this a "locus of subjectivity".[23] We can think of it as an additional ontological ingredient that is needed, in conjunction with the third-personal world, to determine all first-personal facts. Let me use the letter ω to denote the third-personal world ("omega" for "world") and the letter π to denote a locus of subjectivity ("pi" for "perspective"). We can then represent a "first-personally centred world" by an ordered pair $<\omega,\pi>$ consisting of a third-personal world ω and a locus of subjectivity π.[24]

Structurally, this notion is an instance of the standard notion of a "centred world" in the tradition of W.V. Quine and David Lewis: a world paired with some "location" or "centre".[25] However, centred worlds are usually interpreted in a thinner way than required for present purposes: centres are often understood simply as spatio-temporal coordinates, akin to the dot indicating your current location on your smartphone map, or alternatively as specific individuals in the world. Centres in this conventional sense – locational coordinates or pointers to who you are – may not be rich enough to fix a subject's full conscious perspective on the world; or even if they were, this would be a substantive claim that we shouldn't presuppose from the outset. More than one distinct stream of conscious experience might be compatible with occupying the same centre in the world, on a thin understanding of what a centre is. Plausibly, for example, the total facts about the third-personal world, together with my location pointer as picked up by my smartphone's global-positioning system, still under-determine my conscious experiences.[26]

I therefore interpret "centres" in a thicker way here, as indicated by the term "locus of subjectivity". I take any such "locus" to be whatever ingredient is needed to fix the relevant subject's first-person perspective on the world. This must be specified as richly as required in order to ensure that any first-personally centred world in which the given locus of subjectivity occurs – i.e., any pairing of it with a third-personal world – leaves no first-personal facts underspecified.

[23] The term also appears in Fine (2005), but not with the exact same meaning.

[24] As noted, this is structurally similar to Hare's (2007, 2009) and Honderich's (2014) notions of a "subject world" and a "subjective physical world", respectively, each of which may be understood as a world present to a conscious subject, though Honderich doesn't make the connection with centred worlds. For more on this structural similarity, see note 32. The notion of "first-personally centred worlds" may also be compared with Vacariu's (2005) notion of "epistemologically different worlds", in which the subject and the conditions of observation are key elements. However, the latter notion doesn't seem to be defined around the template of centred worlds. Moreover, the emphasis of Vacariu's theory appears to be more on the thesis that "[b]y changing the conditions of observation, the human observer can pass from one epistemological world to another" (537), where such a change may occur, for instance, if we shift from using our eyes to observe something to using an electron microscope that allows us to see its microphysical composition, and the thesis that "the mind and the brain belong to epistemologically different worlds" (515) than on the thesis – which would be more analogous to the one discussed here – that different conscious subjects, even under otherwise similar observational conditions, are associated with different worlds. Finally, Johnston (2007, 270) introduces the notion of an "arena of presence and action": a subject's "whole centered pattern of presence, existing at a particular time, and perhaps over time". But he adds, "[t]he implied center is just a virtual center… The world is not in fact centered in this way" (ibid.). For him, the appearance of subjectivity is due to the fact that things come in different modes of presentation of which we "sample" only some, though they are all objective. Crucially, my own interpretation of a first-personally centred world is ontic, not epistemic.

[25] See Quine (1969) and Lewis (1979); for a recent discussion, see Liao (2012).

[26] Chalmers (1996, 144) argues that a centred world, in the standard thin sense, is insufficient to capture a subject's phenomenal experiences. He writes: "indexicals accompany facts about conscious experience in their failure to supervene logically on physical facts, but they are all settled by the addition of a thin 'indexical fact' about the location of the agent in question. But even when we give [the agent] perfect knowledge about her indexical relation to everything in the physical world, her knowledge of [e.g.] red experiences will not be improved in the slightest. In lacking phenomenal knowledge, she lacks far more than someone lacking indexical knowledge." See also Chalmers and Jackson (2001). These observations underline the need to adopt a richer interpretation of a locus of subjectivity.

The facts to be fixed must include, in particular, all of the subject's phenomenal experiences at that world.

The ordered pair $<\omega,\pi>$ thus encodes the totality of facts that hold at the world ω with π placed inside it as the locus of subjectivity. Returning to Wittgenstein's characterization of "the world", we can say that "a subject's first-personally centred world is everything that is the case for that subject": everything that is the case at $<\omega,\pi>$. Crucially, this includes both

- the totality of third-personal facts which would feature in a complete and exhaustive third-personal description of the world that the relevant subject inhabits, and
- the totality of first-personal facts that hold for that subject: how the subject relates to the world, what the subject feels and experiences, and so on.

Those who like to think of worlds as representable by maximal consistent sets of sentences may think of a first-personally centred world as representable by a maximal consistent set of sentences expressible in third-person *and* first-person language. My own first-personally centred world would then be represented by the set of all first-personal and third-personal sentences that are true from where I stand. By contrast, a third-personal world would be representable by a maximal consistent set of sentences expressible in third-person language alone, and the actual third-personal world would be represented by the set of all true such sentences.

Importantly, the notion of a first-personally centred world is an ontic notion, not an epistemic one. My first-personally centred world $<\omega,\pi>$ includes

- not just those facts at $<\omega,\pi>$ of which I have knowledge or awareness,
- but all the facts that *hold* at $<\omega,\pi>$, which may include many facts of which I am oblivious.

This ontic rather than epistemic understanding of a first-personally centred world matters because the goal is to provide an account of what the ontology of the world must be in order to accommodate both objective and subjective features. This must subsume *everything that is the case, both third-personally and first-personally*. We are not merely providing an account of a subject's epistemic state. Facts that the subject knows or is aware of will be among the facts that hold relative to the subject, but they do not exhaust them.

On the present account, we must think of conscious experience not as something that is located at the third-personal world, but as something that occurs only at a first-personally centred world. To say that I am conscious, on this picture, is to say that some first-personally centred world is "present" for me. I am implying that I am conscious as soon as I acknowledge that some first-personally centred world is present for me, just as – in the more familiar third-personal case – we are implying that some third-personal truths hold once we accept that there is an actual third-personal world. Another way of making this point is to say that we should not treat phenomenal consciousness as a property whose mode and locus of instantiation are on a par with those of a physical property. Rather, I suggest – echoing some phenomenologists – that consciousness is tied to a first-personal "mode of being":[27] my phenomenal consciousness lies in there being a first-personally centred world that is present for me.[28]

[27] The notion of a "mode of being" was discussed by Heidegger (1927). Using this terminology, one might say: consciousness is not simply a property of a being, but rather a mode of being (or an aspect thereof). But I do not here commit myself to any further ideas from Heidegger's (controversial) philosophy.

[28] This is consistent with taking what Chalmers (1996) calls "awareness" to be instantiated at the third-personal world.

It is worth saying a little more about why this picture is best viewed as a *many-worlds* picture of consciousness, not as a *one-centred-world* picture. Unless we wish to accept a strong form of solipsism, we may reasonably assume that the same third-personal world can be paired with different loci of subjectivity which correspond to different conscious subjects. Suppose ω is the actual third-personal world and π and π' are two possible loci of subjectivity, which represent your subjective perspective and mine, respectively. Then the ordered pairs $<\omega,\pi>$ and $<\omega,\pi'>$ are each possible first-personally centred worlds, one of which is present for me, while the other is present for you.

There is no "first-personal world simpliciter" – one that we all share: you, I, and everyone else. As conscious subjects, we are experientially located in different and parallel first-personally centred worlds. Of course, our physical organisms and environment exist in a shared third-personal world. The third-personal facts instantiated at each of our first-personally centred worlds coincide. But the first-personal facts differ.

The best interpretation of this – and certainly the best *non-solipsistic* one – seems to be a "modal realist" interpretation. On this interpretation,

- there are many parallel first-personally centred worlds, all of which are real,
- but only one of them is present for each subject.

Thus your first-personally centred world is no less real than mine. It just so happens that my first-personally centred world is present for me, while yours is present for you. This picture is broadly analogous to David Lewis's realism about possible worlds, albeit applied to first-personally centred worlds, rather than third-personal ones. According to Lewis's own modal realism, all possible worlds – by which he meant uncentred, third-personal worlds – are real, even though only one world is actual; different possible worlds thus do not differ in their *reality* status; they differ only in their *actuality* status.[29] Analogously, on the present picture, different first-personally centred worlds do not differ in whether or not they are *real*, only in whether or not they are *present* for me.

Lewis considered modal realism a theoretically useful hypothesis in the philosophy of modality, suggesting that there are "many ways in which systematic philosophy goes more easily if we may presuppose modal realism in our analyses".[30] He wrote: "I take this to be a good reason to think that modal realism is true, just as the utility of set theory in mathematics is a good reason to believe that there are sets".[31] Regardless of whether we accept Lewis's modal realism with respect to uncentred worlds, embracing the proposed analogous position with respect to first-personally centred worlds seems natural if we are persuaded by a centred-worlds picture of consciousness but want to avoid solipsism. A salient criticism of Hare's "egocentric presentism", which – as noted – is a precursor of the many-worlds theory but lacks this modal realist commitment, is precisely its solipsistic flavour. Hare writes: "an egocentric presentist believes that only one subject world

[29] See Lewis (1986). A modal realist must hold that what is real goes beyond what is actual. So, "the world" and "reality" must not be taken to be synonymous: aside from the actual world, other possible worlds are real too, albeit separate and non-actual. Modal realism in general raises many further philosophical questions (see, e.g., Yagisawa, 2010), which I must set aside here. Suffice it to say that some solution strategies offered for certain challenges raised by standard modal realism in the case of third-personal worlds might carry over also to the case of first-personally centred worlds.

[30] See Lewis (1986, vii).

[31] Ibid.

exists. There are no other subject worlds."[32] By taking the view that other subjects' first-personally centred worlds are real, albeit not present for me, the many-worlds picture can avoid this. It can, on the one hand, point to what is special about my own conscious experience (its presence for me) but, on the other hand, retain a modal realist commitment to the existence of other subjects' conscious experiences too (insofar as those subjects' first-personally centred worlds are also real, in parallel to mine). So, just as Lewis considered his own modal realism about uncentred worlds theoretically useful in the philosophy of modality, a proponent of the many-worlds theory of consciousness may consider an analogous modal realism about first-personally centred worlds theoretically useful in the philosophy of consciousness.

6 | FIRST-PERSONAL FACTS AND INDEXICALITY

The totality of facts that hold at a first-personally centred world $<\omega,\pi>$ can be usefully partitioned into three categories:

> **Pure third-personal facts**: These are facts determined by the third-personal world ω alone. For any two distinct loci of subjectivity, π and π', that might be paired with ω, any such fact holds at $<\omega,\pi>$ if and only if it holds at $<\omega,\pi'>$. Pure third-personal facts are invariant under changes in the locus of subjectivity. In that sense, they are fully "objective".
>
> **Pure first-personal facts**: These are facts determined by the locus of subjectivity π alone. For any two distinct third-personal worlds, ω and ω', with which that same locus of subjectivity π might be paired (if any), any such fact holds at $<\omega,\pi>$ if and only if it holds at $<\omega',\pi>$. Pure first-personal facts are invariant under changes in the third-personal world. In that sense, they are fully "subjective".
>
> **Mixed facts**: These are facts determined only by the combination of the third-personal world ω and the locus of subjectivity π. Whether any such fact holds at the first-personally centred world $<\omega,\pi>$ depends not only on the third-personal world ω but also on the locus of subjectivity π. An example may be the fact that some feature of the world is present to me in a particular way or that I relate to the world in such-and-such a way – whether perceptually, attitudinally, or locationally.

The "mixed" category includes ordinary indexical facts. My mentioning of indexical facts might raise questions about the relationship between phenomenal facts and indexical facts. In particular, one might worry that, by characterizing first-personally centred worlds in analogy to centred worlds in the literature on indexicality and *de se* content, I conflate the problem of conscious experience with that of indexicality. After all, centred worlds – in the standard sense of a world paired

[32] See Hare (2009, 41). Hare defines the notion of a "system of subject worlds (*S*-worlds)". This is "a set of physically identical *S*-worlds such that for any functionally sentient creature in an *S*-world in the set, there is an *S*-world in the set in which that very creature has present experiences" (2007, 366–367). If we think of that set as containing all the distinct first-personally centred worlds that are deemed real on my modal realist picture, then Hare's theory is structurally similar to mine. Yet, Hare seems to reject a modal realist, "many worlds" interpretation. Honderich's (2014) understanding of "subjective physical worlds", another precursor, has more of a many-worlds flavour, but in tension with this, he describes "objective" and "subjective physical worlds" as "parts" of a single physical world (226).

with a "centre" or spatio-temporal location marker – are key formal tools to represent indexical content.[33]

My response is this. I accept that we should not think of phenomenal experience as being the same as indexicality. Not every indexical fact is plausibly a phenomenal fact. The fact that I am in Munich right now is indexical, but not by itself phenomenal. However, while not every indexical fact is a phenomenal one, phenomenal consciousness is arguably an indexical phenomenon. Conscious experiences do not occur third-personally, at the world simpliciter, but first-personally, for a particular subject, and thus at a world that is centred around a first-person perspective. So, it should be no surprise that the proposed framework for representing conscious experience is structurally similar to a standard framework for representing indexicality.

Furthermore, even ordinary indexical facts, such as the fact that I am in Munich right now, seem to require something like a subject for which they hold – perhaps a subject in a coarse-grained sense. The locus at which any indexical fact is instantiated is not the world simpliciter, but *the world together with some perspective on it*. Without any occupied perspective, no indexical fact could be instantiated. At most, we might say: such-and-such indexical fact *would hold if one were to occupy a particular perspective inside the world, such as a particular spatio-temporal perspective*. But no indexical fact would hold simpliciter – "monadically" and not just relative to some hypothetical centre that is actually unoccupied. A subject-free world – one not paired with any occupied centre – has no room for monadically instantiated indexical facts. By contrast, once I occupy the particular centre at which I am (i.e., I am in my suitably centred world), the fact that I am in Munich now holds monadically.[34]

I am open to the suggestion that consciousness is, in some sense, a special case of indexicality, namely, a particularly fine-grained and rich case.[35] First-personally centred worlds may be viewed as sufficiently enriched refinements of centred worlds as understood in the literature on indexicality. The structural parallel between the present account of consciousness and existing accounts of indexicality is therefore a feature of that account, not a bug.

7 | HOW THE MANY-WORLDS PICTURE IMPROVES UPON THE STANDARD PICTURE

Having completed my first sketch of the many-worlds picture, I would like to explain briefly how that picture avoids the identified shortcomings of the one-world picture.

My first criticism of the one-world picture was that, by taking all conscious subjects' experiences, just like all physical properties, to be features of one and the same world, the picture does not fully capture the perspectival, first-person, and subjective character of conscious experience. As should be clear, the many-worlds picture avoids this problem by associating each subject with

[33] See, e.g., Liao (2012) and Milano (2018). Note that indexicality can itself be represented in different ways. Consider the indexical "I am in Munich". One could locate its indexicality either in its *content* (by taking the content to be a centred proposition, i.e., a set of centred worlds, centred around an individual in Munich) or in the *mode* with which the content is represented (by taking the content to be an uncentred proposition, such as "Christian is in Munich", but taking that content to be represented in a special mode). The parallel between my treatment of consciousness and treatments of indexicality arises under the *content* (rather than *mode*) approach to indexicals.

[34] I have here benefitted from Hare's (2007, 2009) distinction between "monadic" and "relational" senses of presence. One might say: at a world that isn't paired with a centre, indexical facts can hold at most relationally (relative to some centre that isn't fixed by the world itself), but not monadically. At a centred world, indexical facts can hold monadically.

[35] I am grateful to Silvia Milano for a helpful discussion of this point.

a subject-specific first-personally centred world. It thus implies that the locus of consciousness is not the third-personal world, but a first-personally centred world, a world that is centred around a "locus of subjectivity". In this way, the many-worlds picture accommodates the perspectival, first-personal, and subjective character of consciousness.

My second criticism of the one-world picture was that it does not capture the centrality of the subject within any conscious experience. Moreover, I noted that the very notion of "the subject" has remained elusive in the analytic philosophy of mind. Since we have no good reason to regard "the subject" as an entity on a par with other entities in our ontology, it seems reasonable to avoid any reification of "the subject". But how do we account for the apparent "subject-centredness" of our conscious experience? The many-worlds picture can say something useful here. On the one hand, the picture suggests that we shouldn't think of first-person experiences as occurring at the (third-personal) "world simpliciter", and it does not treat "the subject" as an entity in that world. On the other hand, it still includes "the subject" as a building block of our ontology, via the concept of a "locus of subjectivity" that is a key constituent of a first-personally centred world. Crucially, however, a "locus of subjectivity" is best viewed, not as an entity, but as a "locus of being" – or as constituting a "locus of being" in conjunction with the third-personal world with which it is paired. In short, "the subject" is not an ordinary entity – either in the third-personal world or in the first-personally centred world – but a locus around which a first-personally centred world is centred.[36]

My third criticism of the one-world picture was that it lacks the resources to address the question of why I am having *my* conscious experiences rather than someone else's – Hellie's "vertiginous question". According to the one-world picture, there is no fact that holds at the world simpliciter which could settle that question. By contrast, the many-worlds picture can make recourse to a richer set of facts in answering it. Given the indexical nature of the question, any answer must point to a fact that isn't purely third-personal. If the locus of my conscious experiences is my first-personally centred world, then I am able to point to such a fact. There is a first-personal fact that holds at my first-personally centred world to the effect that I am having *my* conscious experiences rather than anyone else's.

My fourth criticism of the one-world picture was that it is not well placed to account for the unity of consciousness. Why do some phenomenal properties belong to a unified conscious perspective, while others belong to distinct perspectives? The many-worlds picture implies that what makes different first-personal facts belong to the same first-person perspective is that they hold at the same first-personally centred world. The unifying feature of all the phenomenal facts that constitute my conscious experiences is that they all hold at my first-personally centred world. The unifying feature of all the facts that constitute your conscious experiences is that they all hold at your first-personally centred world. To be sure, this does not settle all questions about the unity of consciousness. For instance, not every fact that holds at my first-personally centred world qualifies as a phenomenal fact. We may need to say more about which facts – among those that hold at my first-personally centred world – count as phenomenal facts. Even so, by locating conscious experiences at first-personally centred worlds, the present picture offers some structural resources for capturing the unity of consciousness.

[36] Relatedly, Fine (2005, 312) distinguishes between the "metaphysical self" and the "empirical self". He describes the former as "the implicit subject of the egocentric facts": "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 describes the latter as "the explicit subject of non-egocentric facts". Arguably, only the latter but not the former can be an entity of an ordinary sort in the world.

My fifth criticism of the one-world picture was that it does not satisfactorily explain why the conscious experiences of others are first-personally inscrutable to us. Why is it impossible to gain direct epistemic access to the first-person experiences of others? The many-worlds picture gives us a principled answer to this question. The conscious experiences of others are located at distinct first-personally centred worlds, which are not present to us. Only my own first-personally centred world is present to me, and the facts about others' conscious experiences are not located at that world. Any references that I am making to the conscious experiences of others are therefore references to certain "parallel" worlds, distinct from my own: namely the first-personally centred worlds of different subjects.

My final criticism of the one-world picture was that it doesn't give us a fully compelling diagnosis of why exactly the hard problem of consciousness is hard. I will return to this issue in Section 10.

I should emphasize that, just as my objections to the one-world picture weren't intended as definitive knock-down arguments but as motivating reasons for considering an alternative picture, so my brief explanations as to why the many-worlds picture avoids the identified problems should also be viewed as somewhat tentative. More work is needed on each of the issues raised.

8 | THE RELATIONSHIP BETWEEN THE THIRD-PERSONAL AND THE FIRST-PERSONAL FACTS

What does the present picture imply for the debate about what, if any, metaphysical dependence there is between consciousness and physical features of the world? How, in particular, should we think about the metaphysical relationship between third-personal facts and first-personally centred facts, both of the "pure" and of the "mixed" sorts distinguished earlier? My suggestion is that we can associate them with two different ontological levels. Crucially, the level of first-personally centred facts turns out to be subvenient and the level of third-personal facts supervenient, not the other way around. This is consistent with the rejection of physicalism by philosophers such as David Chalmers, but the many-worlds theory goes further, by rejecting the one-world picture too.

To explain this, I need to introduce the notion of "ontological levels". Colloquially, we often treat the world as being stratified into levels. For instance, we talk about the level of physics, the level of biology, the level of psychology, and so on. We can make this more precise by associating different levels with different classes of facts. The physical facts are distinct from the biological facts, which are distinct from the psychological facts, and so on. We can then recognize certain relationships of metaphysical dependence between different such classes of facts. For instance, the totality of physical facts is sufficient to determine the biological facts, and so we say that the biological level supervenes on the physical. (Sometimes this is alternatively spelt out in terms of "grounding", but I will here focus on supervenience.) Some levels stand in a supervenience relationship to each other, others not. To illustrate, the geological and biological levels are unrelated by supervenience – the facts at neither of these levels suffice to determine those at the other – but each supervenes on the physical.

Generally, we can think of a levelled ontology as encompassing a class of levels that is partially ordered by supervenience.[37] For each level, we can further introduce the notion of the "set of all

[37] Formally, a levelled ontology is an ordered pair <L,S>, where L and S are formal representations of the class of levels and the supervenience relations, respectively. I here draw on the approach in List (2019).

possible worlds at that level", where a "possible world at a particular level" is a possible specification of the totality of facts at that level. For instance:

- For the physical level, we introduce the set Ω_{phys} of all possible "physical-level worlds". Each element of Ω_{phys} represents a possible way the totality of physical facts could be.
- For the biological level, we introduce the set Ω_{bio} of all possible "biological-level worlds". Each element of Ω_{bio} represents a possible way the totality of biological facts could be.

We can use the sets Ω_{phys} and Ω_{bio} as formal representations of these two levels. Now, to say that the biological level supervenes on the physical is to say that there is a mapping from Ω_{phys} to Ω_{bio} which assigns to each physical-level world the biological-level world that supervenes on it. Supervenience mappings, which map worlds at the subvenient level to worlds at the supervenient level, are:

- "surjective", in the sense that, for each supervenient-level world, there is at least one subvenient-level world that is mapped to it, and
- (typically) "many-to-one", in the sense that more than one subvenient-level world may be mapped to the same supervenient-level world.

In the example of the physical and biological levels, "surjectivity" means that every biological-level world has at least one possible "physical realizer", where a "realizer" of the biological-level world is a physical-level world that gives rise to it. "Many-to-one" means that a biological-level world may be "multiply realizable" at the physical level: more than one physical-level world can give rise to the same biological-level world. Of course, proponents of a levelled ontology can still disagree about which levels there are, and which pairs of such levels are related by supervenience.

With these definitions in place, let me return to the many-worlds picture of consciousness. It is easy to see that we can associate third-personal facts and first-personally centred facts with two different ontological levels. Let Ω_{3rd} denote the set of all possible third-personal worlds, and let Ω_{1st} denote the set of all possible first-personally centred worlds. (Depending on the intended interpretation, "possible" could mean either "metaphysically possible" or "nomologically possible".) Now consider the mapping that maps each first-personally centred world $<\omega,\pi>$ to the corresponding third-personal world ω. This clearly has the properties of a supervenience mapping as just defined. And insofar as it maps from Ω_{1st} to Ω_{3rd}, the first-personally centred level qualifies as subvenient and the third-personal as supervenient. This should be unsurprising, since first-personally centred worlds include strictly more facts than third-personal worlds. Each first-personally centred world, by being an ordered pair of the form $<\omega,\pi>$, includes both third-personal and first-personal facts.

Indeed, we can think of any first-personally centred world $<\omega,\pi>$ as a "first-personal realizer" of the third-personal world ω. If, as I have assumed, different loci of subjectivity – π, π', π'', and so on – are compatible with the same third-personal world ω, we can think of the third-personal world ω as being "multiply realizable" at the first-personal level. On this picture, whenever a new conscious being comes into existence, such as a new-born human, there will be a new first-personal realizer of the third-personal world. Each new site of conscious experiences thus gives rise to a new way the third-personal world may be first-personally realized. A peculiar feature of this multiple realizability, entailed by the proposed modal realist interpretation, is that the different first-personally centred worlds that are "possible realizers" of the third-personal world ω are all real, even though only one of them is present for each subject.

Furthermore, as third-personal facts do not depend on the locus of subjectivity at all, we can treat them as holding at the third-personal level simpliciter, while first-personal facts – both "pure" and "mixed" – hold only at the first-personal level. Unlike in a traditional physicalist picture, the first-personal facts reside at a subvenient level relative to the third-personal facts. In that sense, phenomenal consciousness does not supervene on physical facts. I will now briefly comment on how we can situate scientific theories of consciousness in this two-level metaphysical structure.

9 | PSYCHO-PHYSICAL LAWS AND SCIENTIFIC THEORIES OF CONSCIOUSNESS

Scientific theories of consciousness, as proposed by neuroscientists and psychologists, are typically attempts to identify the neural or, more generally, physical correlates of consciousness. That is, they entail hypotheses to the effect that an organism or entity is phenomenally conscious (or in such-and-such experiential state) if and only if its brain, or cognition, or functional make-up, satisfies such-and-such conditions. When asked why I am conscious now while during general anaesthesia I am not, they will point to certain features of my brain state that explain the difference. Likewise, when asked why a dog is conscious and a washing machine is not, they will point to certain features present in the dog and absent in the washing machine. Technically, such theories hypothesize "psycho-physical laws": laws specifying which physical or biological conditions are such as to give rise to conscious experiences. How can we think about such laws in the present terms?

We begin by noting that since each first-personally centred world is formally an ordered pair $<\omega,\pi>$ consisting of a third-personal world ω and a locus of subjectivity π, the set of all possible first-personally centred worlds, Ω_{1st}, must either coincide with or be a subset of the set of all logically possible such pairs. Formally,

$$\Omega_{1st} \subseteq \Omega_{3rd} \times \Pi,$$

where Ω_{3rd} is the set of all possible third-personal worlds and Π is the universal set of all logically possible loci of subjectivity. Now, if we wish to interpret Ω_{1st} as the set of all *nomologically possible* worlds at the first-personal level, then Ω_{1st} is presumably smaller than the set $\Omega_{3rd} \times \Pi$ of all *logically possible* pairs of third-personal worlds from Ω_{3rd} and loci from Π. Not every logically possible locus of subjectivity needs to be compatible with every third-personal world in Ω_{3rd}. The third-personal facts may impose constraints on the first-person perspectives that can be paired with them. We can now think of a "psycho-physical law" as a specification of which pairs of the form $<\omega,\pi>$ are contained in Ω_{1st} and which are not. Different scientific theories of consciousness postulate different psycho-physical laws.

It may be, for instance, that a locus of subjectivity must be suitably associated with an entity with a particular consciousness-supporting make-up, such as a living organism with a normally functioning brain. Or perhaps, as panpsychists argue, loci of subjectivity are more ubiquitous: first-person perspectives could be attached to many other places in the world as well, beyond complex organisms like us. As I will now explain, we can express existing scientific theories of consciousness in this way, at least schematically, namely as specifications of which kinds of pairs $<\omega,\pi>$ are contained in the set Ω_{1st} and which not.

In the 1990s, for example, Francis Crick and Christof Koch influentially proposed a theory according to which phenomenal consciousness occurs in any biological brain that displays certain patterns of synchronized neural firing activity in a particular frequency range.[38] Others proposed that consciousness is associated with appropriate cognitive capacities, such as the presence of a global workspace or certain forms of higher-order cognition.[39] Each of these theories can be interpreted as a specification of which physical or functional properties must be instantiated in the third-personal world so as to give rise to a corresponding locus of subjectivity. Hence, we can think of them as specifications of which loci of subjectivity π can be associated with any given third-personal world ω. They thereby tell us which pairs of the form $<\omega,\pi>$ are nomologically possible, i.e., contained in Ω_{1st}.

The theory that arguably comes closest to fitting the suggested two-level (first-person / third-person) architecture is integrated information theory ("IIT"), as developed by Giulio Tononi and colleagues.[40] This theory offers an account of when a physical system gives rise to conscious experiences. It says that consciousness occurs in any physical system which instantiates a local maximum of "informational integration", where this is an information-theoretic property that can, in principle, be defined for any physical system. Translated into the present terminology, IIT asserts that a locus of subjectivity π is paired with a third-personal world ω (i.e., Ω_{1st} contains the pair $<\omega,\pi>$) if and only if π is appropriately associated with some site of locally maximal informational integration in ω. A functionally awake human cortex is an example of such a site, while an ecosystem or a fridge is (presumably) not.

IIT can be naturally situated within the metaphysical picture I have sketched. First of all, IIT's starting point, unlike that of most other scientific theories of consciousness, is phenomenological rather than physicalist. In presenting the theory, Tononi and colleagues begin by specifying so-called "axioms" that are intended to characterize the nature of first-person experience. They then seek to derive from those first-personal axioms some corresponding third-personal "postulates" concerning the consciousness-supporting physical conditions. This is where the claim that consciousness is associated with maximal informational integration comes into the picture. I do not take a stand on whether IIT's first-personal foundations or its derivation of the associated third-personal, information-theoretic correlates of consciousness are correct. All I want to note is that IIT is consistent with the metaphysical picture I have sketched, in which the third-personal level supervenes on the first-personally centred level. From the perspective of the present metaphysical picture, IIT's compatibility with it is a congenial feature of IIT.

10 | THE HARD PROBLEM REVISITED

To show how the many-worlds picture allows us to think about the hard problem of consciousness, let me return to the issue of zombies. Recall that a "zombie" is a hypothetical entity that is behaviourally and neurally indistinguishable from a conscious human being, but which lacks first-person experiences. A zombie has the same third-personal properties as its conscious counterpart: the same behaviour, bodily make-up, and brain functioning. From the outside, we would therefore be inclined to attribute the same psychological states and dispositions to it. Yet it lacks

[38] See Crick and Koch (1990).

[39] See, e.g., Baars (1988) and Carruthers (2016).

[40] See, e.g., Tononi (2015) and Tononi and Koch (2015).

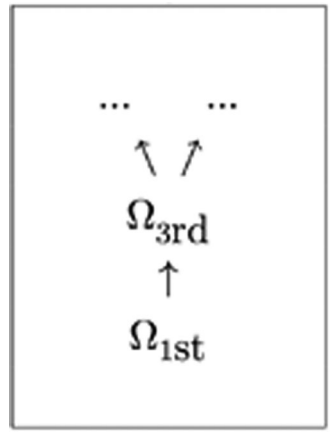

FIGURE 1 The standard scenario

phenomenal consciousness; there is nothing it is like to be such an entity. Importantly, no-one in the mainstream debate suggests that there are zombies in the actual world. Rather, the point of contention is whether the notion is coherent. Are zombies metaphysically possible? Could there be a duplicate of our world which is identical to the actual world in all physical respects, but in which no-one has any first-person experiences?

Let me use the term "zombie scenario" to refer to a scenario in which things are physically and third-personally indistinguishable from the actual world, but there is no first-personal consciousness. Is this scenario coherent? My analysis suggests that there is a sense in which it is coherent and another in which it isn't. Perhaps the existence of these two senses is one of the reasons why the hard problem of consciousness is so intractable.

Let me begin with the sense in which the zombie scenario is coherent. I have argued that we can represent consciousness and its relation to the third-personal world in terms of a levelled ontology in which there is a first-personally centred level in addition to the third-personal level. Thus, we have a levelled ontology in which there are (at least) two levels: the one corresponding to Ω_{3rd} and the one corresponding to Ω_{1st}. As noted, the first-personally centred level is subvenient, and the third-personal supervenient, not the other way around. In principle, there could also be other, higher levels that supervene on Ω_{3rd}, but this does not matter for present purposes. The structure is shown in Figure 1, with arrows indicating mappings from subvenient to supervenient levels. Now, the zombie scenario, where there are only zombies and no conscious beings, corresponds to a different levelled ontology, in which the lower one of the two levels – the first-personally centred one – is absent. In other words, the zombie scenario is represented by a levelled ontology that is truncated at the third-personal level, as shown in Figure 2. There is nothing perspectival, subjective, or first-personal that is subvenient here. Insofar as such a levelled ontology is coherent, the zombie scenario is also coherent.

But there is another sense in which the zombie scenario is not coherent. On the picture I have sketched, no possible *world* at any level – whether third-personal or first-personally centred – could be said to instantiate the zombie scenario. On the one hand, it is not meaningful to speak of "a world in which there are zombies" if by "world" we mean "third-personal world". Whether or not there are zombies depends, not on the features of any particular world in Ω_{3rd}, but rather on whether the third-personal worlds in Ω_{3rd} are "underwritten", or realized, by any first-personally centred worlds in Ω_{1st}. By definition, no features of a third-personal world could allow us to

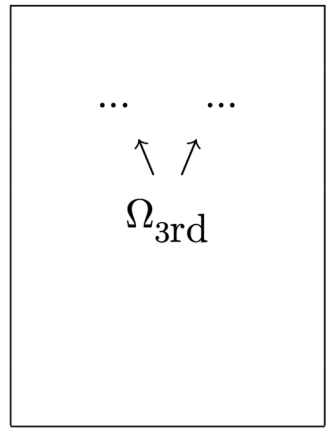

FIGURE 2 The zombie scenario

distinguish between zombies and non-zombies. Indeed, the third-personal worlds in the levelled ontology of Figure 1 are indistinguishable from those in the levelled ontology of Figure 2. So, if we focus on third-personal worlds alone, there is nothing that would justify calling them "zombie worlds" in Figure 2 but not in Figure 1. On the other hand, once we step inside a first-personally centred world, there is, by definition, a subject in that world: a first-personally centred world is a world with a subjective perspective. So, no such world could represent the zombie scenario either.

In sum, there are two ways in which we could interpret the question of whether the zombie scenario is coherent. We could either interpret it as asking whether there could be a *world* – whether third-personal or first-personal – in which there are zombies. Here, the answer must be "no". At the third-personal level, the distinction between zombies and non-zombies cannot be drawn, and at the first-personal level, there is necessarily a conscious subject, namely the one around which any given first-personal world is centred. Or we could interpret our question as asking whether there could be a *levelled ontology* in which there are zombies. Here, the answer is, in principle, "yes", insofar as the levelled ontology shown in Figure 2 is coherent. But despite its coherence, this levelled ontology is not the correct one for the predicament we find ourselves in. We have no reason to doubt the existence of consciousness: we are conscious subjects.

11 | CONCLUDING REMARKS

The aim of this paper has been to sketch the "many-worlds theory of consciousness". It combines three ingredients:

(i) the phenomenologically inspired idea that each conscious subject is associated with a first-personally centred world that is present for that subject;
(ii) the Lewis-inspired idea of modal realism about such first-personally centred worlds; and
(iii) an overall picture of a levelled ontology with a subvenient first-person level and a supervenient third-person level, where third-personal worlds are "multiply realizable" at the first-personal level and consciousness doesn't supervene on third-personal properties.

Although the resulting theory is somewhat heterodox, it stands in the tradition of other subjectivist and first-personal theories, and as noted, this work is not the first to criticize the traditional "one-world" ontology. Moreover, versions of each core ingredient (i) to (iii) can be found in earlier contributions to the philosophical literature. Something similar to the first ingredient, albeit with certain philosophical differences, can be found especially in Hare's work on "subject worlds" and "egocentric presentism" but also, without reference to centred worlds, in Honderich's work on "subjective physical worlds". Something similar to the second ingredient, though again in a different form, can be found in discussions of how consciousness fits into the many-worlds interpretation of quantum mechanics; the present picture, however, is not tied to quantum mechanics.[41] The third ingredient is perhaps least familiar, though the non-supervenience claim implied by it can be found in works on the explanatory and metaphysical gap between physical phenomena and conscious experience. Chalmers, for instance, has influentially argued that conscious experience does not supervene on physical properties, but he doesn't frame his argument in terms of a levelled ontology.[42]

I would like to close with a technical remark. We could also set up the present formal framework in a slightly different way. Instead of representing first-personally centred worlds explicitly as ordered pairs consisting of a third-personal world and a locus of subjectivity, we could take first-personally centred worlds as basic or primitive. We could then introduce two equivalence relations on the set Ω_{1st} of first-personally centred worlds. One partitions Ω_{1st} into equivalence classes of worlds that are third-personally equivalent. We could treat the set of such equivalence classes as the set of third-personal worlds, Ω_{3rd}. A second relation partitions Ω_{1st} into equivalence classes of worlds that are centred around the same subjective perspective. In effect, this would encode a criterion of personal identity. We could treat the set of such equivalence classes as the set of loci of subjectivity, Π. This yields a similar structure as before, but without defining first-personally centred worlds explicitly as pairs of the form $<\omega,\pi>$. Those phenomenologically oriented scholars for whom the first-personally centred level is the fundamental starting point might prefer this alternative setup.

[41] Chalmers (1996, ch. 10) suggests that his dualistic theory may be combined with Everett's "many-worlds" interpretation of quantum mechanics. He distinguishes between the "splitting-worlds" variant of the Everett view (a genuine "many-worlds" interpretation, which he rejects as a misinterpretation of Everett) and the "one-big-world" variant (which he prefers). According to the latter, "[t]here is only one world, but it has more in it than we might have thought" (1996, 347). Each conscious mind "perceives a separate discrete world, corresponding to the sort of world that we perceive – call this a miniworld, as opposed to the maxiworld of the superposition. The real world is a maxiworld, and the miniworlds are merely in the minds of the subjects" (ibid.). There would then still be a single world underlying all the different first-personal (mini)worlds. This differs from the many-worlds picture described in this paper. An Everett-inspired theory would also suggest that different conscious subjects correspond to the different constituent states within a quantum superposition. I need not assume this here. As noted, Honderich's characterization of each "subjective physical world" as one among many (2014, 192) also has a many-worlds flavour, but he describes subjective and objective physical worlds as "parts" of one physical world (226).

[42] There are also similarities with the "subjective physicalism" of Tim Crane (2003) and Robert Howell (2016). As Crane (2003, 78) notes, some well-known arguments against physicalism, such as Jackson's knowledge argument (1982), target "the view that all facts are … 'book-learning' facts: *facts the learning of which [does] not require you to have a certain kind of experience or occupy a certain position in the world*". Insofar as physicalism is committed to that view, those arguments speak against physicalism. Subjective physicalism, however, abandons the claim that all-facts are book-learning facts while retaining the claim that all facts – even those outside the book-learning category – are physical. Crane and Howell also discuss some parallels between indexical and phenomenal facts. While I agree with the cited observations about not all facts being book-learning facts, my analysis pushes me further away from physicalism. The subjective physicalist theory still upholds the one-world picture. For a helpful discussion of the knowledge argument, see also Nida-Rümelin (1995).

Although I have not offered a fully committed defense of the many-worlds theory of consciousness, I hope to have said enough to motivate further discussion of it.

ACKNOWLEDGEMENTS

The paper builds on Section 4.5 from the 2016 working paper version (http://philsci-archive.pitt.edu/12040/) of List (2019), which I did not include in the published paper due to space constraints. I presented the paper at a workshop on "Layers of Collective Intentionality", University of Vienna, August 2018. I am grateful to the participants and especially to Luke Roelofs, Glenda Satne, and Hans Bernhard Schmid for helpful feedback. I also thank an editor and a referee as well as Jonathan Birch, Zsuzsanna Chappell, Vincent Conitzer, Kristina Musholt, Marcus Pivato, and Daniel Stoljar for detailed written comments and Matteo Bianchin, David Chalmers, Dean Moyar, Silvia Milano, Claudia Passos-Ferreira, Ian Phillips, Robert Prentner, Jonathan Schaffer, Kai Spiekermann, and Laura Valentini for helpful conversations. Finally, I gratefully acknowledge my continuing affiliation with the Department of Philosophy, Logic and Scientific Method at the London School of Economics as a Visiting Professor.

Open Access funding enabled and organized by Projekt DEAL.

APPENDIX A: BRIEF OVERVIEW OF SOME STANDARD THEORIES OF CONSCIOUSNESS

We can distinguish between at least five different kinds of theories:[43]

Reductive or eliminative physicalist theories. Such theories deny that there is any gap between the physical and phenomenal realms. They assert that once we have explained all the physical and functional properties related to consciousness, we have explained everything there is to be explained. Phenomenal properties can be either eliminated from our ontology or reduced to physical or functional properties.

Non-reductive physicalist theories. Such theories accept that there is an explanatory gap – i.e., the explanation of consciousness requires more than the explanation of physical processes and functions – but deny that there is any further metaphysical gap: phenomenal properties still supervene on (and are grounded in) physical properties. According to those theories, the appearance of an explanatory gap is due to the fact that phenomenal properties, despite supervening on physical properties, are not "reducible" to them, so that explaining consciousness requires the use of concepts and categories distinct from those we use in physics or neuroscience. The irreducibility claim, in turn, can be spelt out in different ways. One common suggestion is that despite the metaphysical supervenience of phenomenal properties on physical ones, there is no relation of *a priori entailment* between the two. Another suggestion might be that phenomenal properties don't admit a *finite re-description* in physical-level terms.

Dualist theories. Such theories accept that there is both an explanatory and a metaphysical gap and assert that phenomenal properties do not supervene on (and are not grounded in) physical properties. On this picture, there is at most some weaker nomological (not metaphysically necessary) relationship between physical and phenomenal properties: relative to some contingent psycho-physical laws, phenomenal properties may depend on physical properties.

Idealist theories. Such theories are the mirror images of physicalist theories. They assert that there is a metaphysical dependence relationship between physical and phenomenal properties, but that physicalists have got its direction wrong, hence the appearance of an explanatory gap. It is physical properties that supervene on (and are grounded in) phenomenal ones, not the other

[43] This overview draws on Chalmers (1996, 2019).

way around. Idealist theories can also come in reductive and non-reductive forms. Non-reductive theories would assert that there is a "reverse explanatory gap": physical properties supervene on phenomenal properties but cannot be explained in terms of them. Reductive theories would deny this.

Dual-aspect or Russelian monist theories. Such theories assert that what we call "physical" and "phenomenal" properties are two different aspects of a single reality. There is a single class of fundamental properties grounding everything, including consciousness. One way to develop this idea, though perhaps not the only one, is to suggest that physics and the ordinary sciences only ever study relational, extrinsic, or structural properties while being silent on any underlying categorical, intrinsic, and non-structural properties, but that consciousness has to do with the latter. Further, since there could not be anything relational, extrinsic, or structural if there wasn't something categorical, intrinsic, and non-structural, the latter properties qualify as fundamental.

ORCID


Christian List https://orcid.org/0000-0003-1627-800X

REFERENCES

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. (2003). The global brainweb: An update on global workspace theory. Editorial, *Science and Consciousness Review*, 10/2003.

Carruthers, P. (2016). Higher-order theories of consciousness. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2016 Edition. https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219. https://doi.org/10.1093/acprof:oso/9780195311105.003.0001 (reprinted version)

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., pp. 1111–1120). Cambridge, MA: MIT Press. https://doi.org/10.1111/nyas.12166 (reprinted version)

Chalmers, D. (2019). Idealism and the mind-body problem. In W. Seager (Ed.), *The Routledge Handbook of Panpsychism* (pp. 353–373). London: Routledge. https://doi.org/10.4324/9781315717708

Chalmers, D., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *The Philosophical Review*, *110*, 315–360. https://doi.org/10.1215/00318108-110-3-315

Crane, T. (2003). Subjective facts. In H. Lillehammer & G. Rodriguez-Pereyra (Eds.), *Real metaphysics* (pp. 68–83). London: Routledge. https://doi.org/10.4324/9780203164297

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Fine, K. (2005). Tense and reality. In *Modality and tense: Philosophical papers* (pp. 261–320). Oxford: Oxford University Press. https://doi.org/10.1093/0199278709.003.0009

Gabriel, M. (2015). *Why the world does not exist*. Cambridge: Polity.

Gallagher, S., & Zahavi, D. (2019). Phenomenological approaches to self-consciousness. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2019 Edition. https://plato.stanford.edu/archives/sum2019/entries/self-consciousness-phenomenological/

Hare, C. (2007). Self-Bias, time-bias, and the metaphysics of self and time. *Journal of Philosophy*, *104*, 350–373. https://doi.org/10.5840/jphil2007104717

Hare, C. (2009). *On myself, and other, less important subjects*. Princeton: Princeton University Press. https://doi.org/10.1515/9781400830909

Heidegger, M. (1927). *Sein und Zeit*. Tübingen: Max Niemeyer.

Hellie, B. (2013). Against egalitarianism. *Analysis*, *73*, 304–320. https://doi.org/10.1093/analys/ans101

Honderich, T. (2014). *Actual consciousness*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198714385.001.0001

Howell, R. (2016). *Consciousness and the limits of objectivity: The case for subjective physicalism*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654666.001.0001
Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, *32*, 127–136. https://doi.org/10.2307/2960077
Jackson, F., & Stoljar, D. (2020). Understanding self-ascription. *Mind and Language*, *35*, 141–155. https://doi.org/10.1111/mila.12237
Johnston, M. (2007). Objective mind and the objectivity of our minds. *Philosophy and Phenomenological Research*, *75*, 233–268. https://doi.org/10.1111/j.1933-1592.2007.00075.x
Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x
Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*, 513–543. https://doi.org/10.2307/2184843
Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.
Liao, S. (2012). What are centered worlds? *The Philosophical Quarterly*, *62*, 294–316. https://doi.org/10.1111/j.1467-9213.2011.00042.x
List, C. (2019). Levels: Descriptive, explanatory, and ontological. *Noûs*, *53*, 852–883. https://doi.org/10.1111/nous.12241
Merlo, G. (2016). Subjectivism and the mental. *Dialectica*, *70*, 311–342. https://doi.org/10.1111/1746-8361.12153
Milano, S. (2018). De se beliefs and centred uncertainty. PhD thesis, London School of Economics. https://doi.org/10.21953/lse.b2wsi6xbghk6
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435–450. https://doi.org/10.2307/2183914
Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
Nida-Rümelin, M. (1995). What Mary couldn't know: Belief about phenomenal states. In T. Metzinger (Ed.), *Conscious experience* (pp. 219–241). Paderborn: Ferdinand Schöningh.
Prior, A. N. (1968). Egocentric logic. *Noûs*, *2*, 191–207. https://doi.org/10.2307/2214717
Quine, W. V. (1969). Propositional objects. In *Ontological Relativity and Other Essays* (pp. 139–160). New York: Columbia University Press.
Solomyak, O. (2020). Temporal ontology and the metaphysics of perspectives. *Erkenntnis*, *85*, 431–453. https://doi.org/10.1007/s10670-018-0034-4
Tononi, G. (2015). Integrated information theory. *Scholarpedia*, *10*, 4164. https://doi.org/10.4249/scholarpedia.4164
Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society, B*, *370*. https://doi.org/10.1098/rstb.2014.0167
Vacariu, G. (2005). Mind, brain, and epistemologically different worlds. *Synthese*, *147*, 515–548. https://doi.org/10.1007/s11229-005-8366-4
Wallace, D. (2012). *The emergent multiverse: Quantum theory according to the Everett interpretation*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199546961.001.0001
Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Kegan Paul.
Yagisawa, T. (2010). *Worlds and individuals, possible and otherwise*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199576890.001.0001
Zahavi, D. (2017). *Husserl's legacy: Phenomenology, metaphysics, and transcendental philosophy*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199684830.001.0001

How to cite this article: List C. The many-worlds theory of consciousness. *Noûs*. 2022; 1–25. https://doi.org/10.1111/nous.12408