Modal Inertness and the Zombie Argument

Tristan Grøtvedt Haze
University of Melbourne
tristan.grotvedt@unimelb.edu.au

This paper proposes a way of blocking the zombie argument against materialism. The central idea—which can be motivated in various ways, but which I will motivate by drawing on recent work by Wolfgang Schwarz—is that sentences reporting conscious experience are *modally inert*, roughly in the sense that adding them to a description of a metaphysically possible scenario always results in a description of a metaphysically possible scenario. This is notable in that it leads to a way of blocking the zombie argument which is perfectly compatible with modal rationalism and with the view that conceivability entails possibility.

Keywords: consciousness, zombie argument, materialism, conceivability, modal rationalism.

1. Introduction

The idea that we have certain knowledge about conscious experience exerts a powerful hold on the philosophical imagination. I am looking at a desk, or at least I think I am, but I might be wrong about that. It might be a hologram set up to trick me, or my brain may be being stimulated by someone to give me the experience of seeing a table, or an evil demon may be deceiving me. But I can't be wrong about the fact that I am seeing *this*—having *this* experience of seeing a table.

But what if this claim I think I know for sure to be true isn't really in the business of describing the world at all? What if my "beliefs" about what I am experiencing are actually playing a quite different role in my cognition from the beliefs I have which represent the world? An interesting view of this kind has been developed by Wolfgang Schwarz.[1] On this view, our feeling of certainty about our experiences is explained in terms of the special functional role that these "judgements" about experiences play in our cognitive architecture.

[1] See Schwarz (2018), (2019).

Furthermore, once we appreciate this functional role we seem no longer to have much reason to think that the "judgements" involved are judgements about how things are in reality. And so, this sort of explanation of our apparent certitude about conscious experience lends itself to a sophisticated sort of anti-realist view about sentences which appear to report such experience. On this view, it might be fine to assert and affirm and even call 'true' such sentences, as long as we keep in mind that they aren't in the business of describing reality.

Given such a view, it is plausible that these sentences are *modally inert* in a certain sense. This idea will be refined in what follows, but at a first approximation, a sentence is modally inert when adding it to a description of a possible scenario always yields another description of a possible scenario. The idea is that, since these sentences aren't in the business of describing reality, whenever you have a description of a possible reality, adding them won't change that. In a sense they add nothing to the scenario that is being described.

I want to show that the view that conscious experience sentences are modally inert is highly defensible (along the lines just indicated), and that it blocks the well known zombie argument against physicalism about the mind. Furthermore, it blocks the argument at a refreshing point. Much ink has been spilled about the process of *getting to* the sort of possibility claim that may then be thought to undermine physicalism almost automatically—especially, about whether conceivability entails possibility in the relevant sense. But the modal inertness response can grant the possibility claim. This may be of interest to physicalists, but also to philosophers with views about the epistemology of modality which they would like their physicalist colleagues to be at least *able* to agree with. If the modal inertness response to the zombie argument is the right response, then hostility between modal rationalism and physicalism about the mind can be dialled down.

To be clear, it is the view that conscious experience sentences are modally inert here that is doing the immediate work, rather than anti-realism about such sentences or any kind of explanation of our apparent certitude about experience. But the only compelling way of *motivating* the modal inertness thesis I am aware of is via anti-realism about such sentences. As for anti-realism's motivation in turn, I see much more room here for different approaches, but I take Schwarz's work to show that there is at least one credible way of motivating anti-realism.

In the next section I sketch Schwarz's explanation of our apparent certitude regarding conscious experience. In section 3 I discuss how this explanation motivates anti-realism about conscious experience sentences. In section 4 I explain how anti-realism leads to the thesis that conscious experience sentences are modally inert. In section 5 I explain how the thesis blocks the zombie argument, before concluding briefly in section 6.

2. Schwarz's Imaginary Foundations Model

In recent work, Wolfgang Schwarz has developed a model designed to explain, and in a sense vindicate, our feeling of certitude about experience. Notably, this model does not require that there be a realm of facts about the world which we are certain about in such cases—indeed, it strongly suggests that there is not. Here I will just sketch Schwarz's idea. The role it plays in the present paper is to show by example that there are ways of explaining our feeling of certitude about experience which do not appeal to a realm of facts about the world that we are certain about.

Schwarz's key idea is that a cognitive agent receiving sensory input can benefit from having, in addition to credences in various propositions about their environment, full credence in "imaginary propositions" which play the quite different role of recording incoming sensory experience. These then facilitate cognitive updating on the part of the agent. Imagining designing a robot, Schwarz writes:

> To optimally deal with sensory input, I suggest, we need to extend the robot's probability space by new sentences such that whenever a sensory signal arrives, the robot becomes certain of one of these sentences. But there is no good reason why these sentences must be correct and detailed descriptions of the relevant electrochemical signal. In principle, the update works just as well if the new sentences are bare tags, 'A', 'B', 'C', etc.[2]

And later:

[2] Schwarz (2018), p. 771.

> I have described a model of how subjective probabilities change under the impact of sensory stimulation. The model requires an agent's doxastic space to be extended by an "imaginary" dimension whose points are associated with sensory signals in such a way that when a given signal arrives, the agent assigns probability 1 to the corresponding imaginary proposition; the probability of any real proposition is then set to its prior probability conditional on that imaginary proposition. As I mentioned in passing, this general approach is hardly new: it closely resembles standard treatments in artificial intelligence. It is also well-known in the neuroscience of perception, where similar models have proved a useful paradigm (see Yuille & Kersten 2006). In these areas, the propositions on which an agent or her perceptual system is assumed to conditionalize are called 'percepts', 'sense data', or 'input strings', and people rarely pause to reflect on their representational features or on what the postulated models imply for the epistemology of perception.[3]

The hypothesis that *we* are agents whose cognitive architectures make use of this method, which Schwarz calls 'the method of sensor variables'[4], explains our feeling of certitude about experience. On this hypothesis, we *do* have full credence in "imaginary propositions" which record our incoming sensory experience. And further, given that our credences are more typically credences in propositions which represent the world as being one way rather than another, and given that, historically at least, we have not usually had in mind any hypothesis about our cognitive architectures such as this one about "imaginary propositions", it is not surprising that we as philosophers are tempted to take this felt certainty to be certainty about things happening in the world, i.e. certainty about propositions which do represent the world as being one way rather than another.

3. From Schwarz's Model to Anti-Realism

A *conscious experience sentence* is a sentence used to report, perhaps negatively, on phenomenal conscious experience. 'I am feeling pain' and 'I am not feeling pain' are arguably such sentences. 'I am seeing *this*', said while "pointing inwardly" at one's visual

[3] Schwarz (2018), p. 773.
[4] In Schwarz (2019).

experience, is also arguably such a sentence.[5] If Schwarz's Imaginary Foundations model is correct, conscious experience sentences express "imaginary propositions".

Let *anti-realism* about a kind of sentence K be the view that sentences of kind K do not express propositions which represent the world as being one way rather than another. So for instance, anti-realism about command-sentences like 'Shut the door!' is *prima facie* highly plausible; it seems like such sentences are playing a role quite different from that of representing the world.

Now, the Schwarzian view that conscious experience sentences express "imaginary propositions", while it may not strictly *entail* anti-realism about these sentences, nonetheless strongly suggests it. Thus we have a case for anti-realism.

4. From Anti-Realism to Modal Inertness

If conscious experience sentences are not in the business of describing the world, what is the effect of conjoining them with sentences that describe metaphysically possible scenarios, i.e. ways the world could have been?

A natural answer, I suggest, is that the effect is nil. Call a sentence metaphysically possible just in case it describes a metaphysically possible scenario. Take a sentence S that is metaphysically possible in this sense. And now conjoin with it a conscious experience sentence C. Granting, for the sake of argument, anti-realism about conscious experience sentences, what scenario does S & C describe? It must describe some scenario—as well, perhaps, as doing other things—since it contains a conjunct which describes some scenario, i.e. some way for the world to be. And since the other conjunct is not in the business of describing the world, it would seem that the scenario it describes is the very one that S describes, which we have supposed to be a metaphysically possible one. Hence S & C is metaphysically possible.

Now, you might think that, in order for a sentence to be metaphysically possible, it must also satisfy a logical requirement along the lines of *not harbouring a contradiction*. If all

[5] Such "inward pointing" is discussed critically in Wittgenstein (1953).

meaningful declarative sentences are in the business of describing the world, then this is arguably not a further requirement; a sentence which harbours a contradiction and is in the business of describing the world will not correctly describe any possible scenario. But once we allow anti-realism about some meaningful declarative sentences, you might think that it is a further requirement. You might think for instance that a sentence like S & C & ~C ought to count as metaphysically impossible since it is logically contradictory, even if the part of it which is in the scenario-describing business does describe a metaphysically possible scenario.

There are ways to resist this line of thought, but that doesn't matter for our purposes—we can simply grant it, and weaken our modal inertness thesis accordingly. It is enough to hold that, whenever you conjoin to a metaphysically possible sentence S a sentence that is not in the business of describing the world, *and* the resulting sentence is not logically contradictory, then the resulting sentence is metaphysically possible as well. This is the principle I will use below to block the zombie argument.[6]

5. From Modal Inertness to Blocking the Zombie Argument

Here is Chalmers on the zombie argument:

---

[6] I have tried to bring out the intuitive plausibility of this principle and to keep it as minimal as possible for present purposes. However, there are interesting questions here, reminiscent of the Frege-Geach problem for non-cognitivism in metaethics, about how it might be generalized. How should we understand the compositional semantics of compounds involving sentences not in the business of describing the world? For the case of sentences regarded as expressing Schwarzian imaginary propositions at least, we could proceed by augmenting standard possible worlds semantics so that, instead of truth-at-a-world, we use a notion of truth-at-a-*point*, where a point is regarded as answering all real and imaginary questions. We can then say that a sentence is true at a metaphysically possible world *w* iff it is true at some point *p* whose real aspect is *w*. If, however, we want to maintain the stronger principle that you *always* get a sentence describing a metaphysically possible scenario when you conjoin a sentence describing a metaphysically possible scenario with a sentence not in the business of describing the world, the truth-at-a-point proposal does not give us what we want, since S & C & ~C will come out false at all worlds on that proposal. This stronger version of the modal inertness idea seems to resist such generalization. To get a feel for the difficulty, suppose you are a logician wanting to regiment some sentences using a logical language whose sentences are *all* in the business of describing the world. Whether you are given S & C to regiment, or given S & C & ~C, you might find it natural simply to ignore everything except the S, which factor *can* perhaps be regimented in your logical language, and just try to regiment that. But if you were given, say, the *disjunction* of S with C, you might feel justified in throwing up your hands and saying 'I can't work with this!'. (Many thanks to Wolfgang Schwarz for raising the generalization issue, supplying the truth-at-a-point proposal, and pointing out its clash with the stronger version of the modal inertness idea.)

> The most straightforward form of the conceivability argument against materialism runs as follows.
>
> (1) P&~Q is conceivable
> (2) If P&~Q is conceivable, P&~Q is metaphysically possible
> (3) If P&~Q is metaphysically possible, materialism is false.
> (4) Materialism is false.
>
> Here P is the conjunction of all microphysical truths about the universe, specifying the fundamental features of every fundamental microphysical entity in the language of microphysics. Q is an arbitrary phenomenal truth: perhaps the truth that someone is phenomenally conscious, or perhaps the truth that a certain individual (that is, an individual satisfying a certain description) instantiates a certain phenomenal property. P&~Q conjoins the former with the denial of the latter.
>
> If Q is the truth that someone is phenomenally conscious, then P&~Q is the statement that everything is microphysically as in our world, but no-one is phenomenally conscious.[7]

It is natural to regard Q as a conscious experience sentence (or as the proposition expressed by such a sentence). And so given Schwarz's Imaginary Foundations model, it is natural to regard Q as expressing (or being) an imaginary proposition. There is a hint already in Schwarz about how his model might lead to a way of blocking the zombie argument:

> Similarly, if I_R is an imaginary proposition associated with red experiences, and P is the totality of all physical truths, we can explain why both P & I_R and P & $\neg I_R$ are *a priori* conceivable (see Chalmers 2009), even if the world is completely physical.[8]

But with modal inertness on board, we can do more than that. We can even explain why—and allow that—both conjunctions are *metaphysically possible*. That is, we do not need to stop at conceivability and then invoke worries about the move from *a priori* conceivability to

[7] Chalmers (2009), Section 1.
[8] Schwarz (2018), p. 784.

metaphysical possibility to block the zombie argument. We can block the argument by denying premise (3).

Before I say more in defense of blocking the argument in this way, I want to respond to the following objection regarding the dialectical situation: It is a *presupposition* of the zombie argument that Q is a genuine statement in the business of describing the world, and hence from the point of view of the kind of anti-realism under discussion, the argument doesn't even get started. Hence, we should not be discussing where and how to block it.[9]

In reply to this objection, I am happy to grant that the zombie argument in some relevant sense has a false presupposition (given anti-realism), and in that sense doesn't even get started. But I do not agree that that means we shouldn't run the argument and consider where best to block it by our lights. The zombie argument, in various forms, is well known and often repeated: in that sense, it *has* gotten started, whether the anti-realist likes it or not. And the presupposition is just that: a presupposition, not an explicit premise (or inferential step) in the argument as usually formulated. As I see it, one can object to an argument by identifying and attacking presuppositions, but one can also try letting the argument run, and then blocking it at some point. This latter strategy is what Schwarz is tentatively pursuing in the above quotation, and I am trying to improve on this. This type of strategy might be dialectically effective, and theoretically revealing, in a way that simply protesting that the argument has a false presupposition does not deliver (correct as such a response may be). Hence I should be allowed to try.

Now, blocking the argument by denying premise (3) is not merely a dialectical move which "follows the logic where it leads" but which lacks independent support. Anti-realism about conscious experience sentences, which we used to motivate the modal inertness thesis, undermines the reasons for finding (3) plausible in the first place. Why have people found (3) plausible? It is because physicalism is regarded as entailing the supervenience of facts about the world upon physical facts—or in more recently in-favour ideology, entailing that facts about the world are *grounded in* physical facts. As Chalmers says:

[9] Thanks to an anonymous referee for raising this objection.

> The third premise is relatively uncontroversial. It is widely accepted that materialism has modal commitments. Some philosophers question whether materialism is equivalent to a modal thesis, but almost all accept that materialism at least *entails* a modal thesis. Here one can invoke Kripke's metaphor: if it is possible that there is a world physically identical to our world but phenomenally different, then after God fixed the physical facts about our world, he had to do more work to fix the phenomenal facts.[10]

But given anti-realism, conscious experience sentences are not in the business of reporting facts about the world. So there is no failure of supervenience or grounding. (3) turns out to be false in light of modal inertness, a natural consequence of anti-realism. The zombie argument leaves physicalism (a.k.a. materialism) unscathed.

At this point a defender of the zombie argument might want to object that, in their argument, Q is *not* supposed to be anything like a Schwarzian "imaginary proposition". Rather, it is supposed to be a proposition about the world, e.g. that someone has a particular property, namely the property of being phenomenally conscious. This is fair enough as far as it goes, but the Schwarzian anti-realist has a ready reply: insofar as Q is *not* merely the kind of proposition which we intuitively feel certain about, and which the Imaginary Foundations model explains, then—granting that explanation—we need not believe it to be true. Some philosophers may believe such a thing, and may take it to be almost undeniable, but to take it to be almost undeniable stems from misidentifying it with the "imaginary propositions" we have full credence in. Put provocatively: given a realist understanding of Q we are *all* zombies.

So the defender of the zombie argument faces a dilemma. Either the Q in their argument can be explained in anti-realist terms, or it is a robustly representational claim. (Which it is in a given instance will depend on the Q, and there may be subtle interpretative concerns about it. Admittedly, Chalmers' initial suggestion above, namely 'the truth that someone is phenomenally conscious', being quite theoretical-sounding, may push one towards a robustly representational interpretation. However, if you take your Q to be something like 'I am seeing *this*' while "pointing inwardly", then the anti-realist explanation becomes more clearly

[10] Chalmers (2009), Section 1.

fitting. As I see it, Q's of the latter category are more basic, giving rise to the philosophical problem and to the more theoretical-sounding Q's.) If Q can be explained in anti-realist terms, then it is plausibly modally inert, rendering premise (3) of the zombie argument false. If Q is a robustly representational claim, then—having explained that the feeling of certitude that might have been thought to lend support to Q actually attaches to an "imaginary proposition"—we are free to regard P & ~Q as, not just metaphysically possible, but actually true. And so again premise (3) is false; on *this* understanding, P & ~Q is entirely consonant with physicalism.

6. Conclusion

I have proposed a way of blocking the zombie argument which doesn't require entering the debate about the relationship between *a priori* conceivability and metaphysical possibility, and therefore does not push one toward modal empiricism. That such an option exists should be good news for physicalists, but also for those who think there is something right about modal rationalism (which, they might think, is crucial to our whole conception of philosophy); favouring a materialist metaphysic does not force you out of the philosophical paradise of having *a priori* insight into what's metaphysically possible. Physicalists and non-physicalists alike may happily cohabit there.

References

Chalmers, David (2009). The Two-Dimensional Argument Against Materialism. In Brian P. McLaughlin & Sven Walter (eds.), *Oxford Handbook to the Philosophy of Mind*. Oxford University Press.

Schwarz, Wolfgang (2018). Imaginary Foundations. *Ergo: An Open Access Journal of Philosophy* 5.

Schwarz, Wolfgang (2019). From Sensor Variables to Phenomenal Facts. *Journal of Consciousness Studies* 26 (9-10):217-227.

Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Wiley-Blackwell.

Yuille, Alan and Daniel Kersten (2006). Vision as Bayesian Inference: Analysis by Synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308.