The many-worlds theory of consciousness

Christian List*

Discussion paper, this version posted in March 2020

Abstract

This exploratory paper discusses a somewhat heterodox metaphysical theory of consciousness: the "many-worlds theory". The theory gives up the common assumption that all conscious experiences are features of one and the same world and asserts instead that different conscious subjects are associated with different "first-personally centred worlds". We can think of these as distinct and "parallel" first-personal realizers of a shared "third-personal world". 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 relationship between first-personally centred and third-personal worlds can in turn be captured in a levelled ontology, where the first-personal level is subvenient and the third-personal supervenient. The described setup is intended to capture the irreducibly subjective nature of conscious experience without lapsing into solipsism. The paper also looks at some existing scientific theories of consciousness, such as integrated information theory, through the lens of the present metaphysical theory and discusses its implications for the hard problem of consciousness.

1. Introduction

The aim of this paper is to explore a somewhat heterodox metaphysical theory of consciousness: the "many-worlds theory". The theory seeks to offer 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 gives up the common assumption that all conscious experiences – yours, mine, and those of everyone else – are features of one and the same world and asserts instead that different conscious subjects are associated with different "first-personally centred worlds". 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 theory is metaphysical insofar as it is concerned with the *metaphysical* relationship between consciousness and other features. It is thus relevant to the debate about which, if any, kind of metaphysical dependence there is between consciousness and physical properties, and whether consciousness can be accommodated in a naturalistic ontology. The theory does not address the *scientific* question of what the physical and/or neural correlates of consciousness are. That question is important too, and different answers to it may be compatible with the metaphysical picture to be sketched. I will briefly comment on how some existing scientific theories of consciousness, such as integrated information theory, may be framed in terms of the present metaphysical picture.

The theory is heterodox insofar as it differs from the mainstream theories in the analytic philosophy of consciousness. Although its motivating ideas are similar to those behind the classic critiques of physicalism given by Thomas Nagel, Frank Jackson, Joseph Levine, and David Chalmers,[1] I am aware of only few contributions with which it shares some core ideas, namely some works on

* This 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 it 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 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, Jonathan Schaffer, Kai Spiekermann, and Laura Valentini for helpful conversations.

[1] See, e.g., Nagel (1974), Jackson (1982), Levine (1983), and Chalmers (1996).

"presentism" or "subjectivism".[2] Caspar Hare and Ted Honderich, in particular, have introduced the notions of a "subject world" and a "subjective physical world", respectively, which, despite philosophical differences, are structurally similar to the notion of a "first-personally centred world". The present theory is also similar in spirit to some phenomenological theories in the tradition of Edmund Husserl.[3] But, in its current form, the theory does not seem to have been considered yet.

Crucially, while some earlier "presentist" theories, such as Hare's, invite a solipsistic interpretation by treating only a single subject's world as real, the many-worlds theory avoids this solipsism 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, even though only one of them is present for each subject, just as different uncentred worlds are real in David Lewis's modal realism, even though only one is actual.[4]

The paper is structured as follows. In Section 2, I will introduce the basic challenge of explaining how the first-person, subjective character of conscious experience fits into the rest of the world. In Section 3, I will review how analytic philosophers of mind typically frame the issue and point out that implicit in most existing theories is a "one-world picture". In Section 4, I will 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 and 6, I will sketch the alternative "many-worlds picture" and explain how it arguably avoids those shortcomings. In Section 7, I will discuss how what I call the "first-personal level" and the "third-personal level" are related to one another. In Section 8, I will revisit the "hard problem" of consciousness, and in Section 9, I will conclude.

---

[2] The closest precursors are the works by Caspar Hare (2007, 2009) on "egocentric presentism" and by Ted Honderich (2014) on "actualism". Hare (2007) 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*" (p. 366). In note 39, I further compare his theory with mine. 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" (p. 192), though he doesn't develop a centred-worlds architecture and instead conceives of both the "objective" and the "subjective physical worlds" as being "parts" of the physical world (p. 226). As Jacquette (2015) notes, "[t]he exact ontology of this remarkable relation is mentioned but not further explained by Honderich". Relatedly, Mark Johnston (2007) introduces the notion of an "arena of presence and action", defined as a subject's "whole centered pattern of presence, existing at a particular time, and perhaps over time" (p. 260). But he adds, "[t]he implied center is just a virtual center, a virtual point of convergence of perspectival modes of presentation. The world is not in fact centered in this way" (ibid.). Indeed, he suggests that there is no subjectivity at all: the appearance of subjectivity is due to the fact that things can come in different modes of presentation, but these are in a sense all objective, though we only "sample" some of them. Other related contributions are Kit Fine's discussion of a metaphysical view he calls "first-personal realism" (2005), Benj Hellie's "inegalitarian" approach to consciousness (2013), and Giovanni Merlo's "subjectivist view of the mental" (2016). Specifically, Fine (further citing Prior 1968) develops a taxonomy of different forms of realism in relation to tense and/or the first person. Hellie's "inegalitarianism" stresses the difference between each person's own consciousness that is directly present to him or her and the consciousness of others that lacks that presence. Merlo's subjectivism asserts that "some propositions are true simpliciter without being true from all points of view" (p. 318) but doesn't involve a many-worlds structure. My critique of the mainstream theories is similar to Hellie's and Merlo's critiques, but I develop a different response. For a discussion of what is at stake in the debate on presentism (mainly in temporal ontology), see further Solomyak (2018).

[3] For contemporary discussions, see Gallagher and Zahavi (2015) and Zahavi (2017). Developing these connections, however, is beyond the scope of this paper.

[4] See Lewis (1986).

This is a "broad strokes" paper, which will inevitably omit some qualifications, elaborations, and objections that some readers might like to see. There is a tradeoff between expositional simplicity and hedging of all the details, and my priority will be to give a sense of the many-worlds theory in a compact and accessible manner, rather than to offer a conclusive defence. My aim is to show that the theory is worthy of further consideration.[5]

2. The challenge

As conscious subjects, we are not merely biological organisms that function in certain ways and can be described from some external, third-person perspective, just as we describe the solar system or some other physical process. Rather, we ourselves 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 famously puts it (and Nagel emphasizes both "be" and "for").[6] By contrast, many entities and systems in the world, including some fairly complex ones, have no conscious experiences. The weather system, an eco-system, the global economy, and a smartphone are each systems of considerable complexity, and yet – for all we know – there is nothing it is like to be such a system. Everything that can be said about these systems can be said from the outside, from a third-person perspective. They have no "inner life": there is no first-person perspective attached to them.

How, then, can we study consciousness? David Chalmers describes the explanatory challenge as follows:

> "The task of a science of consciousness, as I see it, 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. When a conscious system is observed from the third-person point of view, a range of specific behavioral and neural phenomena present themselves. When a conscious system is observed from the first-person point of view, a range of specific subjective phenomena present themselves. Both sorts of phenomena have the status of data for a science of consciousness."[7]

And so, a science of consciousness must explain both. Crucially, the third-person data and the first-person data raise very different explanatory problems. The third-person data – the data concerning behavioural and neural phenomena – can in principle be explained using ordinary scientific methodology, such as in psychology and neuroscience. The data at issue here concern, for instance, a subject's wakefulness and sleep, observable attention, cognitive processing, and reasoning

---

[5] Here my expositional approach echoes that of Johnston (2007), who says (on p. 233): "In presenting the hypothesis, I shall not mince words; I shall write as if it is true, so that the exposition goes more smoothly. My original hope was that after the outrageous hypothesis had been set out, orthodoxy would be vindicated, for it then would be clear just why the hypothesis must be rejected. Well, it may be clear why the hypothesis should be rejected, but I must confess that it is not clear to me. So this is a plea for help: Here follows a hypothesis, help me to see just why it couldn't be so!"

[6] 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".

[7] See Chalmers (2004, p. 1111). My framing of the challenge throughout this section draws on Chalmers (1996).

capacities, as well as associated patterns of neural activity. We can study all of this in much the same way in which we study other phenomena in science. That is, we can formulate and empirically test hypotheses about the relevant phenomena from an observer's external perspective, where those hypotheses are expressed in third-person language, like all standard scientific hypotheses. David Chalmers calls the explanation of such phenomena the "easy problems of consciousness". "Easy" is meant to be a relative term. Chalmers's point is that the required explanatory strategy is not different *in principle* from the strategy we use to explain other complex phenomena in science.

By contrast, the first-person data – the data concerning subjective experience – raise a very different explanatory problem. As Thomas Nagel notes:

> "[The subjective character of experience] is not captured by any of the familiar ... reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved liked people though they experienced nothing."[8]

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 like 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. To be sure, nobody in the debate claims that there are any zombies in the actual world, and even those who think that zombies are metaphysically possible tend to concede that there could not be zombies in the world *as it is*. But the question of why this is so seems to require an answer. If a standard third-personal scientific approach cannot give that answer, then it is unclear whether such an approach can satisfactorily explain first-person experience at all. Indeed, if the phenomenon to be explained is irreducibly first-personal, a purely third-personal approach cannot even adequately *describe* the explanandum, let alone provide a satisfactory explanation. At best, it seems, 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.

There appears to be an "explanatory gap", as Joseph Levine puts it, between what the most extensive third-person scientific explanation can in principle deliver, and what we need to account for when we are trying to explain consciousness.[9] Another way of making this point is to say that consciousness involves not just functional properties, like other physical and biological phenomena, but also phenomenal ("experiential") properties and to argue that an ordinary scientific explanation, given from a third-person perspective, might account for the functional properties, but not for the phenomenal ones. The challenge is to explain how something as inherently first-personal as consciousness fits into a world that science usually describes in third-personal terms.

---

[8] See Nagel (1974, p. 436).
[9] See Levine (1983).

3. Some standard theories and the picture implicit in them

The standard metaphysical theories of consciousness in the analytic literature can be interpreted as attempts to answer the question of how physical and functional properties – roughly speaking, the properties studied in the sciences, ranging from physics to neuroscience and psychology – are related to phenomenal properties, which encode conscious experiences. Different theories can be characterized in terms of their positions on whether there is a gap between physical and phenomenal properties, and, if so, what the nature of this gap is.

At the risk of some oversimplification, we can distinguish between five different kinds of theories:[10]

> **Reductive or eliminative physicalist theories:** Such theories deny that there is any gap of the suggested sort. They assert that once we have explained all the physical and functional properties related to consciousness – something which falls under the rubric of the "easy problems" – we have explained everything there is to be explained about consciousness. On this account, phenomenal properties – if they exist at all and are not merely illusory – supervene on and are reducible to physical or functional properties.[11]
>
> **Non-reductive physicalist theories:** Such theories accept that there is an explanatory gap – in the sense that the explanation of consciousness requires more than the explanation of physical processes and functions – but deny that there is any further gap of a metaphysical kind. They assert that phenomenal properties supervene on (and are grounded in) physical properties. The appearance of an explanatory gap is due to the fact that phenomenal properties are distinct from and irreducible to physical properties, so that, despite the supervenience relationship, explaining consciousness requires the use of concepts and categories distinct from those we use in physics or neuroscience.[12]
>
> **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 relationship between physical and phenomenal properties, such as one of nomological (rather than metaphysical) supervenience: relative to some contingent psycho-physical laws, phenomenal properties may depend on physical properties.
>
> **Idealist theories:** Such theories can be viewed as 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,

---

[10] My brief overview of this theoretical landscape draws on Chalmers (1996, forthcoming).

[11] The difference between reductive and eliminative theories is that while the former merely claim that phenomenal properties are reducible to physical or functional properties, the latter eliminate them altogether from their ontology.

[12] There are different ways a non-reductive physicalist might spell out the nature of the gap in question. One suggestion would be that despite the metaphysical supervenience of phenomenal properties on physical ones, there is no relation of *a priori entailment* between the two. Another would be that phenomenal properties do not admit a *finite re-description* in physical-level terms. The details do not matter for the purposes of this paper.

hence the appearance of an explanatory gap. It is physical properties that supervene on (and are grounded in) phenomenal ones, not the other way around. Idealist theories can also come in reductive and non-reductive forms, but I set this distinction aside.[13]

Dual-aspect or Russelian monist theories: Such theories assert that what we conventionally call "physical" and "phenomenal" properties are two different aspects of a single reality. More precisely, there is a single class of fundamental properties grounding everything, including the "physical" and the "phenomenal" aspects of the world. One way to develop this idea, though 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, we can treat the latter properties as fundamental.

There is much to be said about all of these theories, and it is impossible to review this rich theoretical landscape in detail here. What I want to note, however, is that, despite their differences, the various theories all have one key presupposition in common. Implicit in all of them is what I call a "one-world" picture. That is, they all 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 within that world).

The different theories just disagree on which kinds of properties there must be in the world – which properties "populate" it – so as to accommodate conscious experiences along with physical phenomena. To illustrate this point, note that:

- According to physicalism, whether of a reductive or non-reductive sort, the world is fully specified by its physical properties. All other properties, if there are any, are supervenient on (or grounded in) physical properties.

- According to dualism, the world is fully specified not by its physical properties alone, but only by its combination of physical and phenomenal properties. Phenomenal properties do not supervene on (and are not grounded in) physical properties.

- According to idealism, the world is fully specified by its phenomenal properties. All other properties, if there are any, are supervenient on (or grounded in) them.

- According to other forms of monism, the world is fully specified by its fundamental properties of the relevant kind (and, if we take the idea of monism literally, there is just one such kind). Everything – both physical and phenomenal – is supervenient on (or grounded in) them.

[13] Indeed, one could imagine a non-reductive idealist theory according to which there is an explanatory gap of a reverse sort: physical properties supervene on phenomenal properties but cannot be explained in terms of them.

In slightly different terms, we can interpret these theories as giving us different answers to the question of which properties we would have to bring into existence in order to generate the world as it is. According to a physicalist theory, it would suffice to bring into existence the physical properties: they alone would be enough to generate everything there is in the world, including all conscious experiences. Something similar is true for other monist theories, according to which one would need to bring into existence only the properties of one fundamental kind in order to generate everything else. According to a dualist theory, by contrast, bringing into existence the properties of one kind alone would be insufficient: a hypothetical replica of our world that included only the physical properties would lack conscious beings.[14]

But despite these disagreements about which kinds of properties there are and how they depend on each other, the basic assumption that there is a single world in which the properties in question are to be found is not generally in dispute. 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 there are in that one world – that we may not just leave that picture unquestioned but fail to recognize it as an assumption in the first place.

However, for reasons that I will now explain, it is far from clear that the one-world picture does justice to the nature of conscious experience. I will give an overview of what I take to be the main such reasons and will then move on to a tentative alternative proposal: the many-worlds picture. I do not expect every reader to be equally persuaded by each of the critical remarks that I will make. They are not necessarily knock-down arguments against the one-world picture, but rather critical considerations that may motivate us to explore the alternative picture, even if we are not yet sure whether we should abandon the one-world picture. I would therefore ask readers to approach the following discussion with an open mind. What matters ultimately is not whether the one-world picture can be defended against some of the criticisms – it probably can – but rather whether the alternative picture to be sketched offers a more compelling overall theoretical package. This is the question that I would like to put on the table for debate.

4. What's wrong with the standard picture

So, what are the shortcomings of the one-world picture?[15] First, and most importantly, by taking all conscious subjects' experiences, just like all physical properties, to be features of one and the

[14] To illustrate how different theories can disagree about which properties populate the world, Chalmers (1996, pp. 127–128) gives an analogy from the history of science. Compare the ontology of classical Newtonian physics, prior to the discovery of electromagnetic fields, with that of Maxwell's electrodynamics. According to the former, the world includes only mechanical properties. According to the latter, it also includes electromagnetic properties, which do not supervene on mechanical properties. Maxwell's theory thus takes the world to contain more properties than Newton's. And so, if Maxwell's theory is right, replicating the mechanical properties alone is not enough to replicate everything there is in the world. Electromagnetism would be missing in a purely mechanical replica of the world. To replicate the world according to Maxwell, we would have to replicate both the mechanical properties and the electromagnetic ones.

[15] As noted, my critique of the mainstream approaches to consciousness echoes those offered by Hellie (2013) and Merlo (2016), although it is framed differently and my response will also be different. I agree that the sense in which the one-world picture is what Hellie calls "egalitarian" doesn't fully capture the nature of conscious experience. Merlo points out, relatedly, that the mainstream view does not adequately account for the unity of consciousness, the contents

same world, the one-world picture does not fully capture the inherently perspectival, first-person, and subjective character of conscious experience. The one-world picture lacks the resources to tell us *in structural terms* what the source of the subjectivity of conscious experiences is. When asked why some properties – so-called "physical" ones – correspond to objective features of the world while other properties – so-called "phenomenal" ones – correspond to subjective features, the one-world picture can only give a rather stipulative answer and assert that this is how it is.[16] In effect, the one-world picture is still a third-personal picture of the world: a picture of the world as it would be seen from an Olympian perspective, the "view from nowhere", as Nagel calls it.[17] 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.[18]

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 accepted that there isn't any good support for treating "the subject" or "the self" as an entity on a par with other more familiar entities such as organisms, rocks, and armchairs. The ontological inventory of the world as supported by 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 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 *relative to 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

---

of self-awareness, and experiential knowledge. A philosopher who has criticized the "one-world" picture of ontology more generally, albeit from a very different angle, is Gabriel (2015), who argues that the world as such does not exist.

[16] Among the theories of consciousness based on the one-world picture, the ones that are perhaps best placed to respond to this criticism are those 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. Even so, given the one-world assumption, those theories in effect imply that the "locus" at which facts of conscious experience hold is the same as the "locus" at which ordinary physical or functional facts hold: namely the world *simpliciter*.

[17] See Nagel (1986).

[18] One might object that if the world is third-personally such that Christian feels hungry, then we can easily get something first-personal out of this: it will follow that "I feel hungry" holds for Christian. But the fact that *I* feel hungry holds only relative to the first-personal fact that I am Christian. "I feel hungry" does not hold *at the world simpliciter*, without this subject relativization. Another way of making this point is to say that the third-personal proposition "Christian feels hungry" entails the first-personal proposition "I feel hungry" only in the presence of a first-personal premise, namely that "I am Christian". (This can be made precise by noting that the content of a third-personal proposition such as "Christian feels hungry" can be captured by a set of possible worlds simpliciter, while the content of a first-personal proposition such as "I feel hungry" needs to be captured by a set of centred worlds; more on this later.)

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 of course 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 rather than those of the Chalmers-subject be the ones relevant to what *I* feel?"[19]

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?"[20] 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.[21]

Fifthly, the one-world picture does not satisfactorily capture the way in which others' conscious experiences are first-personally inscrutable to 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". Indeed, as is generally acknowledged, we cannot have certainty about the presence of conscious experiences in others at all. My attribution of conscious experiences to others always involves an inferential leap of faith, even though the inference is a reasonable one.[22]

---

[19] See Hellie (2013, p. 309–310).

[20] See Merlo (2016, p. 333).

[21] 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). While this may be a *functional* explanation of the unity of consciousness from a third-person perspective, it arguably does not explain the first-person *experience* of unity. Conceivably, an information-processing system could play the functional role of pooling and integrating information from multiple sources so as to create a "global workspace" without giving rise to any unified experience. A zombie in Chalmers's sense could still have a "global workspace" in information-processing terms.

[22] One might say that skepticism about others' conscious experiences is no harder to reject than skepticism about the external world in general. Insofar as we usually reject the latter, we should also reject the former. However, the two

Finally, the one-world picture does not fully clarify what is distinctive about phenomenal properties and why the hard problem of consciousness is hard.[23] 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", then we fail to capture the perspectival, first-person, and subjective character of consciousness. The core idea of the alternative picture to be explored is the following:

> **Many worlds:** The "locus" of each subject's conscious experiences is not the world as such, in some third-personal sense, but a subject-specific "first-personally centred world". The first-personally centred worlds of different subjects can, in turn, be viewed as distinct and "parallel" first-personal realizers of a shared "third-personal world". If the third-personal world admits more than one conscious subject (as we usually assume), then there can be many first-personally centred worlds. For each subject, one of them is "actual" or "present".[24]

To explain this idea, let me begin with the notion of a third-personal world. A "third-personal world" can be defined as the totality of all facts that hold at that world from a third-personal perspective. We can think of a "third-personal fact" 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 from some "objective" perspective – 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".[25] 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 that

---

forms of skepticism are distinct, and one could consistently accept the existence of the external world while denying the existence of others' consciousness; this would be a form of solipsism. Many will intuitively feel that our epistemic access to others' conscious experiences is somehow 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).

[23] Again, perhaps the qualifications from note 16 apply.

[24] I should note that the idea of "many worlds" has previously been invoked in discussions of consciousness in the context of quantum mechanics, but usually in a very different way. I briefly return to this at the end of the paper.

[25] See Wittgenstein (1922).

supervene upon those "physical" facts. (I assume that any facts that supervene on third-personal facts are 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 presumably feature in a complete description of the relevant world from an Olympian perspective.

At the same time, a third-personal world does not determine any "first-personal facts". By a "first-personal fact", I mean a fact that holds only from a 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 rather than those of someone else, or the fact that some object is present to me in a particular way. In his discussion of "first-personal realism", Kit Fine also 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".[26] First-personal facts are left indeterminate by "the world as such" and are fully determined only relative to a particular subjective perspective. 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-personal perspective is related to a given third-personal world – say, where *I* fit into the 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."[27]

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-personal 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.

---

[26] See Fine (2005, p. 311).

[27] See Lewis (1979, p. 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 (2019).

To place a subject inside the world, we need to specify something above and beyond the third-personal world, namely a first-personal perspective on it. I call this a "locus of subjectivity".[28] We can think of it as an additional ontological ingredient needed in order 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 define a "first-personally centred world" as an ordered pair ⟨ω,π⟩ consisting of a third-personal world ω and a locus of subjectivity π.[29]

Formally, this definition is an instance of the standard definition of a "centred world": a world paired with some "location" or "centre".[30] 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 – basically, 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, at least 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.[31]

I therefore interpret "centres" in a thicker way here, as indicated by the term "locus of subjectivity". I take any such "locus" to encode a subject's entire first-person perspective on the world. This must be specified as richly as needed 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. The facts to be fixed must include, in particular, all of the relevant subject's phenomenal experiences at that world. The ordered pair ⟨ω,π⟩ 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" once again, we might say that "a subject's first-personally centred world is everything that is the case relative to that subject".

Crucially, this includes both

---

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

[29] As noted, this is structurally similar to Hare's (2007, 2009) notion of a "subject world". It also echoes Honderich's (2014) notion of a "subjective physical world", though Honderich doesn't make the connection with centred worlds.

[30] Centred worlds go back to Quine (1969) and Lewis (1979).

[31] Chalmers (1996, p. 144) argues that a centred world, in the standard thin sense, would be insufficient to capture a subject's full first-personal perspective, including his or her phenomenal experience. He writes: "indexicals [whose content may be represented by a set of centred worlds] 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 totality of third-personal facts that would feature in a complete and exhaustive third-personal description of the world the subject inhabits, and

- the totality of first-personal facts that hold relative to the subject: how the subject relates to the world, what the subject feels and experiences, and so on.

So, a first-personally centred world $\langle \omega, \pi \rangle$ encodes everything that is encoded by the third-personal world ω plus everything that is the case at ω relative to the locus of subjectivity π. Note that this notion is an ontic notion, not an epistemic one. My first-personally centred world $\langle \omega, \pi \rangle$ includes

- not just those facts of which I have knowledge or awareness,

- but all the facts that *hold* at $\langle \omega, \pi \rangle$, which may include facts of which I am oblivious.

This ontic rather than epistemic understanding of a first-personally centred world is important. After all, our 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 the features of the world the subject knows. Of course, facts about what the subject knows 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 "actual" or "present" for me. I am implying that I am conscious as soon as I acknowledge that some first-personally centred world is actual or present for me, just as – in the more familiar third-personal case – we are implying that some third-personal truths hold (at a minimum, tautological ones) once we accept that there is an actual world. Another way of making this point is to say that we should not treat consciousness as a property whose mode and locus of instantiation are on a par with those of a physical property. Rather, I would suggest – echoing some phenomenologists – that consciousness is tied to a first-personal "mode of being":[32] my consciousness is grounded in there being a first-personally centred world that is present or actual for me.

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

> **Pure third-personal facts:** These are facts settled by the third-personal world ω alone. This implies that, for any two distinct loci of subjectivity, π and π', that might be paired with ω, any such fact holds at $\langle \omega, \pi \rangle$ if and only if it holds at $\langle \omega, \pi' \rangle$. Pure third-personal facts are invariant under changes in the locus of subjectivity. In that sense, they can be said to be fully "objective".

[32] 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.

> **Pure first-personal facts:** These are facts settled by the locus of subjectivity π alone. This implies that, 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 ⟨ω,π⟩ if and only if it holds at ⟨ω',π⟩. Pure first-personal facts are invariant under changes in the third-personal world. In that sense, they can be said to be fully "subjective".
>
> **Mixed facts:** These are facts settled only by the combination of the third-personal world ω and the locus of subjectivity π. This implies that the question of whether or not any such fact holds at the first-personally centred world ⟨ω,π⟩ depends not only on the third-personal world ω but also on the locus of subjectivity π. An example may be the fact that the world, or something in the world, is present to me in a particular way or that I relate to the world in such-and-such a way – whether locationally, attitudinally, or perceptually. The present category of facts thus includes ordinary indexical facts.

My mentioning of indexical facts might invite the following objection. By defining a first-personally centred world in a way that is analogous to a centred world in the literature on indexicality and *de se* content, so the objection goes, I run the risk of conflating the problem of conscious experience with that of indexicality. After all, centred worlds – in the standard sense of an ordered pair of a world and a "centre" or spatio-temporal location marker – are key formal tools to represent indexical content.[33] My response to this objection is this. I accept that we should not think of phenomenal experience as being the same as indexicality. Not every indexical fact needs to be a phenomenal fact of the sort characteristic of conscious experience. The fact that I am in London right now is indexical, but arguably, it is not by itself a phenomenal fact. However, while not every indexical fact needs to be a phenomenal fact as conventionally construed, I would make two points. First, phenomenal consciousness is an indexical phenomenon. As I have emphasized, 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 formal framework needed to represent conscious experience is structurally similar to the formal framework needed to represent indexicality.[34]

Second, even ordinary indexical facts, such as the fact that I am in London right now, do not hold at the world simpliciter, *without some subject for whom those facts hold*. In that sense, indexicality presupposes a form of subjectivity. In a world without subjects, there could not be any genuine instantiated indexicality. Imagine a universe without any subjects. There can certainly be all sorts of third-personal facts in such a universe, but just as there would not be any phenomenal facts in the absence of any subjects for whom they hold, there could be no any indexical facts either. At most, we might say: such-and-such indexical fact *would hold if one were to occupy a particular*

---

[33] For recent discussions, see, e.g., Liao (2012) and Milano (2018).

[34] I should note that indexicality can itself be represented in different ways. Consider the indexical "I am in London". 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 London) or in the *mode* with which the content is represented (by taking the content to be an uncentred proposition, such as "Christian is in London", but taking that content to be represented in a special mode). The structural parallel between my treatment of consciousness and treatments of indexicality arises when one opts for the *content* approach to indexicals (rather than the *mode* approach).

place or time in that universe. But no such indexical fact would hold simpliciter – "monadically" and not just relative to some hypothetical centre that is actually unoccupied. It seems that a subject-free universe – one not paired with an occupied centre – has no room for monadically instantiated indexical facts. By contrast, the fact that I am in London right now holds monadically in my first-personally centred world.[35]

I am open to the suggestion that consciousness is, in some sense, a special case of indexicality.[36] First-personally centred worlds can be viewed as sufficiently enriched refinements of centred worlds as understood in the literature on indexicality.[37] The structural parallels between my account of consciousness and the established accounts of indexicality should therefore be a welcome feature of my approach, not a bug.

It is also worth saying a little more about why the present picture is best viewed as a *many-worlds* picture of consciousness, and not simply 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.[38] 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 $\langle\omega,\pi\rangle$ and $\langle\omega,\pi'\rangle$ are each possible first-personally centred worlds, one of which is actual or present for me, while the other is actual or present for you. This means that, in a sense, you and I "inhabit" different worlds, and there is no "first-personal world simpliciter" – one that we all share: you, I, and everyone else. Rather, as conscious subjects, we live, in effect, in parallel first-personally centred worlds. Of course, our physical organisms and all other features of our physical environment exist in a shared third-personal world. The third-personal facts instantiated at each of our first-personal worlds coincide. But the first-personal facts are different.

The best interpretation of all of this – and certainly the best *non-solipsistic* one – seems to be a many-worlds interpretation. On this interpretation, there are many parallel first-personally centred worlds, all of which are real in some appropriate modal sense, but only one of which is present for each subject. This picture is analogous to David Lewis's realism about possible worlds, albeit applied to the case of first-personally centred worlds, rather than third-personal ones. In the case

---

[35] 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.

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

[37] While the indexical fact that I am in London right now under-determines my conscious experiences, and the π-component of my first-personally centred world alone (without the ω-component) under-determines my relationship with the third-personal world, the entire first-personally centred world that is present for me – the pair $\langle\omega,\pi\rangle$ – does determine all indexical facts that hold for me, along with all other first-personal and third-personal facts. On this picture, a first-personally centred world determines not just all of the subject's conscious experiences, together with all third-personal facts about the surrounding environment, but also all indexical facts that hold for that subject.

[38] In principle, different loci of subjectivity could also be interpreted as encoding different perspectives that the same subject could have on the world at different points in time. My formal definitions do not rule out such an interpretation. But just as the third-personal world ω is best interpreted as being temporally extended, so each locus of subjectivity π may also be best interpreted as being temporally extended.

of first-personally centred worlds, this modal realism seems justified, provided we do not wish to accept solipsism.[39] (Whether a similar modal realism is also justified in relation to third-personal worlds is a separate question that I set aside here.)

This completes my first sketch of the many-worlds picture of consciousness. Of course, more needs to be said about the metaphysical relationship between third-personal worlds and first-personally centred worlds. Before I turn to this, however, I would like to explain briefly how the many-worlds picture avoids the problems I have identified with the one-world picture.

6. How the many-worlds picture avoids the identified problems

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 a subject-specific first-personally centred world. It thus implies that the locus of consciousness is not "the world simpliciter", 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 treat "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? I think the many-worlds picture can say something useful here. On the one hand, the picture implies that we should not treat consciousness as something that occurs at the 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

[39] See Lewis (1986). Honderich's (2014) understanding of "subjective physical worlds" has a many-worlds flavour, too, though he also suggests that "objective" and "subjective physical worlds" are "parts" of a single physical world. It is also worth comparing the present many-worlds picture with Hare's "egocentric presentist" theory. 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, pp. 366–367). If we think of that set as containing all the distinct first-personally centred worlds that are deemed real on the present modal realist picture, then Hare's theory is structurally similar to mine. Yet, Hare's theory is more solipsistic. He emphasizes that "an egocentric presentist believes that only one subject world exists. There are no other subject worlds" (2009, p. 41), thereby apparently rejecting a modal realist interpretation. He also argues that one should not take this solipsism too far. He writes: "egocentric presentism is quite neutral about what sorts of things there are. To say that the world is a subject world is just to say that, of the things there are, some of them are monadically present, and there is a sentient creature with all and only those things as perceptual objects. This in no way threatens the … idea that many people exist, and they are all equally real … People are not any more or less real, fleshy or concrete for having or failing to have monadically present perceptual objects" (2009, p. 42).

a "locus of being" in conjunction with the third-personal world with which it is paired. In short, "the subject" is not an 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.[40]

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 those of someone else: the "vertiginous question", as Benj Hellie calls it. 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 has no difficulties answering the vertiginous question. Given the indexical nature of the question, any answer to it must point to a fact that isn't purely third-personal. If the locus of my conscious experiences is indeed a 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-personal perspective is the fact 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 would qualify as a phenomenal fact experienced by me. 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.

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 – in some sense – references to certain "parallel" worlds, distinct from my own: namely the first-personally centred worlds of different subjects.

[40] Relatedly, Fine (2005, p. 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 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 8, when I revisit the hard problem of consciousness.

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 exploratory and tentative. There is no doubt that more work will need to be done on each of the issues raised.

7. The relationship between the third-personal and the first-personal levels

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 worlds and first-personally centred worlds? I want to suggest that we can associate third-personal worlds and first-personally centred worlds with two different ontological levels. Crucially, I will argue, the level of first-personally centred worlds is subvenient and the level of third-personal worlds supervenient – not the other way around. This is consistent with the rejection of physicalism by philosophers such as David Chalmers, but goes a step further, by rejecting the one-world picture too.

To explain this, I need to say more about the notion of "ontological levels". It is often said that the world is stratified into levels. For instance, we commonly distinguish between the level of physics, the level of chemistry, the level of biology, the level of psychology, and so on. In a levelled ontology, different levels can be associated with different classes of facts. The physical facts are distinct from the chemical facts, which are distinct from the biological facts, and so on. At the same time, we can recognize certain relationships of metaphysical dependence between different levels. For instance, the biological level plausibly supervenes on the physical, in that the totality of physical facts is sufficient to determine the biological facts. Some levels stand in such a relationship to each other, while others don't. To illustrate, the geological level and the biological level are probably unrelated to each other by supervenience, but each supervenes on the physical.

Generally, a levelled ontology can be represented by an ordered pair $\langle \mathcal{L}, \mathcal{S} \rangle$, where $\mathcal{L}$ is the class of all "levels" in that ontology (e.g., physical, chemical, biological, and so on) and $\mathcal{S}$ is the class of all "supervenience mappings" between levels.[41] Each such mapping has a "source level" and a "target level", and the existence of a mapping from one level (its source) to another (its target) means that the target level supervenes on the source.[42] So, the source level is the subvenient one, and the target level is the supervenient one. Of course, not all pairs of levels need to be related by supervenience.

[41] I have developed this formalism in List (2019).

[42] Supervenience is further assumed to be (i) reflexive, i.e., for each level in $\mathcal{L}$, there is a (trivial) supervenience mapping from that level to itself in $\mathcal{S}$; (ii) transitive, i.e., whenever $\mathcal{S}$ contains a mapping from one level to a second, and a mapping from the second level to a third, then it also contains a mapping from the first level to the third; and (iii) unique, i.e., $\mathcal{S}$ never contains more than one mapping from one given level to another given level.

The present formalism can be spelt out further by identifying each level with the set of all possible worlds at that level. A "possible world at a particular level" is simply a possible specification of the totality of facts at that level. For instance, the physical level can be identified with the set Ω_{phys}, where each element of Ω_{phys} represents one possible way the totality of physical facts could be. Similarly, the chemical level can be identified with the set Ω_{chem}, where each element of Ω_{chem} represents one possible way the totality of chemical facts could be. We can interpret the elements of Ω_{phys} as possible worlds at the level of physics, and we can interpret the elements of Ω_{chem} as possible worlds at the level of chemistry. Now, to say that the chemical level supervenes on the physical is simply to say that there is a mapping in $\mathcal{S}$ from Ω_{phys} to Ω_{chem} which assigns to each physical-level world the chemical-level world that supervenes on it.

Supervenience mappings are:

- "surjective", in the sense that, for each world at the supervenient (target) level, there is at least one world at the subvenient (source) level which is mapped to it, and

- (at least typically) "many-to-one", in the sense that more than one world at the subvenient (source) level may be mapped to the same world at the supervenient (target) level.

Applied to the example of the physical and chemical levels, the "surjectivity" condition captures the idea that every chemical-level world has at least one possible "physical realizer", where a "realizer" of the chemical-level world is a physical-level world that gives rise to it. And the "many-to-one" condition captures the idea that a chemical-level world may be "multiply realizable" at the physical level: more than configuration of the physical facts can give rise to the same body of chemical facts.

With these definitions in place, let me return to the many-worlds picture of consciousness. My claim is that we can associate third-personal worlds and first-personally centred worlds 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. As defined earlier, each third-personal world, of the form ω, is a full-specification of all third-personal facts, and each first-personally centred world, of the form $\langle\omega,\pi\rangle$, is a third-personal world paired with a locus of subjectivity. Depending on the intended interpretation, we can interpret the sets Ω_{3rd} and Ω_{1st} either as sets of metaphysically possible worlds of the relevant kinds or as sets of nomologically possible worlds. Suppose we take $\mathcal{L}$ to contain Ω_{3rd} and Ω_{1st}, and we take $\mathcal{S}$ to contain a mapping that assigns to each first-personally centred world $\langle\omega,\pi\rangle$ the third-personal world ω. Then the ordered pair $\langle\mathcal{L},\mathcal{S}\rangle$ is an instance of a levelled ontology as I have formally defined it.

Now, insofar as the relevant supervenience mapping has Ω_{1st} as its source level and Ω_{3rd} as its target level, the first-personal level qualifies as the subvenient one and the third-personal as the supervenient. This should not be surprising, given that first-personally centred worlds encode strictly more facts than third-personal worlds. Each first-personally centred world, by being an ordered pair of the form $\langle\omega,\pi\rangle$, encodes both third-personal and first-personal facts. In fact, we can think of any first-personally centred world $\langle\omega,\pi\rangle$ 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 – can be 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. Indeed, on this picture, we may say that whenever a new conscious being comes into existence, such as a new-born human, this means that 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.

Note, further, that as third-personal facts do not depend on the locus of subjectivity at all, we can treat third-personal facts as holding at the third-personal level simpliciter, while first-personal facts (both of the "pure" and of the "mixed" sorts introduced earlier) hold only at the first-personal level. In that sense, third-personal and first-personal facts may be said to "reside" at different levels.

In addition to recognizing that there is a supervenience relationship between the first-personal and third-personal levels, we can also make sense of the idea of a psycho-physical law within this framework. Given that each first-personally centred world is an ordered pair $\langle \omega,\pi \rangle$ consisting of a third-personal world ω and a locus of subjectivity π, the set of all possible first-personally centred worlds, Ω_{1st}, should either coincide with, or be some subset of, the set of all logically possible such pairs, formally, the product set $\Omega_{3rd} \times \Pi$, where Ω_{3rd} is the set of all possible third-personal worlds and Π is the universal set of all possible loci of subjectivity. However, if we wish to interpret Ω_{1st} as the set of all genuinely possible worlds at the first-personal level – in either a metaphysical or a nomological sense – then Ω_{1st} may well be smaller than the product set $\Omega_{3rd} \times \Pi$. Not every possible locus of subjectivity needs to be compatible with every possible third-personal world. In particular, we may hold different philosophical and/or scientific views about how the third-personal facts encoded by a third-personal world ω constrain the possible first-personal perspectives that may be paired with ω. Generally, we can define a "psycho-physical law" as a specification of which pairs of the form $\langle \omega,\pi \rangle$ are included in Ω_{1st} and which are not. Different theories of consciousness may give us different such 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, one can re-interpret existing scientific theories of consciousness as specifications of which kinds of pairs $\langle \omega,\pi \rangle$ are included in the set Ω_{1st} and which not. Theories of consciousness proposed by neuroscientists and psychologists are often attempts to identify the neural or, more generally, physical correlates of consciousness. This is to say: they entail hypotheses to the effect that an organism or entity is phenomenally conscious if and only if its brain, or cognition, or functional make-up, satisfies such and such conditions. In the early 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.[43] Others proposed that consciousness is associated with appropriate cognitive capacities, such as certain forms of higher-order cognition.[44] Each of these theories can be interpreted in present terms as a specification of which physically instantiated properties within the third-personal world are such that they give rise to a corresponding locus of subjectivity. Hence, we can think of them as specifications of which kinds of loci of consciousness π can be associated with any given third-personal world ω.

The theory that arguably comes closest to fitting the two-level (first-person / third-person) structure for the analysis of consciousness sketched in this paper is integrated information theory ("IIT"), as developed by Giulio Tononi and colleagues.[45] This theory, too, 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 $\langle\omega,\pi\rangle$) 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 subsumed under 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 associated 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 wish to commit myself to any claims about 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 this picture is consistent with the metaphysical architecture I have sketched, in which the third-personal level supervenes on the first-personally centred level. If the metaphysical picture presented here turned out to be right, then IIT's compatibility with it would be a good-making feature of that theory.

8. Some lessons for the hard problem of consciousness

To show how the many-worlds picture allows us to think about the hard problem of consciousness, let me return to the much-debated question of whether zombies are metaphysically possible. As already noted, a "zombie" is a hypothetical entity that is behaviourally and neurally indistinguishable from an ordinary human being like you and me, but which lacks first-person experiences. A zombie has the same third-personal properties as its conscious counterpart: its bodily make-up is the same, as is its brain-functioning; it behaves in the same way as you or I do.

[43] See Crick and Koch (1990).
[44] For a review and critical discussion, see Carruthers (2011).
[45] See, e.g., Tononi (2015) and Tononi and Koch (2015).

From the outside, we would be inclined to attribute the same psychological states and dispositions to it. Yet it lacks phenomenal consciousness. There is nothing it is like to be such an entity, as Thomas Nagel would put it. As I have also noted, no-one in the mainstream debate suggests that there are zombies in the real world. Rather, the debate concerns the question of whether the notion of a zombie is coherent. Are zombies metaphysically possible or not? In particular, could there be a possible duplicate of our world which is identical to the actual world in all physical respects, but in which no-one has any first-person experience?

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 sense in which it isn't. Perhaps the existence of these two senses is, in part, responsible for the hardness of the hard problem of consciousness.

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 rest of the 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}. Importantly, in this ontology, the first-personally centred level is subvenient, and the third-personal level 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, I suggest that 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 one in which the levelled ontology 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.

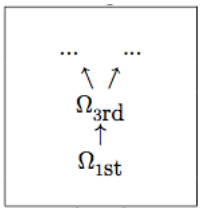

Figure 1: The standard scenario

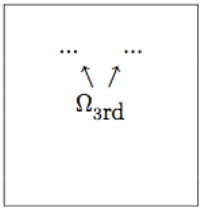

Figure 2: The zombie scenario

But there is another sense in which the zombie scenario is not coherent. On the picture I have sketched, it is clearly 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 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 allow us to call them "zombie worlds" in Figure 2 but not in Figure 1.

On the other hand, once we step inside any 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 properties of such a world could mark the distinction between zombies and non-zombies either. For this reason, the debate about whether there could be zombies is best interpreted, not as a debate about what properties there are in a given level-specific world, but rather as a debate about which levels there are in a levelled ontology; and the question of how phenomenal consciousness relates to the rest of our ontology is really a question about the ontological status of the first-personal level, assuming there is such a level.

In sum, there are two distinct ways in which we could understand 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, around which the first-personal worlds are centred. Or we could alternatively 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 clearly not the correct one for the predicament in which we find ourselves. We have no reason to doubt the existence of consciousness: we are conscious subjects. Of course, why this is so requires an answer, and this answer must ultimately be given by an appropriate psycho-physical law along the lines discussed in the previous section.[46]

---

[46] Philosophers of consciousness sometimes discuss the notion of a "ghost" as the conceptual opposite of a "zombie": a hypothetical entity with phenomenal properties but no physical ones. Though no-one in the debate suggests that ghosts are *nomologically* possible, dualist theories render them *metaphysically* possible, so one might wonder what the present theory has to say about this issue. Here is one way one might approach the issue in analogy with my analysis of zombies. Let me use the term "ghost scenario" to refer to a scenario in which things are phenomenally indistinguishable from how they are in our present first-personally centred worlds, but there is nothing physical. Insofar as each first-personally centred world is a pair of the form $\langle\omega,\pi\rangle$, and the set Ω_{1st} is equal to or included in $\Omega_{3rd} \times \Pi$, we may consider a levelled ontology that contains Ω_{1st} as a subvenient level and Π as a supervenient one (among other supervenient levels, such as Ω_{3rd}), where the mapping from Ω_{1st} to Π maps each first-personally centred world $\langle\omega,\pi\rangle$ to its π-component alone, interpretable as a "pure phenomenal world". The latter is the totality of facts settled by the locus of subjectivity π alone. Figure 3 shows this levelled ontology. One might argue that the ghost scenario, in which there is nothing physical at all, corresponds to the levelled ontology shown in Figure 4, which is truncated at the "pure phenomenal level" Π, where the worlds are of the form π rather than $\langle\omega,\pi\rangle$, and in which the richer first-personally *centred* level Ω_{1st} is absent.

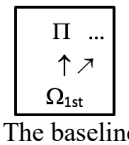

Figure 3: The baseline scenario

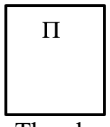

Figure 4: The ghost scenario

I do not claim that such a levelled ontology is plausible, but if it is coherent, one might argue that the ghost scenario is coherent too. I thank Daniel Stoljar for prompting me to address this issue.

9. Concluding remarks

The aim of this paper has been to explore what I have called 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 him or her

with (ii) the Lewis-inspired idea of modal realism about such first-personally centred worlds,

while (iii) representing the overall picture as 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 theory resulting from these three ingredients may be heterodox, versions of each ingredient can be found elsewhere in the philosophical debate. As already noted, something similar to the first ingredient – though developed differently – can be found in Caspar Hare's work on "egocentric presentism" and, without reference to centred worlds, in Ted Honderich's work on "actualism". Something similar to the second – though not in exactly the form proposed here – can be found in discussions of how consciousness fits into the many-worlds interpretation of quantum mechanics.[47] The present many-worlds picture, however, is not tied to quantum mechanics. The many-world picture also fits Honderich's characterization of each "subjective physical world" as being one "among very many, as many as there are sets of perceivings of single perceivers", though, contrary to the picture I have sketched, Honderich thinks of subjective and objective physical worlds as being different *parts* of the physical world.[48] The third ingredient is perhaps the 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. David Chalmers, for instance, has influentially defended the view that conscious experience does not supervene on physical properties, but he doesn't frame his argument in terms of a levelled ontology. That said, the proposed combination of the three ingredients is probably new or at least underexplored in the existing debate.[49]

---

[47] For instance, 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" (p. 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 I have sketched here. Moreover, an Everett-inspired theory would seem to suggest that different conscious subjects correspond to the different constituent states within a quantum superposition. Although I do not rule out that the Everett interpretation of quantum mechanics may be defensible, I need not tie "many worlds" to quantum superpositions here.

[48] See Honderich (2014, pp. 192 and 226).

[49] As noted, the present theory also shares some ideas with Fine's first-personal realism and Merlo's subjectivism about the mental, but those other theories are not developed around an explicit centred-worlds architecture. There are

Finally, I would like to make a couple of technical remarks. One is that the present picture offers some resources for thinking about certain kinds of counteridentical conditionals such as "If I were such-and-such person, I would experience, feel, or do such-and-such", especially in the context of empathizing with and relating to others. Counteridentical conditionals raise a number of difficulties and are sometimes interpreted as instances of counterpossible conditionals – conditionals with a necessarily false antecedent. According to the standard possible-worlds semantics, such conditionals would be vacuously true, which is an undesirable result. The present theory gives us a possible strategy for analyzing them. We might say – at least as a first simplistic stab – that the conditional "If I were such-and-such person, I would experience, feel, or do such-and-such" is true at a given first-personally centred world if and only if its consequent is true at all nearest first-personally centred worlds at which the antecedent is true. In the simplest case, truth-evaluating such a conditional would require shifting the locus of subjectivity from the one at my present first-personally centred world to the locus corresponding to the target subject, and then considering whether the consequent is true at that world. Of course, there are many potential subtleties and complications to which I cannot do justice here. For instance, while projecting myself into someone else's perspective, I might assume that I still retain some of my own psychological or phenomenal characteristics. This would go against defining the nearest first-personally centred world in which the antecedent is true simply as the one where my current locus of subjectivity is replaced by that of the target subject while the third-personal world remains the same. Rather, I would somehow need to shift my current locus of subjectivity to a hypothetical "hybrid" locus that is supposed to be suitably co-located with the target subject while retaining some aspects of my own subjectivity; in addition, some other features of the third-personal world might need to be adjusted too. In any case, I wish to note that the present framework gives us possible resources for thinking about at least some counteridentical conditionals in the context of relating to other minds.[50]

The second technical remark I would like to make is that we could also set up the present formal framework in a slightly different way. Instead of defining first-personally centred worlds explicitly as pairs consisting of a third-personal world and a locus of subjectivity, as I have done, we could take first-personally centred worlds to be 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 subject. In effect, this would encode a criterion of

---

also some similarities between the theory sketched here and the "subjective physicalism" of Tim Crane (2003) and Robert Howell (2016). As Crane notes, some well-known arguments against physicalism, such as Frank 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*" (p. 78). 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 quoted observations about book-learning facts, my analysis, as explained, pushes me further away from physicalism. The subjective physicalist theory still upholds the one-world picture.

[50] For recent studies of counteridenticals, see Kocurek (2018) and Kauf (2018); the latter invokes centred worlds.

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 ⟨ω,π⟩. Those phenomenologically oriented scholars for whom the first-personally centred level is the fundamental starting point might prefer this alternative setup.

Although my discussion has been tentative and exploratory, and I have certainly not offered a conclusive defense of the many-worlds theory of consciousness, I hope to have said enough to motivate further discussion of it.

References

Baars, Bernard J. (1988) *A Cognitive Theory of Consciousness*. Cambridge (Cambridge University Press).

Baars, Bernard J. (2003) "The Global Brainweb: An Update on Global Workspace Theory." Guest editorial, *Science and Consciousness Review*, October 2003.

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

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

Chalmers, David (forthcoming) "Idealism and the Mind-Body Problem." In William Seager (ed.), *The Routledge Handbook of Panpsychism*, London (Routledge).

Chalmers, David, and Frank Jackson (2001) "Conceptual Analysis and Reductive Explanation." *The Philosophical Review* 110(3): 315–360.

Crane, Tim (2003) "Subjective facts." In Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra (eds.), *Real Metaphysics*, London (Routledge), pp. 68–83.

Crick, Francis, and Christof Koch (1990) "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2: 263–275.

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

Gabriel, Markus (2015) *Why the World Does Not Exist*. Cambridge (Polity Press).

Gallagher, Shaun and Dan Zahavi (2015) "Phenomenological Approaches to Self-Consciousness." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition) <http://plato.stanford.edu/archives/spr2015/entries/self-consciousness-phenomenological/>.

Hare, Caspar (2007) "Self-Bias, Time-Bias, and the Metaphysics of Self and Time." *Journal of Philosophy* 104(7): 350–373.

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

Heidegger, Martin (1927) *Sein und Zeit*. Tübingen (Max Niemeyer Verlag).

Hellie, Benj (2013) "Against Egalitarianism." *Analysis* 73(2): 304–320.

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

Howell, Robert (2016) *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*. Oxford (Oxford University Press).

Jackson, Frank (1982) "Epiphenomenal qualia." *Philosophical Quarterly* 32(127): 127–136.

Jackson, Frank, and Daniel Stoljar (2019) "Understanding self-ascription." *Mind and Language*, online early.

Jacquette, Dale (2015) "Review of *Actual Consciousness*." *Notre Dame Philosophical Reviews* 2015.08.35.

Johnston, Mark (2007) "Objective Mind and the Objectivity of Our Minds." *Philosophy and Phenomenological Research* 75(2): 233–268.

Kauf, Carina (2018) "Counteridenticals and dream reports: A unified analysis." *Proceedings of Sinn und Bedeutung* 22(2): 1–18.

Kocurek, Alexander W. (2018) "Counteridenticals." *The Philosophical Review* 127(3): 323–369.

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

Lewis, David (1979) "Attitudes De Dicto and De Se." *The Philosophical Review* 88(4): 513–543.

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

Liao, Shen-yi (2012) "What are centered worlds?" *The Philosophical Quarterly* 62(247): 294–316.

List, Christian (2019) "Levels: Descriptive, Explanatory, and Ontological." *Noûs* 53(4): 852–883.

Merlo, Giovanni (2016) "Subjectivism and the Mental." *Dialectica* 70(3): 311–342.

Milano, Silvia (2018) *De se beliefs and centred uncertainty*. PhD thesis, London School of Economics <http://etheses.lse.ac.uk/3735/>.

Nagel, Thomas (1974) "What Is It Like to Be a Bat?" *Philosophical Review* 83(4): 435–450.

Nagel, Thomas (1986) *The View from Nowhere*. Oxford (Oxford University Press).

Prior, Arthur N. (1968) "Egocentric Logic." *Noûs* 2(3): 191–207.

Quine, W.V. (1969) "Propositional objects." In *Ontological Relativity and Other Essays*, New York (Columbia University Press), pp. 139–160.

Solomyak, Olla (2018) "Temporal Ontology and the Metaphysics of Perspectives", *Erkenntnis*, online early.

Tononi, Giulio (2015) "Integrated Information Theory." *Scholarpedia* 10(1): 4164.

Tononi, Giulio, and Christof Koch (2015) "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society* B 370.

Wittgenstein, Ludwig (1922) *Tractatus Logico-Philosophicus*. London (Kegan Paul).

Zahavi, Dan (2007) *Husserl's Legacy: Phenomenology, Metaphysics, and Transcendental Philosophy*. Oxford (Oxford University Press).