

Richard Brown

## *Deprioritizing the A Priori Arguments Against Physicalism*

**Abstract:** *In this paper I argue that a priori arguments fail to present any real problem for physicalism. They beg the question against physicalism in the sense that the argument will only seem compelling if one is already assuming that qualitative properties are nonphysical. To show this I will present the reverse-zombie and reverse-knowledge arguments. The only evidence against physicalism is a priori arguments, but there are also a priori arguments against dualism of exactly the same variety. Each of these parity arguments has premises that are just as intuitively plausible, and it cannot be the case that both the traditional scenarios and the reverse-scenarios are all ideally conceivable. Given this one set must be merely prima facie conceivable and only empirical methods will tell us which is which. So, by the time a priori methodology will be of any use it will be too late.*

Roughly speaking, physicalism is the view that only physical things exist. Physical things are those things that are postulated by a completed physics. Currently physical things include quarks, electrons and various forces but these may or may not be part of the completed theory. To say that only physical things exist is not to deny that tables, chairs, cars, and brains exist. They do, but only because they are composed of physical things. Physicalism can then be more precisely formulated as the thesis that only physical things, or things which are made from physical things, exist. To put this another way the

Correspondence:  
Richard Brown,

physicalist holds that a complete microphysical duplicate of the actual world is a complete duplicate *simpliciter*. So, the physicalist in the philosophy of mind holds that a complete microphysical duplicate of me would be a complete qualitative duplicate of me as well. For instance, any complete micro-physical duplicate of me in a world physically just like ours would have the very same kind of pain experiences that I have. He would see the very same red when looking at a ripe tomato etc. Given this, and some other assumptions, the physicalist is committed to being able to deduce that these qualitative properties exist from a completed micro-physics (Chalmers & Jackson, 2001).<sup>1</sup> I am a physicalist but I do not intend, in this paper, to argue that physicalism is true. Rather I intend to argue for the more modest claim that no current a priori argument shows that it is false. In particular I will focus on the knowledge argument and the zombie argument as canonically formulated by David Chalmers. This does not mean that physicalism is in fact true. There may be empirical arguments or perhaps other a priori arguments against it. Currently the empirical case for physicalism is very strong, but that may change. Also, there are other a priori arguments against physicalism, like Kripke's well known modal argument, but all of these rely on the same intuitions as found in the knowledge and zombie arguments. What we say here can be easily extended to them. All of this taken together shows that the case for physicalism is quite strong but here I want to focus more on the a priori case against physicalism.

I am willing to grant that physicalism could be shown to be false a priori (if in fact it is false); of course I also hold that it can be shown to be true a priori (if in fact it is true).<sup>2</sup> To see how, it is important to recognize that something is a priori just in case it *can* be known independently of experience and this introduces a new modality (Kripke, 1980, pp. 34–38). So, while it is true that if we idealize the rationality and knowledge of an agent all things would be a priori it does not follow that we can know these things a priori now or that the way we will first come to know them is via a priori means. For instance, before the rise of modern chemistry it was true that the totality of physical facts about our world entail that water is a molecule composed of hydrogen and oxygen and is not a simple substance but people in 1605 CE were

- 
- [1] Not every physicalist believes this is true (see Block & Stalnaker 1999) but I grant it for the sake of the arguments herein.
- [2] An empiricist at heart I am in general a skeptic about the existence of a priori knowledge. I tend to agree with the arguments given by Michael Devitt in his (Devitt forthcoming). But I here set that aside in order to engage the dualist on the battlefield of their choice. I will therefore use 'a priori' in the way that Chalmers does.

not in a position to know that this was the case. They could have known that a certain conditional statement would have been true—if the world turns out a certain way then water is a molecule, if it turns out another way then not—but they could not have known which of these it was. We came to know which was actually the case empirically (see Read, 1995, pp. 133–35) and so historically this is an a posteriori discovery. Now, some 235 years later, we can see that water as molecule facts are entailed by the physical facts and we can arguably see this a priori. In other words an ideal agent who knew all of the actual physical facts would be able to determine that water was a molecule, which makes it knowable a priori, even though historically it was an empirical discovery.

Given this I can restate the aims of my argument more precisely. First I will argue that we are not yet in a position to see a priori whether physicalism is false (or whether it is true). Second I will argue that the way we will come to know which will be via empirical means. An idealized agent would know a priori that physicalism was true or that it was false, but we are not idealized agents. I am willing to grant that we can make some kinds of claims about what ideal reasoning would look like but this is not the main problem in my view.<sup>3</sup> The problem comes from our lack of empirical knowledge. As the water/H<sub>2</sub>O case shows, to be in a position to see that ‘water is a compound’ is entailed by the physical facts required that we know that water is actually H<sub>2</sub>O. Imagine someone in 1768 defending the Aristotelian view of water as a simple element against someone who suggests that it might be a compound of hydrogen and oxygen by asserting that she could imagine a world micro-physically identical to ours with H<sub>2</sub>O but no water. As it turns out this cannot really be conceived, since it is to conceive of something contradictory. It must have only been *prima facie* conceivable to the 18th century objector. So too, in order to know whether physicalism is true or false a priori requires knowing whether it or dualism is *actually* false. Once we know that we will be in a position to a priori deduce, or not, the qualitative facts from the physical facts. But, of course, once we know whether physicalism is actually false we won’t need a priori methods to see that. It is therefore the case that from where we are now, only empirical work will advance the debate between the physicalist and the dualist. From where we are now offering a priori arguments are of little to no use for showing that physicalism is false. All it can do, I

---

[3] Some might find this problematic. For instance, does ideal reasoning include the law of non-contradiction?

argue, is to let one know where one stands with respect to the question but cannot answer it. Since we do not yet know whether any physicalist theory of consciousness will ultimately be vindicated we cannot yet say whether zombies are really conceivable or whether they are only *prima facie* conceivable.

What results is a position that falls under what is known as type-C physicalism. To make this argument I will present 4 parity arguments, two of them familiar from the literature two of them not. Afterwards I will turn to examining a number of objections.

## 1. The Reverse-Thought Experiments

### 1.1. *Zoombies*

Zoombies are creatures that are nonphysically identical to me in every respect and which lack any nonphysical phenomenal consciousness. Put a bit more formally we can say that where NP is the totality of the nonphysical facts about me now and Q is some qualitative fact about me, say that I am now seeing green, it is conceivable that NP &  $\sim$ Q obtain.<sup>4</sup> That is to say that I can conceive of all the actual nonphysical properties being instantiated, in just the way they are now, and yet not including qualitative properties. There is no obvious contradiction that emerges from conceiving of all actual nonphysical properties being instantiated and yet not including phenomenal properties. From this it follows that dualism is false. The reasoning is exactly the same as in the traditional zombie argument. Zoombies closely resemble zombies. The traditional zombie is a creature that has everything I do physically but lacks phenomenal properties, a zoombie is likewise a creature that has everything I do nonphysically but lacks phenomenal properties. We can formalize the zoombie argument as follows,

1. NP and  $\sim$ Q is conceivable
2. If (NP &  $\sim$ Q) is conceivable, then (NP &  $\sim$ Q) is possible
3. If (NP &  $\sim$ Q) is possible then dualism is false
4. Therefore dualism is false

This argument is in every way parallel to the original. So it either shows that dualism is false or it shows that there is the very same problem with the original zombie argument. The problem here is that we

---

[4] I sometimes get asked what kinds of nonphysical properties zombies have. One way this might go is to say that NP is the null set and so what we are really conceiving is that NP is empty. But I think we can make some sense of the idea that zoombies may have other non-physical properties. I can conceive of zoombies as having only non-qualitative cognitive states like thoughts and that this exhausts the nonphysical properties of the actual world. Either way NP &  $\sim$ Q seems just as conceivable.

have just assumed that qualitative properties are not nonphysical properties, which seems unfair to the dualist. This is, of course, the precise complaint that many physicalists make about the traditional zombie argument.

### 1.2. *Shombies*

A shombie is a creature that is micro-physically identical to me, has conscious experience, and is completely physical.<sup>5</sup> Shombie pain is just as painful as my pain is and shombie orgasms are every bit as pleasurable as mine are. My shombie twin and I all have the same experiences. The only difference, if it is a difference, is that shombie pain is completely physical. That doesn't make it any different from the inside. What it is like for me to have a pain and what it is like for my shombie twin to have a pain are identical in all respects. We have stipulated that shombie pain is just like my pain in every respect (qualitatively) and that my shombie twin is a complete micro-physical duplicate of me in a world that is stipulated to lack nonphysical properties of any kind and that this is all there is to a shombie. The shombie is NOT a zombie. A zombie lacks phenomenal consciousness; a shombie doesn't. The qualitative does, therefore, logically supervene on the physical and dualism is false. Zombies are metaphysically impossible.

We can formalize the shombie argument as below; where P is a complete physical description which includes a 'that's all' phrase and Q is the same as before.

1. P and Q is conceivable
2. If (P & Q) is conceivable, then (P & Q) is possible
3. If (P & Q) is possible then dualism is false
4. Therefore dualism is false

(4) follows because in order for dualism to be true there has to be a zombie world but shombies show that there is no zombie world since the world physically identical to ours also has consciousness just like ours.

Again we see that this argument is exactly parallel to the original zombie argument. And again what this demonstrates is that either

---

[5] Shombies have received a lot of attention in the literature (Martin 1998; Sturgeon 2000; Piccinini 2006; Frankish 2007; Balog forthcoming) though none of these authors put them to the use that I do. Usually shombies are used to argue that both zombies and shombies are conceivable and so conceivability doesn't entail possibility whereas I am arguing that both seem *prima facie* conceivable but only one of them is ideally conceivable so as to engage the dualist like Chalmers on their own terms.

dualism is false or that the original zombie argument makes the same mistake that this one does. The mistake here is the same as before. We have simply assumed that qualitative properties are physical properties, which begs the question against the dualist.

### *1.3. Maria*

Maria is Mary's twin sister who was separated from her at birth and so is the intellectual equivalent of her super-scientist sister. However the evil scientists who raised Maria raised her as a super phenomenologist. She was raised in a special room where she was taught from a very early age to focus on her own experience. She learns to master all of the platitudes of folk psychology and so is a master of such things as that red is more like pink than it is like blue and that turquoise is more like blue than it is like red and on and on to a degree that we can only dream of. Maria is able to discriminate between shades of color that we cannot (though perhaps we could with the proper training) also she is able to describe her experience as accurately as humanly possible. She, in short, knows everything there is to know about her own experience. She is, however, kept completely ignorant about all physical theories of our time or anyone else's. She knows that she has a body but does not know anything about the way it works.

Then one day Maria is taught the completed science of her day. This includes everything from Aristotle's theories to the completed physics of her time. She comes to know everything there is to know about the brain and color processing in the brain as well as the physical theory about light and the way it is reflected all the way down to the completed microphysics.<sup>6</sup> I have the intuition that Maria will then learn that her visual experience of red is just a brain state, just as she learns that water is H<sub>2</sub>O. She will learn that her color experience is a physical event in her brain. Maria will learn something that she would express by saying 'oh, so that's what my color experience is!' Once she sees the identities she will be in a position to deduce the qualitative facts from the physical facts a priori.

We can reformulate the Reverse-Knowledge argument as follows,

1. 'P  $\rightarrow$  Q' is a priori (the reverse-Knowledge intuition)
2. If 'P  $\rightarrow$  Q' is a priori, it is necessary
3. If it is necessary, dualism is false
4. Dualism is false

---

[6] To make all this go faster we can imagine that the completed science of Maria's day give them the ability to upload all this information into Maria's brain in the amount of time it takes Mary to see her first red ripe tomato.

This way of formulating it puts it in reverse-form to the way that Chalmers formulates the original knowledge argument (Chalmers, 2009). This argument is to me very compelling. In fact I suspect that something like this describes the way that most physicalists come to be convinced of physicalism. And as with the previous two arguments what it shows is that either dualism is false or that it makes the very same mistake that this one does. The mistake here is the assumption that the qualitative facts are a special subset of the physical facts just as in the original knowledge argument the assumption is that they aren't.

#### *1.4. Mark*

Mark is the name given by Yujin Nagasawa (2002) to Churchland's (1989) Mary-like super-scientist who learns the completed nonphysical science without seeing red. When Mark is let out of his special black and white room and sees his first red ripe tomato there is no reason to think that he won't learn what it is like to see red in exactly the same way that Mary did. This argument exactly parallels Jackson's (1986) original formulation of the knowledge argument. Mark knows all of the nonphysical facts but yet learns something new when he sees red for the first time, therefore phenomenal facts cannot be deduced from the nonphysical facts. Whatever response the dualist gives to Mark can be given to Mary.

## **2. Type-C Physicalism**

It will be useful for us to remind ourselves of David Chalmers' (2003) well-known classification of physicalist theories in terms of how they respond to the a priori arguments against it.<sup>7</sup> As we have seen the a priori arguments for and against physicalism have a common structure. They start with a conceivability claim move to a possibility claim and then make a metaphysical conclusion. Type-A physicalists deny the conceivability claim. They deny that zombies are conceivable and insist that the zombie world is a world where there is consciousness. Typical type-A physicalists include Dan Dennett and Paul Churchland and often claim that it is analytically true that pain is a functional state. Chalmers, in fact, seems to equate type-A physicalism with being an eliminative materialist.

Type-B physicalists deny the possibility claim. They take the intuitive conceivability of zombies and Mary to be compelling but deny

---

[7] See Dave Beisecker's paper in this issue for a new type of physicalism.

that zombies are therefore possible. They typically invoke Kripke's notion of a posteriori identities. So take our empirical discovery that water is in fact  $H_2O$ . We know that water equals  $H_2O$  and that this identity is necessary (since all of them are). But still it is conceivable that water is not  $H_2O$  in the sense that we can imagine a world where the watery stuff is not  $H_2O$ ; this is our old friend Twin Earth. On Twin Earth the stuff that falls from the skies, and which life depends on, and which people bathe in, etc is not  $H_2O$ . It is some other stuff with a very complicated chemical formula we can abbreviate by XYZ. So the Twin Earth world is conceivable but not possible.

The type-C physicalist admits that zombies seem conceivable but then denies that they are ideally conceivable. At the ideal limit we will be able to make the required deductions and we will see that zombies are not ideally conceivable. Type-C physicalism is clearly the most plausible kind of physicalism. It allows us to agree with the dualist that the anti-physicalist thought experiments are intuitively compelling, given what we know now, and also admit the intuitive principle that an ideally rational agent who knew all of the physical facts would be in a position to determine which things were contradictory and which things weren't. Given the traditional rationalist assumption that contradiction is a guide to what is knowable a priori this agent would be in a position to know, a priori, all of the truths.

We can see that there are corresponding types of dualism. The type-A dualist will deny the conceivability claim in the reverse-thought experiments. The type-B dualists will admit that both are conceivable and then deny that they are both possible. The type-C dualist will hold that they both are *prima facie* conceivable but that only one of them is ideally conceivable. Chalmers at times has talked like a type-A dualist (Chalmers, 2009; see especially section 8 objection 15), which is what partially inspires the physicalist's reaction that he is simply refusing to take physicalism seriously. Type-A dualism is plausible only if one assumes a kind of analytic dualism on which our concepts of qualitative properties explicitly note that they are non-physical. But, stipulating that qualitative properties are nonphysical is question begging. This is the heart of the matter. What non question-begging reason can one give to show that zombies are really conceivable and zoombies and shombies aren't? Once we have set aside stipulative answers to the problem of consciousness and as long as we refuse to give up the thesis that conceivability in some sense entails possibility then type-C dualism is the only viable alternative. Since there are no a priori reasons to choose between type-C physicalism and type-C dualism only empirical discoveries will decide the issue.

Thus the way that we will come to know whether physicalism is true or false will be a posteriori even though it can ultimately be known a priori.

### 3. Objections and Replies

#### 3.1. *To Type-C Physicalism*

The main argument against type-C physicalism is that it threatens to collapse into either type-A or type-B physicalism and then will have all of the problems associated with those positions. However, the kind of type-C physicalism advanced here does not fall to any of these objections. Let us take them in turn.

The argument against type-A physicalism is just that they have failed to take consciousness seriously. Consciousness cannot be eliminated or explained away. It is a real phenomena that is perhaps the one thing we know best about the world and there are no conceptual connections between physical and phenomenal concepts. So I am with Chalmers in setting aside analytic functionalism and all other eliminative views about conscious mental states. According to the kind of physicalism I am arguing for what allows us to complete the deduction of phenomenal facts from the physical facts is the (for us) a posteriori discovery of identities between phenomenal and physical properties. It is because of these identities that we are able to deduce the mental facts from the physical facts. How might this deduction go? One possibility might be as follows. Suppose that some kind of higher-order theory of consciousness is right. It will then be the case that a conscious pain is a pain that I am aware of in the right way. We might then find out how it is that we are aware of our own mental states and come to see that this is just a certain brain state. We would have then found out what a conscious pain is. This will be done only after our brain science gets to the point where it can verify or falsify theories of consciousness.

Chalmers thinks that type-C physicalism collapses into type-A physicalism because he can see no way for us to get any conceptual hooks to link the notion of structure and function with the phenomenal. On the present account the hooks come from the discovered identities. This does mean that we will have discovered that phenomenal properties can be explained in broadly functional terms but this does not thereby endorse type-A physicalism. It is just to point out that I cannot really conceive of anything else doing any explanatory work. What else besides structure and function is there? No one has ever given anything like a proper account of what nonphysical properties

are or how they explain phenomenal consciousness. In fact every appeal to nonphysical properties has eventually succumbed to elimination via broadly functional conceptual hooks. Given this and the deeply mysterious nature of nonphysical properties we should err on the side of physicalism. Of course this anti-structure and function argument is exactly the reverse of Chalmers' arguments against structure and function. Which just again illustrates the parity between the two positions.

Since I invoked a posteriori identities one might worry that the view defended here falls prey to Chalmers' arguments against type-B physicalism. Chalmers argues that there is clearly a sense in which 'water is not H<sub>2</sub>O' is primarily ideally conceivable and so metaphysically possible. It is conceivable in the sense that if Twin Earth had turned out to be actual it would have been the case that water was not H<sub>2</sub>O. If Twin Earth were actual, as opposed to counter-factual, then water would have been XYZ. Whether we call the watery-stuff on Twin Earth 'water' or not is irrelevant. When we are imagining Twin Earth we really do have access to some possible situation and if that possible situation had been actual then it would have been true that water was not H<sub>2</sub>O. This invokes the distinction between primary and secondary conceivability and the corresponding kind of intensions. The primary intension of a sentence gives us the truth-value for that sentence in possible worlds that we consider as actual. Thus, if twin Earth were actual water would be XYZ, and so 'water is not H<sub>2</sub>O' is primarily conceivable. The simple way to think about primary intensions is that they are descriptions that pick out different referents in different possible worlds ('water' has the primary intension 'the watery stuff' for instance). The secondary intension of a sentence gives us of the truth-value of that sentence at possible worlds considered as counterfactual. In other words we hold the actual world fixed and then ask what could be true of it. So, 'water is not H<sub>2</sub>O' is not secondarily conceivable since its secondary intension is false. There are no worlds where H<sub>2</sub>O is not H<sub>2</sub>O and given that water is actually H<sub>2</sub>O it is impossible for water not to be H<sub>2</sub>O.

Chalmers then argues that if we are careful to start with ideal primary conceivability then the zombie argument goes through. If the zombie world had been actual it would be the case that physicalism is false. This gets us to the claim that the zombie world is primarily possible, and Chalmers argues that this is enough to falsify physicalism. Therefore physicalists who endorse a posteriori identities between qualitative states and brain states cannot avoid the zombie argument by invoking Kripke. The dualist is in some sense conceiving a real

primary possibility when they imagine a zombie world (just like the person conceiving Twin Earth is conceiving something which is metaphysically possible). Whether we apply our word ‘consciousness’ to it is irrelevant. The realm of possibilities has not shrunk and ideal primary conceivability is still a good guide to what is metaphysically possible.

But what the reverse-thought experiments show is that it is not clear that zombies really are primarily ideally conceivable. Zoombies and shombies seem to me to be just as primarily ideally conceivable as zombies; which is to say that they only seem contradictory when one has tacitly accepted a theory about what qualitative properties are, but doing so begs the question against the other side from the beginning. Chalmers starts with the assumption that zombies or the Mary intuition is in fact ideally and primarily conceivable. But it isn’t obvious that it is, at least it isn’t until one has shown what is wrong with the reverse-thought experiments and why they aren’t primarily ideally conceivable. I can even grant that the zombie world might be negatively conceivable in the sense that there is no obvious contradiction in imagining the zombie scenario to be true — just as Chalmers grants this for physicalism — but it is not positively conceivable. To be positively conceivable is to do more than merely be unable to detect a contradiction. We may simply not know enough to see that there is a contradiction. It is to actively envision the scenario holding in detail. So to positively conceive of the zombie world is to succeed in imagining a world that is physically identical to ours and which lacks qualitative consciousness. But this cannot work since we do not yet know if physicalism is in fact true. If the identity theory is true then there is a contradiction in the specification of the zombie world for it asks us to imagine a world where a certain physical property is present and also not present at the same time; after all if the two properties are really the same property then wherever there is the one there must be the other. So if the identity theory is true zombies are not really ideally conceivable and so cannot yet be evidence against physicalism or used as an argument that tries to show that physicalism is false.

Just as in the shombie case. If dualism is true then there is a contradiction in the shombie world since it would have to both have phenomenal conscious (since it is described that way) and yet lack it (since there are no nonphysical properties). Someone must be begging the question here, but both sets of arguments are exactly parallel so there is no good a priori reason to say who is doing the begging. So Chalmers’ standard objection to type-B physicalism doesn’t apply to me. This is because unlike the type-B physicalist I do not think that

zombies are ideally conceivable. We cannot yet rule out that zombies (or shombies for that matter) merely seem to be conceivable, or that they are merely *prima facie* conceivable and not ideally conceivable. Until we are in a position to do so the zombie argument is question begging.

One other objection to type-B physicalism comes from strong necessities. A crucial premise of Chalmers' argument is that when it comes to pains and other phenomenal properties their primary and secondary intensions are identical. What that means is that the statements in question pick out the same property no matter whether we consider the world as actual or counterfactual. This is supposed to capture Kripke's claim that it is impossible for there to be someone in the same epistemic situation as someone who was in pain and yet for that person not to be in pain (and that it is impossible for there to be a person who was in the same epistemic situation as someone who wasn't in pain and yet be in pain). In short, the idea is that there is no appearance/reality distinction when it comes to pains. So then the upshot here is to try to show that there is a difference between the way 'water is not H<sub>2</sub>O' works and the way 'pain is not C-fiber firing' works that preserves Kripke's general idea but is made precise by the 2-D framework. Kripke's basic idea was that when we think that some identity is contingent what is really going on is that there is some identity statement involving a description that is contingent ('the watery-stuff=H<sub>2</sub>O' is contingent) but this can't be the way we explain away the seeming contingency of 'pain is C-fiber firing' since there is no alternate contingent identity involving a description in the case of pains. This translates into the 2-D framework as the claim that Kripkean a posteriori necessities have a contingent primary intension (i.e. 'water isn't H<sub>2</sub>O' comes out true at some possible world (e.g. Twin Earth) considered as actual) but 'pain isn't c-fiber firing', according to the physicalist, has a necessary primary intension (there are no worlds considered as actual where this comes out true). Chalmers takes this to show that the postulated mind-brain identities do not behave like Kripkean a posteriori necessities. They are stronger in that their primary intensions are necessary but we cannot know them a priori.

But what are we to make of this claim? Is it really the case that the primary and secondary intensions of 'the painful stuff is c-fiber firing' are identical? Or another way of asking the question; can the painful stuff fail to be c-fiber firing at some possible world considered as actual even though the painful stuff picks out c-fiber firing here in the actual world? Chalmers, and Kripke, seem to think that it is a priori

that the answer is no but there is empirical evidence that suggests that it is at least not contradictory to think that the answer is yes. Cases of what is commonly known as dental fear suggest that we can pick out mental states as painful which are not pains; while pain asymbolia also, arguably, shows that we can pick out pain states without picking them out as painful. These kinds of results show us that it is at least not obviously contradictory to think that the primary intension of pain diverges from its secondary intension and if so 'pain' would work just like 'water'. Let us look at these phenomena.

David Rosenthal discusses the phenomenon of dental fear in his 'Sensory quality and the relocation story' (Rosenthal, 2005). Here is what he says,

Dental patients occasionally report pain when physiological factors make it clear that no pain could occur. The usual explanation is that fear and the non-painful sensation of vibration cause the patient to confabulate pain. When the patient learns this explanation, what it's like for the patient no longer involves anything painful. But the patient's memory of what it was like before learning the explanation remains unchanged. Even when what it's like results from confabulation, it may be no less vivid and convincing than a non-confabulatory case. (p. 172)

These patients can't be having a pain sensation since the relevant nerves are anesthetized. Nonetheless they think that they are in pain (when they aren't).<sup>8</sup>

I have always felt that this dental fear stuff was a really convincing way of showing that there really is a reality/appearance distinction for pains, but when I have tried to research this I have not been able to find very much on it (and Rosenthal offers no citations), but it does seem to be a relatively common phenomenon. Here is an excerpt from a paper on dental fear in children that tells dentists how to deal with this;

Problems that a dentist is convinced are associated with misinterpretation of pain may be addressed by explaining the gate theory of pain. A very basic explanation which is suitable for children as young as five is as follows. 'You have lots of different types of telephone wires called nerves going from your mouth to your brain (touch appropriate body parts). Some of them carry 'ouch!' messages and the

[8] One might want to insist that these patients really are in pain. After all, they *seem* to feel pain and, one might argue, that is all there is to being in pain. But it is arguably the case that in the ordinary sense of the word 'pain' we mean to pick out the pain sensation, which is clearly absent in the dental fear cases. Besides which, these cases are here used to show that it is conceptually possible to conceive of the pain sensation and the painfulness coming apart without contradiction. This is all I need to make the argument. See Justin Sytsma's paper for some empirical data on whether or not ordinary people agree.

others carry messages about touch (demonstrate) and hot and cold. The sleeping potion stops the ouch messages being sent, but not the touch and the hot and cold messages. So you will still know that I am touching the tooth and you will still feel the cold of the water. Your brain looks out for messages all the time. If you are convinced that it will hurt, it will. This is because if I make the ouch nerves go off to sleep and I touch you, a touch message gets sent. But your brain is looking for ouch messages and it says to itself, 'There's a message coming. It must be an ouch message.' So you go 'ouch' and it hurts, but all I did was to touch you. It's just that your brain was confused.' (The language may, of course, be adjusted for older children.) If this fails to work, then active treatment should be stopped (Chapman & Kirby, 1999).

This is clearly fool's pain, as evidenced by the fact that the way they treat it is not with more medication, but with an explanation, pitched at the kids level, of why what they are feeling is not pain. So it is conceivable that someone be in the epistemic position of someone in pain and yet not be in pain. When we conceive of pain that isn't c-fiber firing we may be conceiving of someone in a dental fear-like position. This person picks out some state in the same way that we normally pick out c-fiber firing, that is, as painful, even though the state they pick out is not c-fiber firing. Just as in the water/H<sub>2</sub>O case. What this suggests is that 'pain isn't c-fiber firing' does have a contingent primary intension and if this is true then there is no problem with strong necessities.

Next let's look at pain asymbolia (Grahek, 2001). Pain asymbolics have a specific kind of brain damage that leaves them able to feel pain but not as something unpleasant. They are able to discriminate painful stimuli, saying for instance that something hurts and that it is a burning pain, or a pinching pain. They are also able to reliably judge how intense the stimulus is and so can tell the difference between a light pinprick and one that pierces the skin. Yet they fail to be motivated to withdraw and say that they do not find the sensation unpleasant at all. They often laugh at the pain. They seem to know that pain is supposed to be this horrible thing that we want to avoid at all costs, but when the pain actually comes it is a pathetic joke. While it is not exactly clear what we should make of this it is at least consistent with what these people say to hold that they have the pain sensation without painfulness. This gives us a response to the zombie intuition. If this is right then when the dualist thinks that he is conceiving of a zombie world he is actually conceiving of a world where c-fiber firing lacks the contingent property of being painful for the person that has it and if it lacks a

property then there must be some physical difference in that world. Hence the physicalist can acknowledge that the dualist is really conceiving of some possible world without having to worry. The world they are conceiving is one that very closely resembles the actual world but has some crucial element missing (i.e. the element that would entail the inclusion of the contingent property of being painful for the person who has it).<sup>9</sup> This is, of course, not the way that the dualist would interpret what is going on but what we need is some kind of argument that rules this out as a possibility.

If these are consistent interpretations of the experimental and common sense data then why is that so many philosophers resist this idea? Something that Kripke says in another context is useful here. He says,

The fact that we identify light in a certain way seems to be *crucial*, even though it is not necessary; the intimate connection may create the *illusion* of necessity (Kripke, 1980, p. 139).

He is here responding to the objection that a world where something besides photons, say heat, was the cause of visual experiences (rather than photons) would not be a world where light was heat. It is more naturally described as a world where people pick out heat in the same way that we pick out light. So also it seems that being unpleasant is crucial to the way we identify being a pain. But this crucial connection generates an illusion of necessity. The fact that pain is painful for us is a contingent feature of the sensation of pain, as evidenced by pain asymbolia and dental fear. At the very least what we would need is an argument showing that there is more than just the illusion of necessity generated by painfulness seemingly being crucial to the way that we identify pain sensations. Unless this can be done the physicalist has an empirically plausible response to both the objection from strong necessities and the reformulated argument against type-B physicalist in terms of primary conceivability.

I conclude that the present view does not collapse into type-A or Type-B physicalism and so is not threatened by any of the objections to those positions. Let us now turn to addressing objections specifically to the reverse-thought experiments.

### 3.2. *To the Zoombie Argument*

The zoombie intuition relies on there being nothing contradictory in the totality of the actual nonphysical properties not including phenomenal properties. Though as formulated it is valid one might argue

[9] For instance it might not include the higher-order states necessary for the pain sensations to be conscious.

that it is unsound.<sup>10</sup> One might, for instance, think that this is so because nonphysical properties, though necessary for qualitative consciousness (according to the dualist), need not be sufficient. That is, there might be a creature that was identical to me in all nonphysical respects (that is, had all of the nonphysical properties that I in fact do have) but because it lacked a certain *physical* element these nonphysical properties were ‘inert’ and so the creature does not have any conscious experience (a special kind of neuron or a certain kind of firing pattern, might be needed in order to ‘turn on’ the nonphysical properties in such a way as to get consciousness). If this is possible then the existence of zombies does not show that dualism is false since premise (3) of the zombie argument would be false.<sup>11</sup>

But there cannot be inert nonphysical properties in the zombie world as properly conceived. The zombie world is one nonphysically exactly like our world. So if the dualist thinks that they need certain physical properties, or certain nonphysical laws, in order for me to consciously experience, say, pain, then that will be present in the zombie world.<sup>12</sup> This is because the zombie world is stipulated to contain exactly the same nonphysical properties that I actually have and to contain them in exactly the same way that they are actually contained. This is not to deny that worlds like the one in the objection are possible; they may be, but these worlds are not zombie worlds. Compare: the physicalist like myself admits that there are physical worlds where there is no consciousness but these worlds are not physical duplicates of our world and so are not truly zombie worlds. In fact, as just discussed, the physicalist like me thinks that it is one of these worlds that the dualist actually succeeds in conceiving.

Some philosophers argue that the zombie argument is not really a parody of the zombie argument. If this is so then one can argue that the zombie argument is indeed a bad argument but the flaws it has do not infect the traditional zombie argument.<sup>13</sup> The basic reason for this is that there is an alleged important dis-analogy between the original

---

[10] Thanks to Robert Howell for pressing this objection.

[11] What is nice about this is that if this is exactly the same kind of move that a physicalist like me makes about the zombie world. What you are actually conceiving, what is actually possible is the world that looks like ours but is not microphysically identical to it. The exact parity between these two argument is again striking.

[12] In fact zombies may have to be physically as well as nonphysically identical to me on some versions of dualism. If they are physically identical to me but lack qualitative consciousness then one may think that I have accidentally shown that zombies are conceivable. But this isn't right. The zombie world is stipulated to lack nonphysical phenomenal properties but zombies may have physical phenomenal properties.

[13] I am grateful to Richard Chappell for pressing this point.

zombie argument and the zoombie argument. This alleged dis-analogy resides in the fact that no self-respecting dualist can accept NP as a complete list of the nonphysical properties. This is because the dualist will insist that a complete list of the nonphysical properties that I in fact have will explicitly include qualitative properties. NP is thus contentious as it already assumes that dualism is false, they allege, and so the zoombie argument really does beg the question against the dualist. On the other hand all the parties to the zombie debate, this objection proceeds, agree on the physical description specified in P; the dualist is happy to let the physicalist fill in P with any physical theory they like. So the traditional zombie argument does not beg the question in the way that the zombie argument does.

Suppose for a second that we grant that there is this difference between the zoombie and zombie arguments does it then follow that the traditional zombie argument is not a bad argument? It is not obvious that it does. The dualist can presumably make sense of whatever way we fill out NP though they will not think that it is ideally conceivable. That is, the issue here is of the conceivability of  $NP \ \& \ \sim Q$ . I can't see any reason to think that it isn't conceivable unless you already think that qualitative properties are in fact nonphysical. Compare the property of being a table or of being water. Chalmers thinks that water facts are deducible a priori from a complete microphysical description of the actual world so someone who claimed to be able to conceive of all of the physical facts (and so there being H<sub>2</sub>O just like there is here) without water facts would be mistaken. He can grasp what the opponent seems to be conceiving when they say " $P \ \& \ \sim W$ " is conceivable' but he denies that it is actually conceivable because it contains a contradiction. So too in the zoombie case.

But is it right that the zoombie argument begs the question in a way that the zombie argument doesn't? The basic objection was that the way the zoombie argument is set up rules out dualism from the beginning and so is unfair to the dualist. It is true that the dualist allows the physicalist to fill in the placeholder description in P with whatever theory of physics they want but the problem lies not with P but with Q. The upshot of the discussion about dental fear and pain asymbolia was that when the dualist says that  $(P \ \& \ \sim Q)$  is semantically neutral they are either wrong or do not threaten physicalism. When they go to explicitly fill in the placeholder ' $\sim Q$ ' with statements like 'RB is consciously having a pain' they assign a semantics to terms like 'pain' where the primary and secondary intensions are identical whereas a physicalist like me will assign those terms a semantics just like other natural kind terms (where these intensions are not identical).

If you really were to remain neutral on this semantic issue the conceivability of  $(P \ \& \ \sim Q)$  is no threat to physicalism since one cannot then rule out that the  $Q$  which are lacking from the zombie world are not the same qualitative properties we have here. The dualist may be conceiving of a world physically identical to ours and which lacks a different kind of qualitative property than the one we have around here. This does not threaten physicalism since this world does not lack the kind of qualitative properties that the actual world has. So to get the zombie argument off the ground you must assume a semantics for the terms in  $Q$ , just as to get the zoombie argument off the ground you have to assume a semantics for  $NP$ . This may be more obvious in the zoombie case but it was designed to highlight the flaw. Another way to make the point is that they assume that the properties in  $Q$  are not identical to the properties in  $P$ , but this is exactly what is in dispute.

### 3.3. *To the Shombie Argument*

Shombies are creatures that are physically identical to me and which have consciousness and lack all nonphysical properties. One objection to shombies might be that it amounts to no more than asserting

1. It is conceivable that physicalism is true
2. Physicalism is a modal thesis and if true at any world physically just like ours it is true at all possible worlds that are physically just like ours
3. Therefore since it is true at one possible world it is true of our world.<sup>14</sup>

This way of putting it makes the shombie argument sound like a version of the ontological argument as advanced by people like Plantinga. But this kind of argument isn't very interesting, one might think, because it is not as though we have found something from which the truth of physicalism follows. We've simply insisted that it is true. One thing to note before we address this is the fact that this is exactly what is going on in the traditional zombie argument. When one asserts that zombies are conceivable one asserts

1. It is conceivable that physicalism is false.
2. Physicalism is a modal thesis and so if true would be true at all possible worlds physically just like ours.
3. Since there is one world where physicalism is false it is false of our world.

---

[14] Thanks to Robert Howell and Dave Chalmers for pressing this in comments at the conference.

But even if one is not moved by this there is a response that adapts a strategy that Chalmers uses in response to Yablo's meta-modal argument (Chalmers 2009). Yablo argued that it is seemingly conceivable both that a necessary being (God) exists and doesn't exist and if so then conceivability isn't a good guide to possibility.<sup>15</sup> Chalmers argued that what he was doing was merely conceiving of one particular possible world and not the entire space of possibilities. In fact he argues that our intuitions about modality become less trustworthy when we try to make meta-modal claims about the entire space of possibilities (i.e. as we do when we say that it is possible that some necessary truth is true). So too, I am merely conceiving of one possible world. When we conceive of the shombie world we only conceive of a world physically and qualitatively just like ours but which is completely physical in nature. We are not making meta-modal claims about the space of possibilities. True, physicalism is a modal thesis and so if true at one world it follows that it is true for all worlds that are physically identical to it, but I don't need to conceive of the shombie world that way. All that I need to do is to conceive of the shombie world is to conceive of one world physically identical to our world and having a creature there who has conscious experience in exactly the same way that I in fact do. This is not to conceive of physicalism being true and so not to employ the argumentative strategy criticized above. This is because the shombie argument is only designed to show that dualism is false, not that physicalism is true. For dualism to be true there must be a world that is physically identical to ours which lacks qualitative consciousness. Shombies show that there is no such world since the world that is physically identical to ours is a world of conscious experience. So just the conceivability of one possible world is in question and that is enough to show that dualism is false. The modal aspect of physicalism comes from independent considerations about the necessity of identities.

#### *3.4. To the Reverse-Knowledge Arguments*

We see many of the same objections with the reverse-knowledge arguments and so I will discuss both of them together in this final section. For instance one might think that the argument based on Maria is question begging in the sense that it is a one-premise argument. Something like Maria learns that dualism is false, therefore dualism is false. The original Mary argument doesn't do this because, one may think, it

---

[15] I would, of course, suggest that only one of these is really conceivable the other just merely appears to be conceivable.

only claims that Mary learns *something* or other about red and given her circumstance this is supposed to entail that physicalism is false.<sup>16</sup> But this is not exactly fair. The dualist thinks that Mary learns a lot about red when she gets out of her black and white room. She will learn how the nonphysical color properties, say, correlate with the physical properties of the brain for instance. So the original Mary does depend on Mary learning one particular thing, she learns that there is more to color perception than physical properties. So, both of the arguments beg the question this way. This is, of course, the point I'm trying to make! The Mary argument depends on our having a certain intuition that learning what it is like to see red would be truly learning something unaccounted for by the physical facts. So too in the Maria case I have the intuition that Maria will learn that there is nothing more to consciousness than the physical. In this sense the two arguments are on a par. Mary learns that there is more to reality than the physical. Maria learns that there is nothing more to reality than the physical.

So, the reverse knowledge argument from Maria shows that qualitative facts are conceivably deducible from a completed physics. That is, Maria having only phenomenal concepts and then introduced to the completed physical theory will be able to tell a priori when phenomenal concepts apply in other cases. One may object to the notion that Maria's deduction is a priori. After all, she is able to do so only because she has had the relevant phenomenal experience. But this is no objection, at least not for the dualist like Chalmers who have argued that empirical knowledge, as long as it plays only a causal role (for example that needed for concept acquisition) and not a justificatory role, is no bar to a deduction being a priori (Chalmers & Jackson, 2001). According to them as long as the concept only facilitates the deduction and does not justify it the deduction will still count as a priori. So Maria needs to have the experience in order to acquire the phenomenal concepts but once she has those and the completed physics the deduction will be a priori since she will not need to use those concepts to justify any step in the deduction. She will instead use the identities that she learned a posteriori.

Chalmers briefly mentions a strategy like this (Chalmers, 2004) calling it 'one of the more powerful replies available to the materialist.' He there lists several *prima facie* objections to this strategy. His first objection is to fall back on the zombie intuition. But as we have seen this doesn't help. Secondly he wonders whether someone like

---

[16] Robert Howell pressed this objection in his comments at the conference.

Maria will be able to deduce that other creatures, like bats and Martians, have phenomenal experience. But there is no non-question begging reason to think that Maria will not be able to do this. The same is true for his third response along the lines that Mary might have a concept like phenomenally indistinguishable and yet be unable to tell if two people were having the same qualitative experience. Finally he objects that someone like Maria will have to crucially rely on introspection in order to complete the deduction. And since introspection yields a posteriori knowledge the deduction will not be a priori. But it is not clear why she would need to rely on introspection in this way. Once the relevant concept is acquired the deduction goes through just as in the water H<sub>2</sub>O case. At least, there is no a priori reason to think otherwise.

Let us finally, then, address Mark. Nagasawa argues that this argument is successful against any kind of reductive dualism, whether property or substance. A reductive dualist holds that qualitative facts can be reduced to or deduced from a set of nonphysical facts, like the protophenomenal facts that Chalmers mentions in his panprotopsychism (Chalmers 2001). The argument, however, is not successful against a non-reductive version of dualism. The non-reductive dualist claims that qualitative properties cannot be reduced to any other kind of property whether protophenomenal or not. The knowledge argument against dualism, according to Nagasawa, shows that non-reductive dualism is the only viable option for someone who wants to use the knowledge argument against physicalism. One could then reason that Mary merely shows that a reductive version of physicalism isn't viable but that a non-reductive physicalism is (like a Davidsonian Anomalous Monism perhaps).

On the other hand one might think that Mark shows that a dualistic theory that does not hold that the phenomenal concepts must be experienced in order to be fully apprehended is just as doomed as a physicalism that does not hold this. If this is so then Mark can't complete the deduction until he acquires the relevant phenomenal concepts. The two theories are on an a priori equivalency here. I suspect that a dualist like Chalmers will argue that once Mark sees red he will be able to deduce, a priori, the qualitative facts from the nonphysical facts and this is the way that my intuitions lie with physicalism. Thus the traditional and reverse-knowledge arguments together suggest that with the phenomenal concepts in hand we can make the deductions and that without them we are unable to make the deductions. But just because we must actually see red in order to fully acquire the

concept of phenomenal red is no objection to physicalism.<sup>17</sup> Dualism is committed to the same thing. The dualist and the physicalist agree that phenomenal concepts are the kinds of things which depend on our having had the relevant experience. And if physicalism is true then all that means is that the property that we must have in order to have a full concept of it is an interesting physical property; this by itself cannot be an objection unless one has just assumed that these kinds of properties are not ultimately entirely physical.

#### 4. Conclusion

So as we have seen the use of a priori methods to determine whether physicalism is false simply fail to do anything except to tell us where sympathies lie. What theory one accepts, either tacitly or not, as true influences what one finds to be *prima facie* conceivable. This is the only way that we can explain why it is that some people find zombies conceivable while others find zoombies conceivable that is consistent with our desire to maintain a link between conceivability and possibility. Once the parity arguments are on the table there is nothing short of table pounding that can decide which one is really ideally conceivable and which one is merely *prima facie* conceivable. In order to answer that question we first need to know is whether physicalism is true or false and in order to answer that question we need more empirical data. Only once we know whether physicalism is actually true or false will a priori methods yield anything; but of course by then we won't need them!

#### References

- Balog, K. (Manuscript), 'Illuminati, zombies, and metaphysical gridlock'.  
 Block, N. & Stalnaker, R. (1999), 'Conceptual analysis, dualism, and the explanatory gap', *Philosophical Review*, **108**, pp.1–46.  
 Brown, R. (2006), 'What is a brain state?', *Philosophical Psychology*, **19** (6).  
 Chalmers, D.J. and Jackson, F. (2001), 'Conceptual analysis and reductive explanation', *Philosophical Review*, **110**, pp. 315–61.  
 Chalmers D.J. (2003), 'Consciousness and its place in nature', in Stich, S. and Warfield, F. (eds) *Blackwell Guide to the Philosophy of Mind* (Oxford: Blackwell).  
 Chalmers, D.J. (2004), 'Phenomenal concepts and the Knowledge Argument', in Peter Ludlow, Yujin Nagasawa & Daniel Stoljar (eds.), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* (Cambridge, MA: MIT Press).

---

[17] Robert Howell defends a similar view that he calls 'subjective physicalism'.

- Chalmers, D.J. (2009), 'The two dimensional argument against materialism', in McLaughlin, B.P. and Beckermann, A. (eds) *The Oxford Handbook of Philosophy of Mind* (Oxford: Oxford University Press).
- Chapman H.R. & Kirby N.C. (1999), 'Dental fear in children: A proposed model', *British Dental Journal*, **187** (8), pp. 408–12.
- Churchland, P.M. (1989), 'Knowing qualia: A reply to Jackson', in Block, N., Flanagan O. and Güzeldere, G. (ed. 1997), *The Nature of Consciousness* (Cambridge, MA: MIT Press), pp. 571–77.
- Devitt, M. (forthcoming), 'No place for the a priori', in Michael J. Shaffer and Michael Weber (eds.) *New Views of the A Priori in Physical Theory* (Amsterdam: Rodopi Press).
- Frankish, K. (2007), 'The anti-zombie argument', *Philosophical Quarterly*, **57** (229), pp. 650–66.
- Grahek, N. (2001), *Feeling Pain and Being in Pain* (Cambridge, MA: MIT Press).
- Jackson, F. (1986), 'What Mary didn't know', *Journal of Philosophy*, **83**, pp. 291–95.
- Kripke, S. (1980), *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Marton, Peter (1998) 'Zombies versus materialists: The battle for conceivability', *Southwest Philosophy Review*, **14** (1), pp.131–38.
- Nagasawa, Y. (2002), 'The Knowledge Argument against dualism', *Theoria*, **68** (3), pp. 205–23.
- Needham, P. (2002), 'The discovery that water is H<sub>2</sub>O', *International Studies in the Philosophy of Science*, **16** (3), pp. 205–26.
- Piccinini, G. (2006), 'Access denied to zombies', Unpublished manuscript. Presented at Tucson VII - Toward a Science of Consciousness 2006, Tucson, AZ.
- Read, J. (1995), *From Alchemy to Chemistry* (New York: Dover Publications).
- Rosenthal, D. (2005), *Consciousness and Mind* (New York: Oxford University Press).
- Sturgeon, S. (2000), *Matters of Mind: Conscious, Reason, and Nature* (London: Routledge Press).