

Four Views of the First-Person

DRAFT - David Builes

1. Introduction

Imagine that you are an ideal reasoner who is enclosed in a windowless room. You know everything there is to know about physics that can be learned by reading physics textbooks: the initial conditions of the universe, the laws of physics, the distribution of fields and particles throughout the entire universe, etc. From this information, you come to know that there are exactly two agents in the universe, both of whom are subjectively indistinguishable and enclosed in rooms that are also indistinguishable¹. One of the rooms is located next to a lake, and the other is not. From this information, you come to know that you must be one of these two agents. However, even though you know everything about the physical features of the universe, it still seems like you are ignorant about something. For example, you don't know if there's a lake outside of your room.

As has been noted before, this story bears a striking resemblance to Jackson's (1982) famous "knowledge" argument regarding consciousness.² Mary, a brilliant color scientist (and also an ideal reasoner) is locked in a black and white room her whole life. Inside the room, she learns about all the relevant physical truths regarding color, including all the relevant physics and all the relevant neuroscience. In fact, we can imagine that she also knows everything there is to know about the initial conditions of the universe, the laws of physics, the distribution of fields and particles throughout the universe, etc. Still, it seems like she is ignorant about something. For example, she doesn't know what it's like to see red.

These epistemic gaps are surprising. It is surprising that an ideal reasoner can know everything there is to know about what physics says (and what any other science says), yet still be ignorant about a certain class of truths.³

¹ It is controversial whether information about macroscopic objects is knowable given only information about microphysics. For further defense of this claim, see Chalmers (2012), and for criticism of this claim, see Schaffer (2017). Regardless, the example can be modified so that Alice is given all "non-indexical" information about the universe.

² Ismael (1999), Perry (2001), Stalnaker (2008), Kwon (2017), and Builes (2023) also discuss this analogy.

³ Throughout, when I speak of "truths", I mean to be working with a notion that is sensitive to differences in cognitive significance, so that the truth that Hesperus is Phosphorus is a distinct truth from the truth that Hesperus is Hesperus. Such a notion is needed to account for epistemic differences between truths that might nonetheless correspond to the same worldly "fact" (e.g. one can rationally believe that Hesperus is Hesperus without rationally believing that Hesperus is Phosphorus). For further discussion on how to best understand such a notion of truth (e.g. by appealing to sentence tokens/types or Fregean propositions), see Chalmers (2012: 42-47)

Anti-physicalists about consciousness have argued that this kind of epistemic gap between physical truths and truths about consciousness motivates the view that truths about consciousness are not reducible to, or identifiable with, physical truths. The kinds of epistemic gaps that Anti-physicalists appeal to take many forms.⁴ Here are three popular versions:

Knowledge Gap: One can know all the physical truths and not be in a position to know (even ideally) every truth about consciousness (e.g. what it's like to see red).

Explanatory Gap: Physical truths don't explain truths about consciousness (e.g. why are brain states associated with certain conscious experiences rather than other ones, or none at all?).

Conceivability Gap: It is conceivable for the physical truths to be just as they are, but truths about consciousness to be different (e.g. it is conceivable for the physical truths to be just as they are, but no one is conscious).

As it turns out, all three of these kinds of gaps exist between objectively characterized "third-personal" truths⁵ and "first-personal" truths of the form "I am F" (e.g. "I am next to a lake").⁶ This is a point that is often made under the heading of the "essential indexical":

Knowledge Gap: One can know every third-personal truth and not be in a position to know (even ideally) every first-personal truth (e.g. whether I'm next to a lake).

Explanatory Gap: Third-personal truths don't explain first-personal truths (e.g. why am I the person next to the lake, rather than someone else, or no one at all?).

Conceivability Gap: It is conceivable for the third-personal truths to be just as they are, but the first-personal truths are different (e.g. it is conceivable for the third personal truths to be just as they are, but I'm (not) next to the lake).

⁴ See Jackson (1982), Chalmers (1996), and Levine (2001).

⁵ Just as it is a non-trivial task to give an adequate analysis of "physical" truths (e.g. see Stoljar 2010), it is non-trivial task to give an adequate analysis of "third-personal" truths. Intuitively, third-personal truths are meant to be true or false independently of any particular perspective. As a first pass, such truths are meant to exclude expressions that are contextually sensitive to a particular agent and their environment, such as indexicals and demonstratives. More carefully, using the framework of two-dimensional semantics (e.g. see Chalmers 2006a), one could say that a third-personal truth only involves expressions whose primary intension, which is construed a function f from centered worlds to extensions, is such that, for any two centered worlds c_1 and c_2 that only differ with respect to their center, $f(c_1) = f(c_2)$.

⁶ It is natural to further specify that such truths are of the form "I am now F". I further discuss the case of "now" in section 4.1 and note 33.

Because these epistemic gaps seem to be *prima facie* parallel, there is some pressure for the Anti-Physicalist to think that, if the epistemic gaps with respect to consciousness show that truths about consciousness are not reducible to or identifiable with physical truths, then the epistemic gaps with respect to the first-person show that truths about the first-person are not reducible to or identifiable with third-personal truths. Let us call the view that all first-personal truths are reducible to or identifiable with third-personal truths *Objectivism*. Let us call the denial of Objectivism *Subjectivism*.

The goal of this paper is to explore four different ways of responding to the epistemic gap between third-personal and first-personal truths, in a way that parallels four different ways of responding to the epistemic gap between physical truths and truths about consciousness. I'll begin by drawing a parallel between two standard Physicalist responses to the epistemic gaps with respect to consciousness and two kinds of Objectivist responses to the epistemic gap with respect to the first-person (sections 2 and 3). Then, I will introduce Subjectivism and discuss some of its potential advantages and disadvantages (section 4). Lastly, I will introduce and defend a non-standard kind of Objectivism about the first-person (section 5), which is in some ways analogous to a non-standard kind of Physicalism about consciousness (section 6).

2. Type-A Objectivism

One response to the alleged epistemic gap between physical truths and truths about consciousness is to deny that there is such a gap, at least upon rational reflection. This may be because there are no positive truths about consciousness at all (i.e. an “eliminativist” where no one is conscious) or because, while there are positive truths about consciousness, they are all *a priori* entailed by the instantiation of highly complex physical and functional properties. Chalmers (2002) calls this kind of view “Type-A” Physicalism.⁷

The same kind of view can be taken in the first-personal case. According to Type-A Objectivism, there are no first-personal truths for which the above epistemic gaps arise. This may be because there are no positive first-personal truths at all, or because, while there are positive first-personal truths, they are all *a priori* entailed by third-personal truths. I'll first consider how these two versions of Type-A Objectivism might be developed, after which I'll assess their overall plausibility.

2.1 Type-A Eliminativism

According to Type-A Eliminativism, there are no positive first-personal truths. Such a view implies that I don't exist and that every sentence of the form “I am F” is false. Just like eliminativism about consciousness, eliminativism about the first-person is extremely radical.

⁷ Type-A Physicalists include Lewis (1988), Dennett (1991), Frankish (2016), and Kammerer (2019).

One way to motivate Type-A Eliminativism is by way of certain “no-self” views with respect to consciousness.⁸ According to such views, many of the intuitive beliefs we have about the “self” are mistaken. For example, perhaps there is no unified, enduring, indivisible, or immaterial soul that is the subject of my stream of consciousness. However, merely denying that there is anything corresponding to such an inflationary conception of the self does not immediately imply Type-A Eliminativism. For example, I might still exist as a physical biological organism, which instantiates various properties associated with conscious experience. One could insist that “I” must refer to an inflationary notion of the self if it is to refer at all, but then it’s unclear why one couldn’t introduce an alternative concept of “I*” that need not refer to such an inflationary notion of the self. Then, the same first-personal gaps that arise for “I” would plausibly also arise for “I*”.

In response, one might supplement such a “no-self” view with *mereological nihilism*, which is the view that no composite objects (such as biological organisms) exist.⁹ If there are neither immaterial souls nor composite objects, then it becomes very unclear what I (or I*) am supposed to be.

However, it’s still not obvious that such a view avoids any first-personal gaps. If a proponent of this view does not also want to be an eliminativist about consciousness, then there must be some sense in which conscious states can be *plurally* predicated of “simples-arranged-human-wise”. In other words, although there might not be a single entity that is the subject of any conscious experience, perhaps many particles (or space-time points, or whatever mereological simples there are) can in some sense jointly experience pain. So, while it might not be true that “I am in pain”, since this seems to presuppose that a single thing is in pain, perhaps instead it is true that “we are in pain”, where “we” refers to the particles that intuitively “correspond” to me and jointly experience pain. If “I” can be replaced with “we” in this way, then perhaps it still makes sense to ask whether *we* are (say) next to the lake or not. If this does make sense, then the same epistemic gaps that arise for truths about “I” will also arise for truths about “we”.

There is an even more radical view that naturally rejects both truths about “I” and “we”, namely *existence monism*.¹⁰ Existence monism (or monism for short) is the view that only one thing exists, namely the world as a whole.¹¹ For the monist, what accounts for the appearance of ordinary objects are the global properties had by the world as a whole. Following Parsons (2004), these properties are typically thought of as *irreducible distributional properties*. A “distributional” property is a property that specifies how some quality is distributed across a spatially extended object. For example, *being polka-dotted* is a color distributional property: it specifies how colors

⁸ See, for example, Metzinger (2009) and Smith (2021).

⁹ For a defense of mereological nihilism, see Sider (2013). Unger (1979) defends the claim that “I do not exist” on the basis of mereological nihilism.

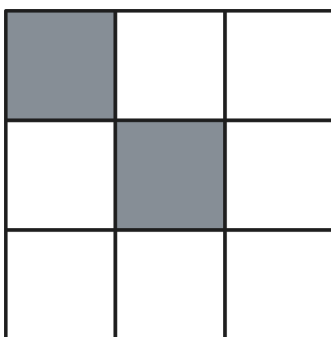
¹⁰ Another view that naturally rejects truths about “I” and “we” is developed by Turner (forthcoming).

¹¹ For defenses of existence monism, see Schaffer (2007), Cornell (2016), Builes (2021, 2023), and Builes and Teitel (2022).

are spatially distributed across some object. Of course, *being polka-dotted* isn't a maximally specific distributional property, because there are a variety of different ways to be polka-dotted. However, for understandable reasons, English does not have single words for all the maximally specific ways to be polka-dotted. To give some other examples, *being rippled* and *being still* can be thought of as distributional properties that specify the shape of the surface of a lake. In the case of our best physical theories, perhaps the most plausible candidate distributional properties are distributional field-theoretic properties, which specify the distribution and magnitude of various quantum fields across space-time. Distributional properties themselves should be uncontroversial: it shouldn't be controversial that some things are polka-dotted (at least for those who believe in the relevant composite objects). However, we ordinarily think that ascriptions of distributional properties to composite objects are "reducible" to the properties and relations of the proper parts of that composite object. However, the monist takes (at least some of) the distributional properties of the world to be *fundamental* properties that are not reducible in this way. This is an instance of the monist's broader revisionary claim that the universe should be thought of in a "top-down" way rather than a "bottom-up" way.

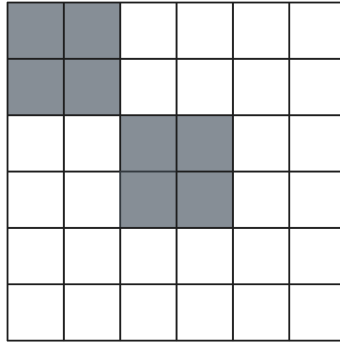
Let us return to our lake case. In the case of nihilism, while there can't be any ignorance about where *I* am located (since I don't exist), there might still be ignorance about where "we" are located. However, for the monist, there are no non-trivial truths about the locations of any object(s). The only object that exists is the world as a whole, and there is no (non-trivial) question to ask about where the world is located. So, ignorance about where I am located cannot be substituted for ignorance of the location of any object(s). In fact, ignorance about where I am located does not seem to be substitutable for any ignorance at all, since third-personal truths seem to fully characterize *all* of the distributional properties of the world.

To further illustrate this point, here is a toy example. Consider a world that seems to contain nine squares, only two of which are grey:

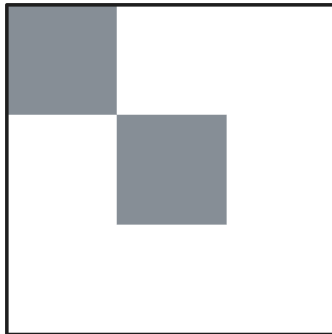


Suppose I take myself to be one of the grey squares (e.g. perhaps my visual field is uniformly grey). Even if I knew exactly how colors are distributed across the world, I might still be ignorant about where I am: am I the square in the corner or the middle?

Now, suppose mereological nihilism is true. For example, suppose none of the nine middle-sized squares exist. In place of each middle-sized square, there are only four smaller-sized squares. Then, while there are no truths about where I am located (since I don't exist), perhaps there can still be truths about where "we" are located.



Lastly, suppose monism is true. Then, once one fixes how the "universal" square is colored, there is nothing left to be ignorant about. There are no non-trivial truths about the location of anything. There is just the colored world:



Of course, none of these views – a "no self" view, mereological nihilism, or monism – show that Type-A Eliminativism is plausible, but they perhaps show that it is at least coherent.

2.2 Type-A Relativism

An alternative version of Type-A Objectivism acknowledges that there are positive first-personal truths but claims that these truths are a priori entailed by third-personal truths.

This view has a hard time handling (say) our initial lake case. For example, it seems like both the person next to the lake and the person not next to the lake are in principle capable of being

omniscient about every third-personal truth¹², but they still wouldn't be able to figure out whether they are next to a lake. Unlike first-personal truths whose content seems to vary based on who is entertaining them (e.g. "I am F" is only true for a speaker if they are F), third-personal truths are not supposed to have this kind of perspectival variation.¹³ So, if they were both omniscient about the third-personal truths, they would be omniscient about *the very same* third-personal truths. However, if they knew the very same third-personal truths, they could not possibly rationally draw *different* conclusions (e.g. "I am next to the lake" vs. "I am not next to the lake") based on these very same third-personal truths.

Still, it is helpful to look at a view that can recognize certain "deflationary" first-personal truths that are knowable on the basis of third-personal truths. Suppose, for example, that one followed Lewis (1979) in thinking that the content of a thought of the form *I am F* is a property, namely *being F*, rather than a proposition. Unlike propositions, properties are not the kind of thing that can be true or false. Properties are "unsaturated" in the sense that they can only be evaluated for truth or falsity when they are taken to be instantiated by a particular object. On this view, the content of I-thoughts can, at best, only be evaluated for truth or falsity *relative* to a particular individual. So, relative to someone who is F, the content expressed by "I am F" is true, and relative to someone who is not F, the content expressed by "I am F" is false. To go back to our previous example, relative to the person next to the lake, the content expressed by "I am next to the lake" is true, and relative to the person not next to the lake, the content expressed by "I am next to the lake" is false. Let *Type-A Relativism* be the view that the only first-personal truths are relativized truths of the form: "relative to D, I am F", where D is any third-personal description that uniquely picks out some individual, and it is true that "relative to D, I am F", whenever whoever is D is F.¹⁴¹⁵

These relativized facts do not give rise to any distinctive epistemic gaps. Given full third-personal knowledge, one will be able to give a third-personal characterization of which speakers and thinkers there are, and which "I" thoughts they entertain. One can then evaluate the truth or falsity

¹² There might be some third-personal truths that are in principle impossible for anyone to know or even entertain (such as truths about quiddities that cannot be expressed, see Lewis 2001), but such truths do not seem relevant to first-personal truths.

¹³ For a more precise definition of "third-personal truth", see note 4. Note that the definition of note 4 implies that demonstrative truths (e.g. if one of them points to the door and says "*that* door is next to the lake") count as first-personal truths, since the content of such demonstratives (e.g. which door is being demonstrated) depends on who is doing the demonstrating and their local environment.

¹⁴ Following Weber (2016), although Type-A Relativists think the content of first-personal thoughts cannot be true or false simpliciter, they can still make sense of there being belief *states* had by particular people that are true or false simpliciter. For example, a person P's belief state that "I am F" is true simpliciter if and only if P is F. Still, truths about which objectively characterized agents have which true or false first-personal belief states will all be knowable given third-personal truths.

¹⁵ Although this view is inspired by Lewis (1976), I do not mean to be attributing this view to Lewis. For one, this view does not give any role to "self-ascription", which is central to Lewis' view (e.g. see Jackson and Stoljar 2020).

of these various I-thoughts relative to the objectively characterized speaker or thinker who entertains them, without having to run into any epistemic gaps.

2.3. *Type-A Objectivism and Absent Worlds*

Both kinds of Type-A Objectivism that we've seen face a similar kind of problem as Type-A Physicalism. In the case of Type-A Physicalism, there seems to be a datum about consciousness that the Type-A Physicalism is not recognizing: there are truths about the qualitative character of experience that are (at least conceptually) "over and above" ordinary physical truths. If we let a "zombie" world be a world without any conscious beings, then Type-A Physicalists struggle with accommodating the following two facts: (i) the actual world is not a zombie world and (ii) it is conceivable for there to be a zombie world that is physically just like the actual world.

Similarly, there seems to be a datum about the first-person that the Type-A Objectivism is not recognizing: there are first-personal truths that are (at least conceptually) "over and above" third-personal truths (and relativized first-personal truths). For example, it is a datum that I exist (simpliciter).¹⁶ If we let an "absent" world be a world where I do not exist, then Type-A Objectivists struggle with accommodating the following two facts: (i) the actual world is not an absent world and (ii) it is conceivable for there to be an absent world that is third-personally just like the actual world.

Is an absent world really conceivable? Well, to start, there are certainly trivial examples of absent worlds. For example, a world with nothing but a few rocks floating in space will be an absent world. After all, it seems inconceivable that I am a rock. More interestingly, consider a third-personal description T that describes a universe that contains no conscious creatures, but it contains a GPS system lying on the floor somewhere. This GPS system in some sense has "indexical" representations, which track where it is relative to its environment (maybe it runs Google Maps). It seems to me that there is no conceivable world where T holds and I exist. Just as it is inconceivable for me to be a rock, it is inconceivable for me to be a (wholly unconscious) GPS system with indexical representations. This shows that the distinctive kind of explanatory gap that arises with "I" (associated with there being different conceivable first-personal truths added on to the third-personal truths) does not merely have to do with there being indexical representations of some sort. Rather, it only arises when a world contains conscious beings. Once there is a third-personal description containing conscious beings, then there are at least two conceivable ways for such a third-personal description to be realized: in an "absent" way, corresponding to a kind of "view from nowhere", or in a "present" way, corresponding to a kind of "view from somewhere".

¹⁶ Even if it is not a datum that I exist (because of, say, the epistemic possibility of mereological nihilism), it still seems to be a datum that certain conscious experiences are "mine" in a way that is not being captured by ordinary third-personal truths.

So, we've seen that absent worlds are conceivable. However, is it really conceivable for David in particular to exist as a conscious being while I do not exist? Yes. Suppose I know that I am David, but it turns out that I have a doppelganger named Bob. Suppose both of our memories are wiped in such a way that neither of us remember whether we are David or Bob (but both of us still remember that David and Bob exist and have certain objectively characterized histories). Then, I would be uncertain whether I am David. So, I would be uncertain about whether what happens to me would also happen to David. As an example, suppose I am told that I will be killed in my sleep tonight. Then, although I believe that I won't exist tomorrow, it is still a live epistemic possibility that David will exist tomorrow. This epistemic possibility corresponds to a conceivable situation in which David exists tomorrow, yet I don't exist tomorrow.

In sum, just as Type-A Physicalism is radically revisionary in that it denies the datum that I am phenomenally conscious in a sense that is not physically or functionally analyzable, Type-A Objectivism is radically revisionary in that it denies the datum that there are first-personal truths that are true *simpliciter*, which differentiate the actual world from its corresponding absent world.

2.4. Debunking

By way of closing, it is worth noting one way that one might defend Type-A Objectivism that “mirrors” one way of defending of Type-A Physicalism. Consider a zombie world. In that world, there seems to be a wholly physical explanation for why zombies are disposed to make verbal reports about how their own states of consciousness, and there seems to be a wholly physical explanation for why zombies are disposed to make verbal reports about how consciousness is mysterious and physically inexplicable. Similarly, in the actual world, there will be some wholly physical brain-based explanation for our own verbal reports about consciousness. Some philosophers have thought that the availability of wholly physical explanations for our judgements about consciousness can be used to “debunk” our beliefs about consciousness. Why are our beliefs about consciousness more justified than the beliefs of zombies?¹⁷

Similarly, consider again the absent world that contains nothing but an unconscious GPS system. Imagine upgrading the GPS system with a sophisticated AI system, which eventually begins outputting sentences about how there are first-personal truths that are true *simpliciter* and that such truths are inexplicable given third-personal truths. There will certainly be some third-personal explanation for why such a GPS system is disposed to output such sentences. Similarly, there will be some third-personal explanation for why I, in the actual world, am disposed to make similar kinds of first-personal judgements. *Prima facie*, one can use this fact to motivate a “first-personal” debunking strategy that is highly analogous to the above debunking strategy in the case of consciousness. There is of course much more that can be said about both of these debunking strategies, but insofar as one is inclined to retain one's beliefs about consciousness in light of the

¹⁷ For further discussion, see Chalmers (2018, 2020).

consciousness debunking strategy, it is at least natural to be inclined to retain one's first-personal beliefs in light of the first-personal debunking strategy.¹⁸

3. Type-B Objectivism

According to "Type-B" Physicalism, there are truths about consciousness for which the above epistemic gaps arise, but nonetheless truths about consciousness are still reducible to or identifiable with physical truths.¹⁹

According to Type-B Objectivism, there is an epistemic gap between third-personal and first-personal truths, but nonetheless first-personal truths are reducible to or identifiable with third-personal truths. A natural version of this view is that the fact that that [I exist] is reducible (or identical) to the fact that [David exists], and more generally, the fact that [I am F] is reducible to or identical to the fact that [David is F]. Moreover, this reduction or identification is true regardless of the fact that there is a conceivable absent world in which David exists, yet it is not the case that I exist.

Type-B Objectivism is analogous to Type-B Physicalism. To take the standard example, Type-B Physicalists might say that the fact that [x is in pain] is reducible to or identifiable with the fact that [x has C-fibers firing], even though there are conceivable zombie worlds in which people have C-fibers firing without pain.

There is a lot to be said about the viability of both Type-B Physicalism and Objectivism, and whether such a view is ultimately viable will depend on a variety of background theoretical commitments. Nonetheless, it is worth looking at two standard arguments against Type-B Physicalism to see if they also apply to Type-B Objectivism.

3.1. *Fundamental Scrutability*

One way that Type-B Physicalism has been defended is by arguing that, in other examples of empirically motivated reductionist theses, where truths about X reduce to truths about Y (e.g. truths about water reduce to truths about H₂O, truths about genes reduce to truths about DNA, various truths about chemistry reduce to truths about physics), such a reduction does not seem to involve

¹⁸ There is another interesting parallel between zombies and (say) unconscious GPS systems. For zombies, one can ask: what is the content of a zombie's thoughts about consciousness? Similarly, one can ask: what is the content of an unconscious GPS system's "indexical" thoughts? Roughly similar options are available in both cases. For example, perhaps the question does not arise because neither zombies nor unconscious GPS systems have any thoughts at all. If they do have thoughts, then their thoughts can be assigned deflationary "type-A" contents or (false) inflationary contents that require the truth of Anti-Physicalism or Subjectivism. For further discussion of the content of "I"-thoughts according to Subjectivism, see the references in note 32.

¹⁹ Type-B Physicalists include Loar (1990), Papineau (2002), Balog (2012), and Levin (2018).

an unbridgeable epistemic gap: there is an *a priori* entailment from truths about X to truths about Y (e.g. see Chalmers and Jackson 2001). However, if it's true that every *other* example of reduction does not involve an unbridgeable epistemic gap, that gives some prima facie reason to think that, if there is to be a reduction from phenomenology to physics, then it should not give rise to an unbridgeable epistemic gap either.

A similar kind of objection can be made to Type-B Objectivism. If first-personal truths are to be reduced to third-personal truths, then this reduction should not give rise to an unbridgeable epistemic gap, contrary to Type-B Objectivism.

In response to this kind of argument, Type-B Physicalists can (and have) given two kinds of responses. First, they can argue that the very same kinds of epistemic gaps that arise between physical truths and truths about consciousness also arise in other paradigm cases of scientific reduction, so that the epistemic gaps that arise with respect to consciousness are not at all exceptional (e.g. see Block and Stalnaker 1999). Second, they can argue that, although there is a unique kind of epistemic gap that exists in the case of consciousness, there is a principled explanation for why such an epistemic gap would arise in the case of consciousness. This principled explanation often appeals to the distinctive nature of our *concepts* of phenomenal consciousness and often goes by the name of the “phenomenal concept strategy”.

A Type-B Objectivist can give the same two responses: either the kind of epistemic gap associated with first-personal truths is commonplace with other reductionist theses, or else there is a principled explanation for why there is a distinctive epistemic gap with respect to first-personal truths (perhaps due to the nature of our “indexical” concepts using an “indexical concept strategy”).

The general thesis that reduction must always be accompanied by *a priori* entailment naturally leads to the following thesis:

Fundamental Scrutability: All truths are *a priori* entailed by metaphysically fundamental truths.²⁰

Both the Type-B Physicalist and Type-B Objectivist reject **Fundamental Scrutability**. However, since our focus is on first-personal truths, we can ask the following question: are there any good reasons for thinking that first-personal truths are an exception to **Fundamental Scrutability**, even though truths about consciousness are not an exception to **Fundamental Scrutability**?

²⁰ For further details about how to precisely formulate this thesis (e.g. to take into account the mode of presentation of fundamental truths), see Chalmers (2012: 405).

Chalmers (2012: 404-409) is someone who has defended **Fundamental Scrutability** with the single exception of indexical truths.²¹ He justifies this exception as follows (in the relevant context, a truth is (in)scrutable if it is (not) a priori knowable given fundamental truths):

I think the case of indexical truths is special, though, in that one can straightforwardly explain why even in a world that is fundamentally objective, one would expect that there are inscrutable indexical truths, and this objective truth (like all others) is itself scrutable... (408)

Chalmers says this response is broadly analogous to the phenomenal concept strategy that Type-B Physicalists employ, but (for independent reasons) he thinks that the phenomenal concept strategy fails (e.g. see Chalmers 2006b):

It then remains open to [the Type-B Physicalist] do the same for phenomenal truths: that is to explain why, even in a world that is fundamentally physical, one would expect there to be inscrutable phenomenal truths...[but] the thesis that there are inscrutable phenomenal truths (or quasi-phenomenal truths, where these truths are cast in topic-neutral nonphenomenal terms) is not scrutable from fundamental physical truths. If this is right, the analogy with indexical truths cannot be maintained, and the opponent must once again postulate a unique and unexplained exception.

However, it is doubtful that Chalmers is right that “even in a world that is fundamentally objective, one would expect that there are inscrutable indexical truths”. Let us grant, for example, that it is an a priori truth that: if someone who is F utters “I am F”, then that utterance is true relative to that speaker. Then, it will be scrutable that various people have uttered sentences using “I”, and it will be scrutable that many such utterances are true relative to the speakers of those utterances. However, none of this shows that we should expect there to be *inscrutable* indexical truths. These relativized “Type-A” indexical truths are all scrutable (similarly: there are scrutable “Type-A” truths about consciousness, which are analyzable in terms of complex physical and functional facts, even in a fundamentally physical world). The only kinds of inscrutable indexical truths are indexical truths that are true *simpliciter*, without being relativized to anything else. However, it’s not true that we should expect there to be such inscrutable indexical truths in a fundamentally objective world. After all, the fundamental objective truths all hold in an “absent” world where there are no (positive and non-relative) indexical truths at all.

An alternative strategy in the vicinity is to argue that it is scrutable from a third-personal description that it will *seem* to various agents as if there are inscrutable indexical truths. However, it’s not clear that this establishes an asymmetry with the case of consciousness. As we saw in

²¹ However, in earlier work, Chalmers tentatively defends fundamental indexical facts (e.g. see Chalmers 1996: 84-86).

section 2.4, various Physicalists have motivated their view on a precisely similar basis: because there will be a wholly physical explanation for why it *seems* as if there are inscrutable truths about consciousness, then this (allegedly) serves to “debunk” Anti-Physicalist arguments about consciousness.²² So, this alternative strategy doesn’t seem to provide a disanalogy between first-personal truths and truths about consciousness.

3.2. *Conceivability and Possibility*

An initial way to motivate Type-B Physicalism is by noting that there are certain widely agreed upon examples where conceivability does not imply possibility. For example, it is conceivable that Water is not H₂O, even though it is impossible that Water is not H₂O. For similar reasons, one might initially think that (i) the conceivability of a zombie analog of the actual world does not imply that it is possible that there is a zombie analog of the actual world and (ii) the conceivability of an absent analog of the actual world does not imply that it is possible that there is an absent analog of the actual world.

However, in response to this argument, Anti-Physicalists have pointed out that, in the case of water, when we are conceiving of a world where water is not H₂O, we are successfully conceiving of a metaphysically possible world such that, *if it turned out to be actual*, water would not be H₂O (e.g. a metaphysically possible world where the clear substance in lakes and rivers had some other chemical composition). In other words, by drawing a distinction between *considering a world as actual* and *considering a world as counterfactual*, one can argue that the fact that water is necessarily H₂O does not undermine the general principle that the conceivability of a truth is strong evidence that there is a metaphysical possibility where that truth obtains, when that metaphysical possibility is considered as actual.²³

However, this general principle makes trouble for Physicalism. In the case of a zombie world, the Physicalist *cannot* say that when we are conceiving of a zombie world, we are successfully conceiving of a genuinely metaphysically possible world such that, when considered as actual, there is no consciousness. This is because the very same physical facts obtain in both the conceivable zombie scenario and the conceivable scenario that corresponds to the actual world, so on a Physicalist view, these two conceivable scenarios must correspond to the very same metaphysically possible world, namely the actual world.²⁴ However, it is certainly false that if the

²² Note that Physicalist who takes this line should interpret the relevant sense of “seems” in a functional sense rather than a phenomenal sense, because no phenomenal seeming will be scrutable from a purely physical basis (because of the conceivability of zombies). The very same functional sense of “seem” can also be operative in the first-personal case (where it is scrutable that it will seem to various agents as if there are inscrutable indexical truths).

²³ For further discussion of this link between conceivability and possibility, see Chalmers (2002).

²⁴ There is a loophole here for “Russellian Monist” views. It is plausible that when we conceive of zombie worlds, we are only conceiving of worlds with the same “structural” physical truths, but it might be that the “intrinsic nature” of fundamental physical properties is responsible for the existence of consciousness. However, conceivability arguments

actual world turned out to be actual, then there would be no consciousness. The actual world did turn out to be actual, and there is consciousness!

As it turns out, the case of an absent world is just like the case of a zombie world. The Objectivist cannot say that when we are conceiving of the absent analog of the actual world, we are successfully conceiving of a genuinely metaphysically possible world such that, when considered as actual, I wouldn't exist. This is because the very same third-personal facts obtain in both the conceivable absent scenario and the conceivable scenario that corresponds to the actual world, so on an Objectivist view, these two conceivable scenarios must correspond to the very same metaphysically possible world, namely the actual world. However, it is certainly false that if the actual world turned out to be actual, then I wouldn't exist. The actual world did turn out to be actual, and I do exist!²⁵

In response to this kind of symmetry, an Objectivist could say that although there are distinct *centered* metaphysically possible worlds that correspond to the conceivability of the actual world and the conceivability of the corresponding absent analog of the actual world (where the "center" of a world picks out who I am), centered metaphysically possible worlds that merely differ in their center do not correspond to distinct metaphysical possibilities. In other words, mere differences in center are not genuine metaphysical distinctions. Only distinct "uncentered" worlds are genuinely metaphysically distinct. However, this response needs to be independently motivated, because an exactly similar response can be given by the Physicalist. The Physicalist can say that the conceivability of P entails that there is a *phenomenal-centered* metaphysically possible world in which P is the case, when that world is considered as actual. However, distinct phenomenal-centered metaphysically possible worlds that merely differ with respect to centered truths or phenomenal truths do not correspond to distinct metaphysical possibilities. At least in the absence of further arguments, these two responses seem to be parallel, so they don't give any reason to favor Type-B Objectivism over Type-B Physicalism.²⁶

are still a problem for "standard" versions of Physicalism, according to which the physical truths responsible for consciousness are merely about the "structure" of the physical world. For defenses of Russellian Monism, see Strawson (2006) and Goff (2017).

²⁵ Note that it is not a premise of this "two-dimensional" argument that phenomenal concepts or the "I" concept have the same primary and secondary intension (i.e. that what they refer to does not depend on which world turns out to be actual). So, it is not a relevant asymmetry to point out that the "I" concept does not have the same primary and secondary intension. If "I" does refer to different things depending on which world turns out to be actual, this argument would still establish that (say) even though what "I" refers to is a wholly objective entity (David), there are irreducible subjective properties associated with the primary intension (or "mode of presentation") of "I" that serve to pick out David. In fact, this is precisely what the views I will be considering later (Subjectivism and Universalism) will go on to say. See Chalmers (2010: 153) for further discussion on why the two-dimensional argument does not assume that phenomenal concepts have the same primary and secondary intension.

²⁶ A third way that Anti-Physicalists have argued against Type-B Physicalism is by appealing to a premise called "Revelation", which very roughly says that phenomenal concepts reveal the "essence" of the properties that they refer to (e.g. Goff 2017). I'll briefly make three points about this argument. First, the premise of Revelation can be

4. Subjectivism

In order to comply with **Fundamental Scrutability**, Subjectivists must believe that the metaphysically fundamental facts are in some way “asymmetrically oriented” towards David, so that it is scrutable that (say) “I am David” or “I am in pain”. However, the world seems to be populated by many “I”s rather than just one “I”. How do Subjectivists capture the multi-perspectival structure of reality?

They typically do so in a way that is structurally parallel to more familiar positions in the metaphysics of modality and time.²⁷ With respect to modality, we can ask: are all possible worlds “on a par” or is one metaphysically distinguished in some way? The standard answer is that one world, the actual world, is metaphysically privileged. Still, it is only *contingently* privileged rather than necessarily privileged. We can ask the same question with respect to time: are all times “on a par” or is one metaphysically distinguished in some way? According to “A-theories” of time, the present time is metaphysically privileged. Still, it is only *temporarily* privileged rather than permanently privileged.

The same question can be asked about “perspectives”: are all perspectives “on a par” or is one metaphysically distinguished in some way? According to a Subjectivist, one perspective is metaphysically privileged. However, it is only *subjectively* privileged rather than objectively privileged.²⁸ A “subjective” fact is analogous to a contingent and temporary fact, and an “objective” fact is analogous to a necessary and permanent fact. Just as a contingent (temporary) fact is a fact that obtains according to some possible worlds (times) and not others, a subjective fact is fact that obtains according to some perspectives and not others. Similarly, just as a necessary (permanent) fact is a fact that obtains according to all possible worlds (times), an objective fact is a fact that obtains according to all perspectives. So, to say that I am subjectively metaphysically

challenged. For example, Stoljar (2018) notes that it is doubtful whether our phenomenal concepts reveal whether (say) the sense-data theory is true, but arguably if the sense-data theory is true it is an essential truth about phenomenology (see Roelofs 2000 for further discussion on how one might qualify the thesis of revelation). Second, the inference from Revelation to the falsity of Physicalism can be challenged. For example, perhaps Revelation can tell us what phenomenal properties essentially are, but they cannot tell us what they are *grounded* in (e.g. whether they are grounded in physical properties), or perhaps as Levin (2018) argues, even the identity physicalist can accept Revelation. Lastly, even if the argument from Revelation is successful, it is plausible that the two views I consider later - Subjectivism and Universalism – will also endorse Revelation for “presence” or “immediacy”, as these are meant to be properties associated with phenomenology. Neither of these views need to endorse Revelation for “I”, since “I” will turn out to be something like a definite description, referring to whatever it is whose experiences happen to be present/immediate.

²⁷ For further discussion of the parallels between time and modality, see Rini and Cresswell (2012), Emery (2018, 2020), Skow (2022), and Builes (2023).

²⁸ The terminology of “subjective” and “objective” facts comes from Merlo (2016).

privileged is to say that I am metaphysically privileged (simpliciter), but there are other perspectives according to which I am not metaphysically privileged.²⁹

Just as there are multiple ways to understand the notion of “metaphysically privileged” in the modal and temporal case, there are also multiple ways to understand the notion in the perspectival case. However, what all Subjectivists are trying to “get at” is the following intuitive distinction. Imagine a world that contains two conscious beings, one of whom is having a phenomenally red experience and the other of whom is having a phenomenally green experience. Now, imagine the following possibility: you are the one experiencing phenomenal red. Now, imagine another possibility: you are the one experiencing phenomenal green. According to the Subjectivist, there are two genuinely distinct metaphysical possibilities here. What does the difference between these worlds consist in?

According to Hare (2009) the difference between them is that, in the first possibility, the phenomenally red experience is *present*, and in the second possibility the phenomenally green experience is *present*. More exactly, Hare says that there is a property of *presence* that applies to whatever my direct objects of awareness are (e.g. perhaps sense-data for sense-data theorists, perhaps external physical objects for naïve realists, etc.). Importantly, for Hare, there is nothing fundamental about “I”. Rather, “I” functions merely as a definite description that refers to whoever’s objects of direct awareness happen to be present. So, for Hare, there are two perspectives in the world, and the “metaphysically privileged” one corresponds to the one whose direct objects of awareness happen to be present.

According to Merlo (2016), the mental states that characterize the “metaphysically distinguished” perspective have a character that no one else’s mental states have. The metaphysically distinguished perspective experiences (say) PHENOMENAL RED, while the other perspective only experiences (say) phenomenal green. In general, the metaphysically distinguished perspective has MENTAL STATES, while other perspectives merely have mental states. In general, someone has a particular mental state in virtue of the fact that, from their perspective, they have the corresponding MENTAL STATE. So for example, others are conscious in virtue of the fact that, from their perspective, they are CONSCIOUS. MENTAL STATES are not meant to be a mysterious posit: from the first-person point of view, my own mental life is wholly characterized by MENTAL STATES. I can know what PAIN is like just by experiencing it directly for myself. The distinctive aspect of this view is that it has a metaphysics that reflects the epistemological point that our primary understanding of MENTAL STATES comes from our acquaintance with the relevant states in our own first-personal case, and we can only understand how others have a mental life in a derivative way: by asking what things are like from their perspective. If someone else is in pain, then this is because, from their perspective, they are in PAIN. On this view, the reason why zombies don’t have

²⁹ For more on the relevant notion of “perspective”, see Merlo (2016: 314-315) and Lipman (2023a, 2023b).

any mental states at all is because there is no “perspective” that corresponds to a zombie (just as, presumably, there is no perspective that corresponds to a chair).

There is a lot to be said about whether Subjectivism is a defensible view. Elsewhere, I have surveyed eight different arguments in favor of Subjectivism (Builes 2024a). I’ll briefly cover a few considerations in favor of Subjectivism below (sections 4.1 and 4.2), but my main goal will be to argue that Subjectivism should ultimately be rejected (sections 4.3, 4.4, and 4.5).

4.1. Three Initial Objections

The first natural worry about Subjectivism is that it seems objectionably arbitrary. Why did my perspective happen to be the metaphysically privileged one? The problem with this worry is that it seems to overgeneralize. Anyone who thinks the actual world or the present time is metaphysically privileged faces an exactly similar worry: why did the actual world (present time) happen to be the metaphysically privileged world (time)? Moreover, in order to make these views seem less arbitrary, one can stress that the actual world is only contingently privileged, the present time is only temporarily privileged, and my own perspective is only subjectively privileged.³⁰

A second natural objection to Subjectivism is epistemological. Even if I am the unique metaphysically privileged perspective, how can I tell? Won’t everyone else think that they are uniquely metaphysically privileged? Why should I think that I am right and they are wrong?³¹

To respond to this objection, first consider the following analogous case. Suppose I am told that there is only one conscious being in the world, and everyone else who seems to be conscious is actually a zombie. Would that make me think that I am a zombie? No, it is natural to say that I can know with (near) certainty that I am conscious, even if zombies “think” (in some deflated functional or behavioral sense) that they are also conscious. Similarly, suppose only I am subjectively metaphysically distinguished. Then, I can know this with (near) certainty because (say) I would be directly acquainted with PAIN. Doubting whether I am really in PAIN would be tantamount to seriously entertaining the possibility that, although *someone* is in pain, I am not in pain (because I am in pain iff there is PAIN). But, surely I can tell whether I’m in pain! Although it’s true that other people would also think they are metaphysically distinguished, no one else would THINK they are metaphysically distinguished (just as zombies would not “think” they are conscious in the same consciousness-involving sense that I would think that I am conscious).³²

³⁰ Two other ways of responding to the objection from arbitrariness are the “many worlds” account of List (2023) and the “fragmentalist” approach inspired by Fine (2005). For further discussion, see Lipman (2023a) and Solomyak (2024).

³¹ A similar worry arises for some views that privilege the actual world or the present time. See, for example, Bricker (2006), Forbes (2015), and Builes (2022).

³² Another natural worry: if “I” refers to whoever is present, doesn’t this imply that everyone else’s “I” thoughts are mistaken? See Hare (2009: 52-55, 2010: 765-768) for a number of different ways of thinking about the “I” thoughts

A third natural worry for Subjectivism is that, insofar as it is motivated by the inscrutability of truths involving "I", there seem to be other indexical and demonstrative concepts that also generate truths that are inscrutable from a purely third-personal description: "here", "now", "you", "this", "that", etc. But surely not all of these concepts correspond to some extra metaphysical ingredient in the world!

In response, it should be noted that none of these indexical/demonstrative concepts are independent of "I". For example, consider an absent world containing nothing but a bunch of rocks. In such a world, it doesn't make sense to ask if there is a "further fact" corresponding to which rock is *here* or which is *this* rock, etc. In other words, there can only be truths involving these other indexical and demonstrative concepts in worlds where I exist. Furthermore, as Lewis (1979) and Chalmers (2012: 285-287) have argued, truths involving other indexical and demonstrative concepts seem to be scrutable given first-personal truths and third-personal truths.³³ For example, for the Subjectivist, "here" roughly refers to whatever location has MENTAL STATES, "now" roughly refers to whatever time has MENTAL STATES, demonstratives like "this" and "that" roughly refer to whatever I am (or intend to be) demonstrating, etc.³⁴

4.2. Subjectivism and Personal Identity

Another way to motivate Subjectivism (beyond the "explanatory gap" between first-personal and third-personal facts) involves the metaphysics of personal identity.

of others (none of which imply that others are incorrect for thinking in the standard ways that they do). For further discussion, see Merlo (2016: 329-331).

³³ Chalmers (2012: 286-287) has argued that there are certain truths involving "experiential demonstratives" that, at least in hypothetical symmetric situations (involving variations of Austin's (1990) "two-tubes" case), are not scrutable from third-personal and first-personal truths. Chalmers handles these cases by associating a primitive experiential demonstrative *this_E* for every "atomic" experience E that I have. I myself am skeptical that these kinds of examples generate truths that are inscrutable from first-personal and third-personal truths, for the reasons discussed by Wolfgang Schwarz here: <https://www.umsu.de/wo/2013/591>. However, if they wish, a Subjectivist could think that there are fundamental truths corresponding to these experiential demonstratives (e.g. for every direct object of my awareness O, there might be fundamental truth that "this_O is present", where the concept of "this_O" will figure in other fundamental truths, perhaps involving various spatiotemporal and causal relations that this_O stands in).

³⁴ I say "roughly" because the referents of these terms can also vary based on what I intend to refer to (and other factors such as what is contextually appropriate given the conversation, perhaps facts about what is metaphysically natural given reference magnetism, etc.). For example, "here" can be used to refer to a room, a city, a country, etc. Similarly, "now" can be used to refer to a particular instant, a day, a year, etc. Note that there are difficulties with defining "now" as "whatever time I exist" if (say) there is a four-dimensional block universe and I endure through time and am therefore wholly located at multiple times. However, although defenders of Subjectivism have not been explicit about this, it seems to me that the best version of Subjectivism does not assign *presence* tenselessly to every experience I ever have in the "block universe" (otherwise "now" cannot be defined in terms of "presence"), but rather Subjectivists should only assign *presence* to the experiences that I am having at a single moment (now). Subjectivism can then be supplemented with facts about what *was* present and what *will* be present, where these tensed facts are understood in the way that an "A-theorist" of time would understand them.

First, many find it intuitive that personal identity through time must be binary, precise, and non-conventional. In other words, personal identity through time cannot come in degrees (it doesn't make sense for it to be "half-true" that I exist tomorrow), it cannot be vague (it cannot be vague whether I exist tomorrow), and it cannot be a mere matter of convention whether I exist tomorrow.³⁵ Popular accounts of personal identity in terms of bodily or psychological continuity have trouble accommodating these three desiderata.³⁶ However, if one accepts the view that there can be no fundamental metaphysical indeterminacy, then Subjectivism implies that it must always be a binary, determinate, and non-conventional matter whether I exist tomorrow. I exist tomorrow just in case anyone has MENTAL STATES tomorrow (or just in case anything is *present* tomorrow).³⁷

Second, Subjectivism can make sense of the intuition that it is both conceivable and possible that I am (or can become) someone else, without rejecting the basic metaphysical principle that distinct things are necessarily (and permanently) distinct. On standard views, if "I" functions as a rigid designator that refers to a particular person, then it is not possible for me to be (or become) anyone distinct from who I presently am (David). However, there are certainly situations where there are two numerically distinct people, and it is conceivable that I am either one (e.g. the case above where my memories are wiped and it is epistemically possible that I am either David or Bob). Subjectivists can maintain that conceivability implies possibility in this case, without rejecting the necessity of distinctness (e.g. by saying it is metaphysically possible for David's direct objects of awareness to be *present*, and it is metaphysically possible for Bob's direct objects of awareness to be *present*). Moreover, it seems conceivable that I occupy the perspective of someone other than David in the future. For example, suppose that, while David is sleeping tonight, David will symmetrically fission into Lefty and Righty. Lefty will wake up in a red room, while Righty will wake up in a green room. I might be convinced that David is not numerically identical to either Lefty or Righty (since David can't be identical to both and it would be arbitrary if David was identical to either one), but even grating this view, it seems that I can clearly conceive "from the inside" of waking up as Lefty and experiencing a red room or waking up as Righty and experiencing a green room. A Subjectivist can straightforwardly accommodate both of these metaphysical possibilities. In one possibility, red will be *present* tomorrow, and in the other possibility, green will be *present* tomorrow.

³⁵ However, see Braddon-Mitchell and Miller (2004, 2020) for a defense of the view that personal identity can come in degrees and/or be partly a matter of convention.

³⁶ Defenders of psychological accounts include Lewis (1976) and Parfit (1984) and defenders of bodily accounts include Thomson (1997) and Olson (1997).

³⁷ Another view that accommodate these desiderata is the view that we are enduring immaterial souls (e.g. see Swinburne 2019). However, such a view cannot accommodate the intuition of dissociation in the next paragraph.

4.3. Problem: The Dynamics of Presence

As we've seen, Subjectivism has a number of virtues. Ultimately, however, I think Subjectivism should be rejected.

The first problem with Subjectivism is a consequence of one of the “virtues” in the previous subsection. In particular, Subjectivism allows for the metaphysical possibility of *presence* moving around in arbitrarily complex ways. It might be that, every morning, *presence* shifts from one person to another. In fact, Subjectivism seems compatible with the possibility of presence shifting from one conscious perspective to another every second.

Anti-Physicalism about consciousness seems to have a structurally similar consequence. If there is no necessary connection between physical facts and phenomenal facts, couldn't the connection between the physical and phenomenal be very strange and complicated? Perhaps there is a metaphysically possible world with the same history of physical and phenomenal facts as the actual world, but everyone will become a zombie tomorrow. Perhaps there is another such world in which phenomenal colors will get inverted tomorrow. Perhaps there is another world that is physically identical to our own, but the only conscious experience that anyone has ever had is a gray visual field.

A natural initial response to this problem from the Anti-Physicalist is to posit systematic psycho-physical laws connecting physical facts and phenomenal facts.³⁸ A similar initial response is also natural for the Subjectivist. The Subjectivist can posit “bridge laws” between third-personal and first-personal facts that imply (for example) something is *present* if and only if it is a direct object of awareness of David. However, such a law does not seem remotely like any other fundamental law of physics. For one, in order to avoid proper names like “David” appearing in the fundamental bridge laws, a precise description of David's direct objects of awareness throughout time must be given in fundamental terms. Such a law will be massively complex. Moreover, as various puzzle cases in the literature on personal identity demonstrate, it is not at all clear what the persistence conditions of David are. Does David survive teletransportation, brain transplants, or fission? A precise bridge law determining the dynamics of presence would have to give precise (and therefore seemingly ad hoc and arbitrary) verdicts in all of these cases. Lastly, there are some possible worlds where it might not even be possible to formulate a bridge law in qualitative terms. For example, in a perfectly symmetrical world where there are two people – David and David* – that are qualitatively indiscernible, there can be no qualitatively described fundamental bridge law that determines whether David or David* is present.

³⁸ This response still raises various questions. For example, will the psycho-physical laws be extremely complicated? Is there a sense in which the psycho-physical laws are “lucky” and call out for deeper explanation? For further discussion, see Cutter and Crummett (2025).

4.4. Problem: Causal Exclusion

A second problem for Subjectivism also corresponds to a popular objection to Anti-Physicalism. In the case of Anti-Physicalism, we can ask: what is the causal role of consciousness? On the one hand, it seems highly intuitive that consciousness causally influences the physical world. For example, it is very natural to think that consciousness at least plays a causal role in explaining our own beliefs about consciousness. However, on the other hand, the physical world seems “causally closed”, in that every physical event has a complete causal explanation in terms of other physical events. So, if consciousness does causally influence the physical world, there is a danger that it will violate our best theories of physics. Either way, then, the Anti-Physicalist seems to face problems with accounting for the causal role of consciousness in the physical world.³⁹

A similar problem arises for Subjectivism. Suppose there is a perfect physical duplicate of me. According to Subjectivism, we would both be conscious, but we would differ in that only my experiences are *present* (or only I am CONSCIOUS).⁴⁰ We can now ask the analogous question: what is the causal role of *presence* (or CONSCIOUSNESS as opposed to consciousness)? The Subjectivist seems to have the same two uncomfortable options: either these extra properties are epiphenomenal, or else they play a causal role that might be in tension with our best physical theories.

4.5. Problem: What is it like to be others?

A third and final problem with Subjectivism is that it makes it mysterious what it’s like to be others. According to Objectivist Anti-Physicalists, there are two kinds of possible physical duplicates of me: zombie duplicates and conscious duplicates. According to Subjectivist Anti-Physicalists, there are three kinds of possible physical duplicates of me: zombie duplicates, (mere) conscious duplicates, and CONSCIOUS duplicates.

However, it is difficult to form a positive conception of what it is like to be a merely conscious being, which differentiates them from zombies and CONSCIOUS beings. The only characterization that Subjectivism gives of mere conscious beings is negative and indirect: what it is for x to be merely conscious is for (i) x to not be CONSCIOUS and (ii) from x’s perspective, x is CONSCIOUS. Condition (i) is a negative condition that doesn’t differentiate merely conscious beings from zombies, while condition (ii) seems indirect. While condition (ii) tells us what merely conscious

³⁹ For a defense of epiphenomenalism, see Yetter-Chappell (2022). For a version of interactionist dualism that seeks to be compatible with physics by associating consciousness with the collapse of the quantum wave function, see Chalmers and McQueen (2022). An alternative response is a Russellian Monist view, which accounts for mental causation by identifying the “quiddities” associated with fundamental physical properties with (proto-)phenomenal properties (e.g. see Alter and Coleman 2019).

⁴⁰ For shorthand, I’ll say that an experience is present if and only if the direct objects of awareness associated with that experience are present.

beings are like *relative to their own perspective*, it is natural to want to know what merely conscious beings are like *simpliciter*. What non-perspectival facts distinguish merely conscious beings from zombies? Alternatively, what do merely conscious beings have (and zombies lack) that *grounds* the fact that they have a perspective relative to which they are CONSCIOUS?

A similar kind of “grounding” problem arises for “A-theories” of time, which posit that there is something metaphysically privileged about the present time.⁴¹ According to standard A-theories, there are truths about the past using temporal operators, such as [it *was* the case that: dinosaurs exist]. However, a common worry about A-theories of time is that such truths do not have appropriate “truthmakers”. For example, on a Presentist version of the A-theory of time, according to which only present things exist, what is it about reality that makes it true that [it *was* the case that: dinosaurs exist], given that there are no dinosaurs to be found in reality? Just as A-theorists face the problem of what grounds “tensed” truths involving temporal operators, Subjectivists face the problem of what grounds “perspectival” truths involving perspectival operators. For those who are uncomfortable with ungrounded tensed facts, it is also natural to be uncomfortable with ungrounded perspectival facts.

5. Universalism

As we’ve seen, Subjectivism says that what makes an experience *mine* is that it has a certain qualitative property: either the experience is *present* or it is an EXPERIENCE rather than a mere experience. However, once we say that only some experiences are present and other experiences are not present, Subjectivism faces various difficulties. Among the various experiences in the future, what determines which ones will be present? What is the causal difference between consciousness and CONSCIOUSNESS? And can we really understand what it is to have an experience that is not present?

The fourth and final view that we will be considering is called *Universalism*.⁴² Universalism agrees with Subjectivism in that what makes an experience mine is a certain qualitative property, which we can call “presence”. However, it avoids the above problems by saying that *all* experiences are present. Moreover, it is not a mere accident that all experiences are present. Presence is an essential component to any possible experience. Presence is what makes an experience “live” (Hellie 2013: 309), “immediate” (Zuboff: 1990: 49), “given”, or “for-me” (Fasching 2016: 144, 146).

Because it will be useful to reserve “presence” for Subjectivism, let us stick to using the word “immediacy” for the kind of essential feature of experience that makes an experience mine according to Universalism. It is important to stress that the immediacy of experience is not some

⁴¹ For an overview of this problem, see Caplan and Sanson (2011).

⁴² Zuboff (1990) has defended Universalism and is responsible for the name of the view. For further discussion and defense of views in the vicinity of Universalism, see Kolak (2005), Fasching (2016), and Mørch (2024).

super-added “glow” that Universalists posit to all possible experiences. Rather, it is meant to be a familiar aspect of experience that we are acquainted with from the first-person perspective, and for which it is fully intelligible why all possible experiences, if they are to be experiences at all, must be “live”, “immediate”, “for-me”, “given”, etc.

Of course, Universalism faces an obvious objection: isn’t it clear that I’m not having everyone’s experiences? After all, I can’t read other people’s minds! I’ll start by addressing this objection (5.1), after which I will look at several arguments in favor of Universalism (5.2)

5.1. The Objection From Access

To start, Universalism is not meant to be revisionary of ordinary physical science. Universalists can agree that the world contains many distinct bodies and brains, and certain particular brains are associated with certain particular experiences and not others. Moreover, the Universalists can agree that in some sense there is a plurality of distinct “streams” of conscious experiences that are associated with the plurality of distinct brains. Facts about the differential causal connections between different experiences, such as that experiences associated with the same brain causally influence each other in much more direct, reliable, and speedy ways than experiences associated with distinct brains, will also be endorsed by the Universalist. Lastly, the Universalist will also endorse the same standard views about which experiential contents are “jointly introspectable”, in the sense that there can be a single introspective act that has those experiential contents as its objects. For example, David’s current visual and auditory experiences are jointly introspectable, but the experiential contents associated with distinct brains will not be jointly introspectable.⁴³

The Universalist will argue that the reason I cannot “read other people’s thoughts” is that there is a failure of “integration” between my experiences. Because none of David’s experiences are jointly introspectable with the experiences associated with other brains, there will be no introspective act that contains the experiences of David with the experiences associated with other brains. It cannot be used as an argument against Universalism that I don’t “notice” different people’s experiences, because Universalism *predicts* that there will be no act of “noticing” that has as its objects the experiences associated with distinct brains.

It is tempting to think that if Universalism were true, then things would seem radically different than they do seem. For example, one might initially imagine that, if Universalism were true, my visual field would contain something like a “split-screen” of everyone’s experiences, or maybe I would be radically confused and overwhelmed by the bombardment of billions of seemingly incompatible experiences. However, this is a mistake. Universalism does not add any extra experiential contents to the world. However, adding feelings of being overwhelmed or confused,

⁴³ For further discussion of the different senses in which one’s experiences can be “bounded” or “integrated”, see Roelofs (2024).

or adding an extra visual field as of a “split screen” would change the experiential contents that exist in the world. Universalists should therefore claim that things would seem just as they do seem if Universalism were true.

To supplement this kind of explanation, let us consider a few examples of non-integrated experiences that support Universalism.

First, if one is an Eternalist who believes that past, present, and future entities are equally real, then it is natural to think that there are past and future experiences that are (i) equally *mine*, yet (ii) I do not have joint-access of my past, present, and future experiences. The Universalist is merely claiming that the same sort of thing happens for the experiences of “other people”: those experiences are equally mine (and yours), but I (you) do not have joint-access of David’s experience and the experiences associated with other brains.

Second, consider a time travel case. Imagine you use a time travel machine to meet your former self. Then, you would be having two experiences *at the very same time* that are not jointly accessible. Anyone who thinks that we can rule out Universalism merely by introspection (on the grounds that all my present experiences must be jointly accessible), must also say that they can rule out time travel by introspection.⁴⁴

Third, consider “split-brain” patients, whose corpus callosum has been severed for medical purposes, preventing their left and right hemispheres of their cerebral cortex from communicating directly.⁴⁵ It is a matter of controversy what the conscious life of such patients is like. According to one view, there is a single subject of experience whose experiential contents (associated with their left and right hemispheres) are not jointly integrated. According to another view, there are two distinct subjects of experience with two separate streams of consciousness. Alternatively, perhaps it is indeterminate whether there is one or two subjects of experience. However, so long as the first interpretation is a coherent one, then the kinds of failure of joint-access that Universalists posit seem to be similarly coherent.

Zuboff (1990) appeals to a thought-experiment involving a split-brain case that is meant to motivate the coherence of Universalism. Imagine the two hemispheres of your brain are severed, and one of the hemispheres listens to a podcast, while the other hemisphere listens to a song. Now, imagine reintegrating the two hemispheres. After the re-integration, it is plausible that it will seem as if *you* listened to the podcast and *you* listened to the song. Since these events happened at the same time, it will seem to you as if you had two “incompatible” experiences at the very same time, which is what is happening right now if Universalism is correct.

⁴⁴ One could also be a stage theorist, so that, strictly speaking, your “earlier” self and “later” self in cases of time travel are numerically distinct. Still, one would have to rule out enduring time travelers by introspection.

⁴⁵ For further discussion of split-brain cases, see Schechter (2018).

Fourth, there are arguably also “ordinary” examples of lack of integration, where a conscious subject has access to P, access to Q, but does not have joint-access of both P and Q. A widely discussed example of this is a famous experiment by George Sperling (1960), in which a subject is briefly presented with a matrix (for 250 milliseconds) consisting of three rows with four letters each. After the matrix is presented, a tone sounds, which is meant to indicate whether the subject is supposed to report the contents of the top, middle, or bottom row. When the subjects are asked to report the contents of a particular row, on average they correctly report 3.3 of the four letters in that row. However, when the subjects are asked to report the contents of the entire matrix, on average they can only correctly report 4.5 of the twelve letters. One could naturally describe this situation as one in which the information in any particular row is accessible to the subject, but the subject does not have joint access to all of the information in all of the rows.

Bayne and Chalmers (2003) use the example of the Sperling experiment to illustrate a failure of “access unity”, where a subject can have two mental states that are not jointly-accessible. However, they do not think that anything like the Sperling case shows a failure of “phenomenal unity”, where two conscious states are phenomenally unified when they are jointly experienced: when there is something it is like to be in both states at once. When two conscious states are phenomenally unified, there will be a phenomenal state (corresponding to the conjoint what-it-is-like-ness) that *subsumes* each of the two conscious states. While Universalism implies massive failures of access unity between my conscious experiences, Universalism is compatible with the view that all my experiences are phenomenally unified. That is, it is compatible with (but does not entail) the view that there is a single “global phenomenal state” that I have that subsumes all the conscious states in reality, even though such a unified phenomenal state is not access-unified.

5.2. Arguments for Universalism

Why believe Universalism?

First, unlike Type-A Eliminativist and Type-A Relativist views, Universalism acknowledges non-relative facts about which experiences are mine. For example, Universalism says that I am in pain *simpliciter*.

Second, unlike Type-B Objectivism, Universalism is responsive to the epistemic gap between third-personal and first-personal facts in a similar way that Subjectivism is. According to Subjectivists, it is a priori that if an experience is *present*, then that experience is mine. According to Universalists, the same is true: it is a priori that if an experience is present (or “live”, “immediate”, given”, “for-me”, etc.), then it is mine. The quality of “presence” that Subjectivists and Universalists are getting at is supposed to be a quality of experience that I know directly from my own first-person case. Subjectivists are not willing to grant the quality of presence to all experience, presumably because of something like the “objection of access”. However,

Universalists think there are good reasons for thinking the objection of access is mistaken. Moreover, in not granting “presence” to all experience, there are reasons for thinking that the Subjectivist account of the mental lives of others is problematic (section 4.5).

Third, unlike Subjectivism, Universalism (i) has no need for fundamental bridge laws between third-personal and first-personal facts, (ii) does not face the problem of what is the causal difference between consciousness and CONSCIOUSNESS (since the Universalist collapses the distinction between the two categories), and (iii) avoids the problem of positing a mysterious tripartite distinction between zombies, merely conscious creatures, and CONSCIOUS creatures. According to Universalism, consciousness is the same as CONSCIOUSNESS, because all consciousness is suitably immediate/present. Moreover, insofar as one is worried by the arbitrariness of Subjectivism (why is David the special one?), Universalism avoids this problem as well. Lastly, there is no additional skeptical challenge of how I know that I am “present” given Universalism, over and above the skeptical challenge of how I know that I am conscious.

Fourth, like Subjectivism, Universalism can provide binary, precise, and non-conventional personal identity facts, so long as there must be binary, precise, and non-conventional facts about whether a being is conscious. In any puzzling case of personal identity (brain transplants, fission, teletransportation, etc.), I will be “on the other side” just in case a conscious being is “on the other side”. With respect to the question of whether consciousness is susceptible to vagueness or indeterminacy, any Anti-Physicalist who believes that consciousness is fundamental and who wishes to avoid fundamental indeterminacy is committed to the view that it cannot be indeterminate whether a being is conscious.⁴⁶

Fifth, unlike Subjectivism, Universalism provides principled and non-arbitrary answers to puzzle cases about personal identity. For example, consider a case of fission where I fission into Lefty and Righty. Subjectivism implies that there must be a binary, precise, and non-conventional fact about whether I will have the experiences of Lefty or Righty (or neither), but it leaves open what this fact will be. It is compatible with Subjectivism that I (arbitrarily) become Lefty, and it is compatible with Subjectivism that I (arbitrarily) become Righty. Similarly, if I entered a teletransportation machine, Subjectivism is silent on whether I will come out on the other side. However, Universalism provides specific non-arbitrary verdicts in all of these cases. Whenever there is a conscious being on the other side, I will be there.

Sixth, like Subjectivism, Universalism is compatible with the intuition that it is both conceivable and possible (and indeed actual!) for me to have the experiences of other conscious creatures other than David.

⁴⁶ For further discussion of fundamental indeterminacy, see Torza (2023). For further discussion of whether it can be vague or indeterminate whether a being is conscious, see Schwitzgebel (2023).

However, it should be acknowledged that Universalism rejects our “negative” first-personal conceivings: it seems conceivable (and actual!) that I’m *not* having the experiences of anyone other than David. Here, Universalists must provide an error theory. By responding to the objection of access, Universalists deny that I have good grounds for thinking that I am not having the conscious experiences of everyone else. Ultimately, the cumulative case for Universalism, and the *a priori* arguments against alternative views, is meant to motivate the view that it’s not even conceivable that I don’t have the experiences of others. For example, the Universalist can argue that it is *a priori* that all conscious experiences are equally immediate/present (partly for the reasons in section 4.5), contrary to Subjectivism. However, together with Subjectivists, they can argue that it is *a priori* that an experience is mine if it is immediate/present. They can also give *a priori* arguments against alternative views, such as arguing against Type-B Objectivism by appealing to **Fundamental Scrutability** (or the thesis that conceivability entails possibility “when considered as actual”).

Perhaps one of the strongest error theories against the conceivability of absent worlds that are third-personally just like the actual world is that, on reflection, it’s deeply mysterious what the difference is supposed to be between the actual world and its corresponding absent world. When God was making the world, what *more* did God have to do to make David’s objectively characterized experience *mine*? It’s very unclear. God didn’t have to change the qualitative character of David’s experience to make it mine, since the only difference between the actual world and the absent world is that *the very same qualitative experiences* had by David are being had by me. Moreover, on the standard view, no qualitative change in experience seems sufficient for an experience to be mine, since there could always be a qualitative duplicate twin David* whose experiences are not mine.⁴⁷ As Klawon (1987) puts it, the difference between the actual world and

⁴⁷ One could deny that there is a distinct metaphysically possible absent alternative to the actual world (on the grounds that I = David). However, this would be to accept Type-B Objectivism, which has the problems discussed in sections 3.1 and 3.2. Alternatively, one might reject Type-B Objectivism and say that there is a “haecceitistic” difference between the actual world and its corresponding absent world: in the actual world, my haecceity plays the “David-role” and in the absent world, some other haecceity plays the “David-role”. However, haecceitistic differences don’t seem able to account for first-personal conceivable differences. For one, such a view is in tension with popular views about essences when one takes seriously a conceivability-possibility link: it is conceivable/possible that I have different parents (contrary to origin essentialism), and it is conceivable/possible that I am a conscious being who is not a human (contrary to kind essentialism). Subjectivists and Universalists can understand the conceivability and possibility of such claims (by reference to the qualitative property of presence/immediacy) without endorsing any controversial claims about essences. Second, such a view cannot accommodate that it is conceivable that I have the conscious experience of someone other than David tomorrow, while holding fixed the qualitative truth that no particular body/brain/soul “teleported” from David to someone else. Third, it’s unclear how such a view can accommodate **Fundamental Scrutability**. In order for it to be scrutable that I am (uniquely) David given the fundamental facts, the fundamental facts would seem to have to be “asymmetrically oriented” towards David. However, *everything* has a corresponding haecceity: tables, chairs, dogs, people, etc. So, haecceitistic facts are “symmetric” in a way that does not seem to allow for the asymmetric conclusion that I am David. In other words, there is an explanatory question that this view doesn’t seem to address: out of the many haecceities there are, what explains why one of the haecceities is *mine* rather than some other one (note that Subjectivism/Universalism can explain why a particular haecceity happens

its corresponding absent world is “a paradoxical difference, which makes no difference” (52). Universalism avoids this “paradoxical difference” by saying that there is nothing more that God needed to do to make David’s experiences mine. David’s experiences, just like all other experiences, are mine simply in virtue of their immediacy or first-personal-givenness.

Seventh, and finally, Zuboff (1990) has given an interesting probabilistic argument for Universalism. According to Non-Universalist views, there are various epistemic possibilities consistent with the actual third-personal facts, but which differ first-personally. For example, it is a priori consistent with the third-personal facts that I don’t exist, and it is a priori consistent with the third-personal facts that I have the conscious experiences of any objectively characterized conscious being. There is therefore a sense in which it is incredibly *improbable* that I would be having David’s experiences. Given just the third-personal facts, there are billions of other people whose experiences I might have had instead. However, given Universalism, it is not at all surprising that I would be having David’s experiences given the third-personal facts. Universalism, together with the third-personal facts, *guarantees* that I would be having David’s experiences. So, the fact that I am having David’s experiences is very strong empirical evidence in favor of Universalism.

This kind of argument can be formalized in a Bayesian framework. If we let T represent the totality of my third-personal evidence, U represent Universalism, and F represent the totality of my first-personal evidence (e.g. all facts of the form “I am F” that are part of my evidence), then if we let Cr represent my prior credence function⁴⁸, then:

- (1) $Cr(F \mid T \ \& \ U) = 1$
- (2) $Cr(F \mid T \ \& \ \sim U) \ll 1$
- (3) Therefore, $Cr(F \mid T \ \& \ \sim U) \ll Cr(F \mid T \ \& \ U)$

Using standard Bayesian Conditionalization, (3) implies that, given the background third-personal evidence, my first-personal evidence is *very* strong confirmation for Universalism. This is for the simple reason that Universalism *entails* that I would exist and be having David’s experiences given the third-personal facts, while standard views imply that it is a priori very improbable that I would exist and be having David’s experiences given the third-personal facts.

to correspond to me: it corresponds to me because it is having present/immediate experiences). Lastly, there are strong arguments for Anti-Haecceitism (e.g. see Dasgupta 2017), and even bracketing the truth of Anti-Haecceitism, it is natural to think that first-personal truths should be accommodated in a way that at least remain neutral on the question of whether Haecceitism is true. For further discussion of why conceivable first-personal differences should not be understood as haecceitistic differences, see Lewis (1979: 522-524) and Weber (2025).

⁴⁸ Here I am understanding one’s prior credence function not as one’s “temporally prior” credence function (the credence function one happened to have at the beginning of one’s life), but rather as one’s “hypothetical prior” credence function, which encodes one’s “evidential standards” or one’s assessment of how *a priori* plausible various epistemic possibilities are. For further discussion of the nature of hypothetical priors, see Meacham (2016).

Perhaps the most natural way to resist this argument is to say that part of my first-personal evidence is the “negative” evidence that I’m *not* having the experiences of other people besides David. Then, my first-personal evidence would *refute* Universalism, rather than supporting it. However, this objection goes back to the objection of access. It’s true that if my first-personal evidence includes such “negative” evidence, then Universalism is refuted. However, what this probabilistic argument shows is that, if the Universalist is right that my first-person evidence only includes the “positive” evidence that I exist and have David’s experiences, then not only is my first-person evidence *consistent* with Universalism, but it also massively supports Universalism.

5.3 *What am I?*

There is a lingering question about how Universalism should be understood. The official formulation of Universalism is simply: all experiences are mine, and they are mine in virtue of their immediacy, which is a necessary aspect of all conscious experience. However, this doesn’t address what “I” refers to on a Universalist view. In this section, I’ll briefly consider two potential responses to this question. Which version is ultimately more plausible will depend on one’s background theoretical commitments.

One version of Universalism, defended by Fasching (2016), is a view that requires rejecting Physicalism.⁴⁹ On this view, there is only one subject of experience, and I am the one subject of experience. Strictly speaking, nothing else has experiences except for this one subject. So, if there are physical brains and bodies, they are not subjects of experience (although presumably if there are brains and bodies, then the experiences had by the one subject are correlated with the properties of brains and bodies). This view is closely associated with recently defended versions of “cosmopsychism” or “cosmic idealism”, according to which the physical universe is ultimately grounded in, or is a representation of, the experiences of a single entity – something like a “cosmic mind”.⁵⁰ Historically, this view is closely associated with the Indian philosophy of Advaita Vedanta. One central tenant of Advaita Vedanta is that “Atman is Brahman”. Here, “Atman” refers

⁴⁹ Fasching (2016: 144) denies that physical things can be subjects of experiences when he writes “contrary to what might first suggest itself, it is not e.g. the organism which can be the experiencer in the sense relevant here, because *for* the organism, understood as a purely physical entity,...there is nothing at all (if we reject – as I think we should – a materialistic understanding of consciousness, i.e. if for-ness itself is nothing we can find as some physical feature of reality”. Later, in contrasting his position with Zuboff’s (1990) position, he writes “the I is rather to be understood as a substance (or, in the Vedantic view: *the* substance)” (152). However, Fasching also expresses skepticism that the I can be understood using traditional ontological categories of objects and properties, and he elsewhere describes it as a “dimension”: “the I is not an object in the world but rather the for-whom of all objectivity, the where of the stretching-out of the objective world itself. A where which is not a place within the world but the ‘primal place’ of the world itself with all its ‘places’: the dimension of world-manifestation and therefore not locatable within the world manifesting to me. So in some sense the world is rather in the I than the I in the world.” (148).

⁵⁰ For recent defenses of such a view, see Goff (2017), Kastrup (2018), Chalmers (2019), and Shani (2022). However, it should be noted that some of these contemporary views allow for there to be a plurality of distinct conscious subjects that are “part” of the cosmic mind. Universalists should reject this.

to the “self” or “I”, while “Brahman” refers to something like the single ground of all reality, often described as “universal consciousness”.⁵¹ On this version of Universalism, there is a sense in which “I am David”, because I am the subject of the experiences that are associated with David. Still, however, I am not numerically identical to David, if David is understood to be a local, biological organism.

Another version of the view, defended by Zuboff (1990), does not posit a novel entity that is the “universal subject” of all experiences.⁵² Although it is still true that every experience is mine on this view, the fact that an experience is mine should not be analyzed as there being a single subject of that experience that is identical to me (since this analysis quickly leads to the above “universal subject” view). For example, consider a “Humean bundle theory” account of the self, according to which there are no “subjects” that are the bearers of experiences, but rather experiences are merely “bundled” together by various relations (e.g. perhaps a relation of “co-consciousness” or various causal/spatial relations). One might think that, even on such a Humean view, it is intelligible to say that some experiences are *mine* in a way that does not presuppose the existence of a separate “subject” of experience. If such a view can be made sense of, then a Universalist can say that, in that same non-subject-presupposing sense of *being mine*, all experiences are mine. On this view, there is still a sense in which “I am David”, but this truth should be understood as only meaning that the experiences associated with David are mine (*simpliciter*).

5.4. *Universalism and Reference*

Let us see how Universalism handles our initial “lake” case.

Universalists can of course accept that the case involves two physical bodies, and one of the bodies is next to a lake and the other is not. However, once it is stipulated that both agents are conscious, then, according to Universalism, there are no first-personal facts left to be ignorant about. Given Universalism, the third-personal facts entail that (i) I am having the conscious experiences of the person next to the lake and (ii) I am having the conscious experiences of the person not next to the lake. It is an illusion that I am exclusively having the conscious experiences of only one of them.

Zuboff (2009) discusses an interesting hypothetical case in which even non-Universalists should give structurally similar verdicts. Consider a version of the case where neither of the agents in the two rooms have brains of their own. Instead, there is a single “brain in a vat” that is connected to the two bodies. This brain in a vat receives (identical) sensory signals from the two bodies, and it

⁵¹ For another historical example of a view according to which we are all the same mind, see Hung (2024).

⁵² Zuboff (1990: 51) writes: “When I claim that, throughout the various bodies, minds, and experiential contents of all the world, there is but one self and one consciousness, I am not positing the existence of a strange new substance, any more than I would be positing the existence of a strange new book if I claimed that various books were all instances of writing”. For Zuboff, the “one self” corresponds to something like the “type” or “universal” associated with immediacy.

responds to those signals in the very same way that it would respond if the brain was in either body. Moreover, the brain “controls” both bodies in the same way that it would control them if the brain was in either body: if the brain forms the intention of raising its left arm, then the left arms of the two bodies go up. It will vividly seem to the brain as if it has a single body, and it will vividly seem to the brain that there is a definite fact of the matter about whether the single body that it has is next to a lake or not. However, this would be an illusion. The brain in some sense has “both” bodies.

We can further imagine that, in the brain in a vat case, the brain raises its arm(s) and points to the (both) door(s) of its room(s) and asks: is *that door* next to the lake? It will vividly seem to the brain as if it is pointing to a particular door, but this would also be an illusion. It is pointing to both doors, and so “this door” fails to uniquely refer to a particular door.

Going back to the original lake case, suppose I point to one of the doors and ask: is *that door* next to the lake? The Universalist should say that I have pointed to both doors, and although it vividly seems as if I am only pointing to one door, that is an illusion. The phrase “the door that I am pointing at” does not successfully uniquely refer.⁵³

However, in less idealized settings, indexicals and demonstratives can successfully uniquely refer for the Universalist. Suppose that the agent next to the lake is wearing a red shirt, while the agent not next to the lake is wearing a blue shirt. Then, I can still ask: where am I wearing a red shirt? Even for the Universalist, there will be a definite answer: I am wearing the red shirt next to the lake. Similarly, I can point to one of the doors and ask: is the door that I am pointing to with the red-shirted body next to the lake? Again, there will be a definite answer. It is natural for the Universalist to say that, in ordinary day-to-day contexts, questions about “I”, “here”, “this”, etc. are implicitly qualified in something like this way. In daily life, when I am asking “where am I?”, I am implicitly asking (when using David’s brain) “where am I Davidish?”, where “Davidish” is something like a shorthand for the third-personal truths that I (with David’s brain) associate with a particular body (David). When there aren’t multiple David-ish people, then there will be no failure of reference. It is only in highly idealized cases like our lake case where there will be failures of reference.⁵⁴

⁵³ Similar issues to the ones discussed in note 33 become relevant here, concerning whether there can be demonstrative truths that are not scrutable from both third-personal and first-personal truths.

⁵⁴ There are interesting interactions with Universalism and the “Boltzmann Brain” problem. Popular cosmological theories seem to imply that any human experience will be (nearly) duplicated by many Boltzmann Brains. If this is true, then Universalism implies that I am having ordinary human experiences and Boltzmann Brain experiences, and attempts to uniquely refer to a particular human body/brain with indexicals and demonstratives will not be successful.

6. Universalism and the Metaphysics of Consciousness

Unlike every other view of the first-person that we have considered so far, Universalism does not seem to have a straightforward analogous position in the metaphysics of consciousness. For example, Type-A Eliminativism about the first-person naturally corresponds to a Type-A Physicalist view that eliminates consciousness. Type-A Relativism about the first-person naturally corresponds to a Type-A Physicalist view that accepts a deflationary analysis of consciousness in terms of (say) physical behaviors and functions. Type-B Objectivism naturally corresponds to Type-B Physicalism, and Subjectivism naturally corresponds to standard Anti-Physicalist views. In closing, I'll briefly consider what the natural analog of Universalism is in the metaphysics of consciousness.

What is distinctive about Universalism is that, like Type-A Objectivism, it denies that there is an epistemic gap between the third-person and the first-person (because it is a priori that every third-personally characterized conscious experience is mine). In particular, Universalism denies that there is a conceivable absent world that is third-personally just like the actual world. However, Universalism still tries to “take the first-person seriously” in that it acknowledges that there are non-relative facts about the first-person (e.g. “I exist” is true simpliciter, or at least “David’s experiences are my experiences” is true simpliciter).

It turns out that there is a version of Idealism (the view that all facts about concrete reality are grounded in facts about consciousness) that shares these two distinctive aspects of Universalism. In his *Appearance and Reality*, Bradley (1893) endorses the relevant version of Idealism, which he describes as follows:

We perceive, on reflection, that to be real, or even barely to exist, must be to fall within sentience. Sentient experience, in short, is reality, and what is not this is not real. We may say, in other words, that there is no being or fact outside of that which is commonly called psychical existence. Feeling, thought, and volition (any groups under which we class psychical phenomena) are all the materials of existence, and there is no other material, actual or even possible. This result in its general form seems evident at once...[Any] fact that falls elsewhere seems, in my mind, to be a mere word and a failure, or else an attempt at self-contradiction. It is a vicious abstraction whose existence is meaningless nonsense, and is therefore not possible (144-145).

In other words, it is natural to interpret Bradley as endorsing the view that (i) it is not possible for something to exist without being conscious, and (ii) the only possible fundamental properties are phenomenal properties (“...(any groups under which we class psychical phenomena) are all the materials of existence”). In fact, since he describes these views as “evident at once”, and since he describes denying these views as “meaningless nonsense”, it seems fair to say that Bradley thinks

it is inconceivable for (i) and (ii) to be false. Let us call such a view “Modal Idealism”.⁵⁵ Modal Idealism clearly “takes consciousness seriously”, but nonetheless rejects the conceivability of zombie worlds. If we think of physical truths as only describing “structural” truths about reality, where structural truths are truths that are expressible using only logical/mathematical, causal/nomic, and spatiotemporal notions⁵⁶, then Modal Idealism will reject the conceivability of a world that is physically just like ours yet lacks consciousness.

Still, although zombie worlds might not be conceivable according to Modal Idealism, it might still be the case that there are conceivable possibilities that are physically just like the actual world, yet differ in their phenomenology. For example, perhaps it is conceivable that there is a world just like ours that is “hedonically inverted”, so that pleasurable experiences are associated with aversive dispositions and painful experiences are associated with attractive dispositions.

Supplementing Modal Idealism with the “phenomenal powers” view, according to which there is a metaphysically necessary connection between phenomenal experiences and their causal role (which is inconceivably otherwise), successfully rules out the conceivability and possibility of hedonically inverted worlds, as well as any other worlds where actual phenomenal qualities are associated with different causal roles.⁵⁷

Still, the phenomenal powers view is consistent with the possibility that distinct phenomenal qualities might be necessarily connected with the very same causal role. For example, perhaps it is initially intuitive that phenomenal red and phenomenal green are associated with the very same causal role. If so, then it is consistent with both Modal Idealism and the phenomenal powers view that there might be a conceivable world that is just like the actual world except “phenomenal colors” are inverted.

In response, Strawson (2024) has defended the stronger view that “no difference of quality is possible without some difference of structure” (124). According to this principle, a complete “structural” description of a world metaphysically necessitates a complete description of the underlying qualities that realize that structure. Since elsewhere Strawson also endorses the view that “whatever is genuinely conceivable is possible” (137), Strawson’s view implies that it is inconceivable for there to be two worlds that are structurally identical yet different with respect to their underlying qualities.

⁵⁵ I defend this view in Builes (2024c).

⁵⁶ For further discussion of how to understand “structural” truths, see Alter (2015) and Goff (2017).

⁵⁷ For defense of the phenomenal powers view, see Langsam (2011), Mørch (2017, MS), and Pallies (2022). The phenomenal powers view is one version of the general “grounding view of powers”, according to which powers are grounded in categorical properties or qualities. For a defense of this more general view, see Tugby (2022), Kimpton-Nye (2021), and Builes (2022, 2023, 2024b).

Combining this Strawsonian view with Modal Idealism results in a view that (i) takes consciousness seriously, but (ii) denies that there is an epistemic gap between structural truths and phenomenal truths.⁵⁸ Such a view therefore serves as the analogous view to Universalism with respect to the metaphysics of consciousness.

References

- Alter, Torin. 2015. The Structure and Dynamics Argument Against Materialism. *Noûs* 50, no. 4: 794-815.
- Alter, Torin and Sam Coleman. 2019. Russellian monism and mental causation. *Noûs* 55, no. 2: 409-425.
- Balog, Katalin. 2011. In Defense of the Phenomenal Concept Strategy. *Philosophy and Phenomenological Research* 84, no. 1: 1-23.
- Bayne, Timothy and David Chalmers. 2003. What is the unity of consciousness? In *The Unity of Consciousness: Binding, Integration, and Dissociation*, edited by Axel Cleeremans, 23-58. Oxford: Oxford University Press.
- Block, Ned and Robert Stalnaker. 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review* 108: 1-46.
- Braddon-Mitchell, David and Kristie Miller. 2004. How to be a conventional person. *The Monist* 87, no. 4: 456-474.
- Braddon-Mitchell, David and Kristie Miller. 2020. Surviving, to some degree. *Philosophical Studies* 177, no. 12: 3805-3831.
- Bradley, Francis. 1893. *Appearance and Reality: A Metaphysical Essay*. London: Oxford University Press.
- Bricker, Phillip. 2006. Absolute Actuality and the Plurality of Worlds. *Philosophical Perspectives* 2006, no. 20: 41-76.
- Builes, David. 2020. Time-Slice Rationality and Self-Locating Belief. *Philosophical Studies* 117, no. 10: 3033-3049.
- Builes, David. 2021. The World Just Is The Way It Is. *The Monist* 104, no. 1: 1-27.
- Builes, David. 2022. The Ineffability of Induction. *Philosophy and Phenomenological Research*

⁵⁸ Note that, without Modal Idealism, the Strawsonian view is compatible with a Type-A Eliminativist view about consciousness.

- 104, no. 1: 129-149.
- Builes, David. 2023. A Humean Non-Humeanism. *Philosophical Studies* 180, no. 3: 1031-1048.
- Builes, David. 2024a. Eight Arguments for First-Person Realism. *Philosophy Compass*. e12959.
<https://doi.org/10.1111/phc3.12959>
- Builes, David. 2024b. How to ground powers. *Analysis* 84 no. 2: 231-238.
- Builes, David. 2024c. Modal Idealism. In *Oxford Studies in Philosophy of Mind: Volume 4*, edited by Uriah Kriegel, 46-107. Oxford: Oxford University Press.
- Builes, David and Trevor Teitel. 2022. Lawful Persistence. *Philosophical Perspectives* 36, no. 1: 5-30.
- Caplan, Ben and David Sanson. Presentism and Truthmaking. *Philosophy Compass* 6, no. 3: 196-208.
- Carruthers, Peter and Benedicte Veillet. 2007. The phenomenal concept strategy. *Journal of Consciousness Studies* 14, no. 9-10: 212-236.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, David. 2002. Consciousness and its Place in Nature. In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David Chalmers, 247-272. Oxford University Press.
- Chalmers, David. 2006a. The Foundations of Two-Dimensional Semantics. In *Two-Dimensional Semantics*, edited by Manuel Garcia-Carpintero and Josep Macià, 55-140. New York: Oxford.
- Chalmers, David. 2006b. Phenomenal concepts and the explanatory gap. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. New York: Oxford University Press.
- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chalmers, David. 2018. The Meta-Problem of Consciousness. *Journal of Consciousness Studies* 25, no. 9-10: 6-61.
- Chalmers, David. 2019. Idealism and the Mind-Body Problem. In *The Routledge Handbook of Panpsychism*, edited by William Seager, 353-373. New York: Routledge.
- Chalmers, David. 2020. Debunking Arguments for Illusionism about Consciousness. *Journal of Consciousness Studies* 27, no. 5-6: 258-281.

- Chalmers, David and Frank Jackson. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110, no. 3: 315-361.
- Chalmers, David and Kelvin McQueen. 2022. Consciousness and the Collapse of the Wave Function. In *Consciousness and Quantum Mechanics*, edited by Shan Gao, 11-63. New York: Oxford University Press.
- Cornell, David. 2016. Taking monism seriously. *Philosophical Studies* 173 no. 9: 2397-2415.
- Cutter, Brian and Dustin Crummett. 2025. Psychophysical Harmony: A New Argument for Theism. In *Oxford Studies in Philosophy of Religion Volume 11*, edited by Lara Buchak and Dean Zimmerman, 33-71. New York: Oxford University Press.
- Cresswell. 2012. *The World-Time Parallel: Tense and Modality in Logic and Metaphysics*. Cambridge: Cambridge University Press.
- Dasgupta, Shamik. 2017. Quality and Structure. In *Current Controversies in Metaphysics*, edited by Elizabeth Barnes, 7-23. New York: Routledge.
- Dennett, Daniel. 1991. *Consciousness Explained*. Boston: Little, Brown, & Co.
- Emery, Nina. 2018. Actualism without Presentism? Not by way of the Relativity Objection. *Noûs* 53, no. 4: 963-986.
- Emery, Nina. 2020. Actualism, Presentism, and the Grounding Objection. *Erkenntnis* 85, no. 1: 23-43.
- Fasching, Wolfgang. 2016. The Non-Plurality of the I: On the Question of the Ultimate Subject of Experience. *Journal of Consciousness Studies* 23, no. 1-2: 140-157.
- Fine, Kit. 2005. Tense and Reality. In *Modality and Tense: Philosophical Papers*, 261-320. Oxford: Oxford University Press.
- Forbes, Graeme. 2015. The Growing Block's past problems. *Philosophical Studies* 173, no. 3: 699-709.
- Frankish, Keith. 2016. Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies* 23, no. 11-12: 11-39.
- Glazier, Martin. 2020. Being Someone Else. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, edited by Enoch Lambert and John Schwenkler, 37-51. Oxford: Oxford University Press.
- Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York: Oxford University Press.
- Hare, Caspar. 2009. *On Myself, and Other, Less Important Subjects*. Princeton: Princeton

University Press.

Hare, Caspar. 2010. Realism about Tense and Perspective. *Philosophy Compass* 5, no. 9: 760-769.

Hellie, Benj. 2013. Against Egalitarianism. *Analysis* 73, no. 2: 304-320.

Hung, Jenny. 2024. We are the Same Mind! A Study of Zongmi's Idea of the True Mind.

Philosophy East and West 75, no. 4.

Ismael, Jenann. 1999. Science and the phenomenal. *Philosophy of Science* 66, no. 3: 351-369.

Jackson, Frank. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127-136.

Jackson, Frank and Daniel Stoljar. 2020. Understanding self-ascription. *Mind and Language* 35, no. 2: 141-155.

Kammerer, François. 2019. The illusion of conscious experience. *Synthese* 198, no. 1: 845-866.

Kastrup, Bernardo. 2018. The Universe in Consciousness. *Journal of Consciousness Studies* 25, no. 5-6: 125-155.

Kimpton-Nye, Samuel. 2021. Reconsidering the Dispositional Essentialist Canon. *Philosophical Studies* 178, no. 10: 3421-3441.

Klawonn, Erich. 1987. The "I": On the ontology of first personal identity. *Danish Yearbook of Philosophy* 24: 43-76.

Kolak, Daniel. 2005. *I Am You: The Metaphysical Foundations for Global Ethics*. Dordrecht: Springer.

Kwon, Hongwoo. 2017. Mary and the Two Gods: Trying Out an Ability Hypothesis.

Philosophical Review 126, no. 2: 191-217.

Langsam, Harold. 2011. *The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection*. MIT Press.

Levin, Janet. 2018. Once More Unto the Breach: Type B Physicalism, Phenomenal Concepts, and the Epistemic Gap. *Australasian Journal of Philosophy* 97, no. 1: 57-71.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

Lewis, David. 1976. Survival and Identity. In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 17-40. Berkeley, CA: University of California Press.

Lewis, David. 1979. Attitudes de dicto and de se. *Philosophical Review* 88, no. 4: 513-543.

Lewis, David. 1988. What experience teaches. *Proceedings of the Russellian Society* 13: 29-57.

- Lipman, Martin. 2023a. Standpoints: A Study of a Metaphysical Picture. *Journal of Philosophy* 120, no. 3: 117-138.
- Lipman, Martin. 2023b. Subjective Facts about Consciousness. *Ergo: An Open Access Journal of Philosophy* 10: 530-553.
- List, Christian. 2023. The many-worlds theory of consciousness. *Noûs* 57: 316-340.
- Meacham, Christopher. 2008. Sleeping beauty and the dynamics of de se beliefs. *Philosophical Studies* 138, no. 2: 245-269.
- Meacham, Christopher. 2016. Ur-Priors, Conditionalization, and Ur-Prior Conditionalization. *Ergo: An Open Access Journal of Philosophy* 17: 444-492.
- Merlo, Giovanni. 2016. Subjectivism and the Mental. *Dialectica* 70, no. 3: 311-342.
- Merlo, Giovanni. 2021. The Metaphysical Problem of Other Minds. *Pacific Philosophical Quarterly* 102, no. 4: 633-664.
- Metzinger, Thomas. 2009. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. New York: Basic Books.
- Mørch, Hedda Hassel. 2017. The Evolutionary Argument for Phenomenal Powers. *Philosophical Perspectives* 31: 293 – 316.
- Mørch, Hedda Hassel. 2024. Does Panpsychism Mean that “We Are All One”? *Journal of Consciousness Studies* 31, no. 9-10: 88-112.
- Mørch, Hedda Hassel. MS. Phenomenal Powers.
- Olson, Eric. 1997. *The Human Animal: Personal Identity Without Psychology*. New York: Oxford University Press.
- Pallies, Daniel. 2022. The Pleasure Problem and the Spriggean Solution. *Journal of the American Philosophical Association* 8, no. 4: 665-684.
- Papineau, David. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Perry, John. 2001. *Knowledge, Possibility, and Consciousness*. MIT Press.
- Rini, Adrienne and Max Cresswell. 2012. *The World-Time Parallel: Tense and Modality in Logic and Metaphysics*. Cambridge: Cambridge University Press.
- Roelofs, Luke. 2020. Consciousness, Revelation, and Confusion: Are Constitutive Panpsychists Hoist by their Own Petard? *Dialectica* 74, no. 1: 61-93.
- Roelofs, Luke. 2024. The Varieties of (Un)Boundedness. *Journal of Consciousness Studies* 31, no.

- 9: 42-66.
- Schaffer, Jonathan. 2007. From nihilism to monism. *Australasian Journal of Philosophy* 85, no. 2: 175-191.
- Schaffer, Jonathan. 2017. The Ground Between the Gaps. *Philosophers' Imprint* 17, no. 11: 1-26.
- Schechter, Elizabeth. 2018. *Self-Consciousness and "Split" Brains: The Minds' I*. New York: Oxford University Press.
- Schwitzgebel, Eric. 2023. Borderline consciousness, when it's neither determinately true nor determinately false that experience is present. *Philosophical Studies* 180, no. 12: 3415-3439.
- Smith, Sean. 2021. The Negation of Self in Indian Buddhist Philosophy. *Philosophers' Imprint* 21, no. 13: 1-23.
- Shani, Itay. 2022. Cosmopsychism, Coherence, and World-Affirming Monism. *The Monist* 105, no. 1: 6-24.
- Sider, Theodore. 2013. Against Parthood. In *Oxford Studies in Metaphysics Volume 8*, edited by Karen Bennett and Dean Zimmerman, 237-293. Oxford: Oxford University Press.
- Solomyak, Olla. 2024. How to be a perspectival pluralist. *Philosophy Compass* 19, no. 5: e12980.
- Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*. Oxford University Press.
- Stoljar, Daniel. 2010. *Physicalism*. New York: Routledge.
- Stoljar, Daniel. 2018. Review of Philip Goff, *Consciousness and Fundamental Reality*. *Notre Dame Philosophical Reviews*.
- Strawson, Galen. 2006. Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13, no. 10-11: 3-31.
- Strawson, Galen. 2024. *Stuff, Quality, Structure: The Whole Go*. Oxford: Oxford University Press.
- Spencer, Jack. 2022. Relativity in a Fundamentally Absolute World. *Philosophical Perspectives* 36, no. 1: 305-328.
- Swinburne, Richard. 2019. *Are We Bodies or Souls?* Oxford: Oxford University Press.
- Thomson, Judith. 1997. People and Their Bodies. In *Reading Parfit*, edited by Jonathan Dancy, 202-229. Oxford: Blackwell.
- Torza, Alessandro. 2023. *Indeterminacy in the World*. Cambridge: Cambridge University Press.
- Tugby, Matthew. 2022. *Putting Properties First*. Oxford: Oxford University Press.
- Turner, Jason. Forthcoming. On Doing Without Ontology: Feature-Placing on a Global Scale. In

- The Question of Ontology*, edited by Javier Cumpa. Oxford: Oxford University Press.
- Unger, Peter. 1979. I do not exist. In *Perception and Identity: Essays Presented to A.J. Ayer, With His Replies*, 235-251. Ithaca, New York: Cornell University Press.
- Valberg, Jerome. 2007. *Dream, Death, and the Self*. Princeton: Princeton University Press.
- Weber, Clas. 2016. Being at the Centre: Self-location in Thought and Language. In *About Oneself: De Se Thought and Communication*, edited by Manuel García-Carpintero and Stephan Torre, 246-271. Oxford: Oxford University Press.
- Weber, Clas. 2025. *De Se* Modal Illusions. *Pacific Philosophical Quarterly* 106, no. 1: 33-66.
- Yetter-Chappell, Helen. 2022. Dualism all the way down: why there is no paradox of phenomenal judgement. *Synthese* 200, no. 2: 1-24.
- Zuboff, Arnold. 1990. One Self: The Logic of Experience. *Inquiry: An Interdisciplinary Journal of Philosophy* 33, no. 1: 39-68.
- Zuboff, Arnold. 2009. *Time, Self, and Sleeping Beauty*. Ph.D. dissertation, Princeton University.