

## Illuminati, zombies and metaphysical gridlock<sup>1</sup>

Katalin Balog  
Rutgers Newark

During the last two decades there has been a resurgence of arguments against physicalism and for varieties of metaphysical dualism. The conclusion of these arguments is that phenomenal consciousness is absent from a world that is purely physical. Many contemporary philosophers of mind have found some of these arguments to be persuasive and have opted for some form of anti-physicalism.<sup>2</sup> In this paper I will survey the landscape of these arguments and physicalist responses to them. The anti-physicalist arguments that I discuss start from a premise about a conceptual, epistemic, or explanatory *gap* between physical and phenomenal descriptions<sup>3</sup> and conclude from this – on *a priori* grounds – that physicalism is false.<sup>4</sup> I call these arguments

---

<sup>1</sup> I would like to thank Ned Block, David Chalmers, Troy Cross, Tamar Gendler, Barry Loewer, Raymond Martin and Gilad Tanay, participants of a discussion group at NYU, and an audience at the Rutgers University philosophy department for comments and helpful discussion of this paper.

<sup>2</sup> David Chalmers' recently compiled list of contemporary anti-physicalists includes Joseph Almog, Torin Alter, George Bealer, Laurence Bonjour, Paul Boghossian, Tyler Burge, Tim Crane, John Foster, Brie Gertler, George Graham, W.D. Hart, Ted Honderich, Terry Horgan, Steven Horst, Jaegwon Kim, Saul Kripke, Harold Langsam, E.J. Lowe, Kirk Ludwig, Trenton Merricks, Martine Nida-Rümelin, Adam Pautz, David Pitt, Alvin Plantinga, Howard Robinson, William Robinson, Gregg Rosenberg, A.D. Smith, Richard Swinburne, and Stephen White. From his BLOG "Fragments of Consciousness" September 26, 2005.

<sup>3</sup> Phenomenal descriptions attribute phenomenal properties to experience (and perhaps to thought) in the sense of there being something *it is like* to undergo an experience, something one can normally introspect, e.g., the feeling of my fingers flexing that (partly) characterizes my present bodily sensation. I will assume throughout the paper that there are phenomenal properties in this sense. For eliminativism about phenomenal properties, see, e.g., Rey (2007).

<sup>4</sup> These arguments include, among others, arguments based on conceivability considerations by Kripke (1972), Nagel (1974), Bealer (1994), Chalmers (1996, and 2009), as well as the Knowledge Argument of Jackson (1982), versions of the Property dualism Argument in Robinson (1993), White (2007), and Nida-Rümelin (2007), and the Explanatory Gap Argument in Levine (2001) and (2007).

“conceivability arguments”. Although not all of them have a premise concerning conceivability,<sup>5</sup> they all, if successful, establish that it is *inconceivable* that phenomenal experience exists in a *purely* physical world. My first aim is to develop a master argument to counter these arguments. Along the way I develop a new version of one of these arguments – Chalmers’ Zombie Argument – that is immune from objections that I have previously urged against the original argument, based on a distinction between positive and negative conceivability. However, I will show that there is a master argument that provides an adequate physicalist response both to the new Zombie Argument and to the other anti-physicalist conceivability arguments. The master argument is crucially bolstered by what has come to be known as the Phenomenal Concept Strategy<sup>6</sup>; this strategy – following Brian Loar’s original proposal in (1997) – appeals to the special cognitive features of phenomenal concepts in providing a physicalistically respectable explanation of the aforementioned gaps. What is new in my appeal to the Phenomenal Concept Strategy is to harness the strategy – together with holistic considerations about which metaphysical framework is more virtuous in terms of simplicity and explanatory power – to be part of a formal argument to be used to rebut any gap-based argument. I propose such an argument and consider objections to it from the anti-physicalist side.

In the second part of the paper I assess the dialectical situation created by such a general physicalist reply to the anti-physicalist arguments, and argue that, despite the fact that a satisfactory reply can be given to the conceivability arguments, there is a puzzling symmetry between dualist attacks on physicalism and physicalist replies. Each position can be developed in a way to defend itself from attacks from the other position; I will argue that there are neither a priori nor a posteriori ways to decide between the two.

The plan is as follows. Section I provides background for the conceivability arguments. Section II discusses David Chalmers’ Zombie Argument and my refutation of that argument. In Section III and IV I develop a new interpretation of Chalmers’ idea of positive conceivability<sup>7</sup> and formulate a version of Chalmers’ argument that – along with a number of other conceivability arguments - is immune to my refutation of Chalmers’ original argument. In section V and VI, I develop a physicalist master argument – I will call it the *Counter Conceivability Argument* – that takes into account these refinements and rebuts all versions of the conceivability argument. In Section VII, I discuss a physicalist strategy, the Phenomenal Concept Strategy, which complements and supports the Counter Conceivability Argument. In the final section of the paper I argue that there is a structural similarity between the anti-physicalist and physicalist strategies: they are both able to defend themselves from the attacks of the other

---

<sup>5</sup> I will later introduce and explicate two notions of conceivability.

<sup>6</sup> Stoljar (2005) introduced this phrase.

<sup>7</sup> My explication of positive conceivability differs from that of Chalmers (2002a).

side, but they both can be viewed as question-begging from the other side. In conclusion I will examine what this means for the status of the mind-body problem.

## I. Metaphysical background

The debate between physicalism and anti-physicalism is a debate about fundamental ontology. According to *physicalism*, the world's fundamental ontology is physical.<sup>8</sup> It is not easy to say exactly what makes fundamental entities and properties "physical." But this isn't a problem since it suffices for our discussion that physicalism is understood as requiring that *fundamental* physical properties and entities are "non-mental." So if physicalism is true then fundamental physical properties and entities do not exhibit intentionality and consciousness (and they do not even exhibit proto-mental proto-intentionality or proto-consciousness); intentionality and consciousness is instantiated only in macroscopic systems *in virtue of* immensely complex arrangements of fundamental properties and entities and their causal/nomological features; i.e. for biological individuals in virtue of brain states and processes.

Anti-physicalism comes in a number of different varieties.<sup>9</sup> I will be explicitly concerned with non-interactionist property dualism.<sup>10</sup> Non-interactionist property dualists don't necessarily deny mental causation but they – unlike interactionist dualists – accept the causal closure of physics. According to non-interactionist property dualism the fundamental ontology of the world includes mental and/or proto-mental properties.<sup>11</sup> This means that arrangements of fundamental

---

<sup>8</sup> Contemporary physicalists typically hold that the best account of that ontology is provided by fundamental physics. Physics' best hypotheses about fundamental ontology is that it consists of elementary particles, strings and/or fields occupying a space-time structure, and possessing a limited number of quantitative properties (mass, charge, electromagnetic potential, and so on). Physics also claims that there are only a few fundamental dynamical and perhaps non-dynamical laws that govern the structure of space-time and the evolution of its occupants.

<sup>9</sup> The usual suspects include idealism, interactionist property or substance dualism, non-interactionist property or substance dualism and Russellian Monism.

<sup>10</sup> I find substance dualism problematic for reasons that are beyond the scope of this paper. Interactive property dualism is implausible since it requires a denial of the causal closure of physics. I will not be explicitly concerned with Russellian monism in this paper; but I think the points I make with respect to non-interactionist property dualism can be equally made with respect to Russellian monism as well.

<sup>11</sup> Proto-mental properties are not full-blown mental properties; but they are constituents of full-

physical entities and their physical properties are not metaphysically sufficient for the instantiation of phenomenal properties. Non-interactionist property dualism grants that the physical realm is nomologically closed, and it posits the existence of fundamental vertical laws<sup>12</sup> that connect arrangements of physical entities and properties to phenomenal properties. From now on, by “dualism” I mean this view.

Following Frank Jackson, I will assume that there is a fundamental vocabulary (although not necessarily in our language as it is currently) in which there is a complete fundamental true description of the world. This description specifies the total spatio-temporal distribution of fundamental entities, the totality of instantiations of fundamental properties and relations, and the fundamental laws. If physicalism is true then none of the elementary vocabulary refers to mental entities or properties. Jackson pointed out that a necessary condition for the truth of physicalism in a world is that all positive truths in that world, including, as it may be, positive truths about phenomenal consciousness,<sup>13</sup> are metaphysically necessitated by the complete physical truth about that world.<sup>14</sup> It follows that if physicalism is true in our world, the *Physicalist Entailment*

---

blown mental properties.

<sup>12</sup> I will assume that these laws are contingent; i.e., not metaphysically necessary. If laws are taken to be metaphysically necessary then it is difficult to state the difference between physicalism and dualism since then both would hold that configurations of physical property instantiations metaphysically necessitate mental property instantiations.

<sup>13</sup> A positive phenomenal statement says that a phenomenal property is instantiated; e.g., *Joe is feeling an itch*. Negative truths, like *There are no angels*, and global statements, like *Every gold cube has a volume smaller than one cubic meter*, are not metaphysically necessitated by the complete physical truth about the world *P* although they are necessitated by *P* and a clause that says that *P* is the whole fundamental truth. However, the phenomenal and physical truths we will be interested in are all positive truths so I will ignore this complication for the remainder of the paper.

<sup>14</sup> This formulation is based on Jackson’s (Jackson 1993). The first precise formulation of physicalism along these lines is due to Lewis (Lewis 1983a). Subsequent discussions are variations on the same theme. Many philosophers, among them non-physicalists, accept this formulation as capturing a very important component of the intuitive idea of physicalism. But it doesn’t express the full physicalist commitment – only a necessary condition – because it is apparently compatible with certain ontologies that are intuitively non-physicalist e.g., with one in which there are fundamental mental as well as fundamental physical properties connected by “brute” necessary connections.

*Thesis* is true:

(Phys)  $\forall T \Box (P \supset T)$ .<sup>15</sup>

If there are mental truths – for example, that Mary knows what it is like to see red – that are not necessitated by the complete physical description then physicalism is false.

## II. The Zombie<sup>16</sup> Argument and the Zombie Refutation

There is a line of argument going back at least to Descartes' argument for the distinctness of mind and body that claims to show that physicalism is false. In fact, these arguments can be understood to conclude, on the basis of *a priori* considerations, that *no* world where phenomenal properties are exemplified can be a purely physical world. The descendent of this argument that has received the most attention in the last decade is David Chalmers' "Zombie Argument."

Chalmers' (Chalmers 2009) most recent formulation of the zombie argument is as follows:

### *The Zombie Argument*

- 1)  $P \& \sim Q$  is conceivable.<sup>17</sup>
  - 2) If  $P \& \sim Q$  is conceivable then  $P \& \sim Q$  is metaphysically possible (CP principle)
  - 3) If  $P \& \sim Q$  is metaphysically possible then physicalism is false.
- 
- 4) Physicalism is false.

---

<sup>15</sup>  $\forall$  is a substitutional quantifier,  $T$  is a statement variable for true positive statements,  $\Box$  is the metaphysical necessity operator, and  $P$  is the complete fundamental physical description of the world, including the fundamental physical laws.

<sup>16</sup> Zombies are creatures that are physically identical to normal humans but have no phenomenal experiences whatsoever.

<sup>17</sup>  $P$  is the complete fundamental physical description of the world, including the fundamental physical laws, and  $Q$  is a positive phenomenal truth, e.g., that someone is having a visual experience with a particular phenomenal character at a particular time.

By “statement  $S$  is conceivable” Chalmers (1996) means “ $S$  cannot be ruled out a priori”.<sup>18</sup> Thus conceivability for him is not a psychological but an epistemological notion. It may be psychologically impossible for us to imagine or see clearly that space has 11 dimensions but that doesn’t mean that it is inconceivable since it cannot be ruled out *a priori*. This is the basic notion of “conceivable” that I will be working with throughout much of the paper. Chalmers (2002) distinguishes it from another notion of conceivability – positive conceivability – which will also play a crucial role later on.

Chalmers (2009) introduces some clarifications and emendations to this simple argument. Chalmers employs the two dimensional semantic framework to characterize “primary possibility” and “secondary possibility” and argues for his Master Principle (MP) that conceivability implies primary possibility.<sup>19</sup> The CP Principle (premise 2 of his argument) follows from this, on the plausible assumption that both  $P$  and  $Q$  express the same primary and secondary proposition. On the other hand, if we assume that  $P$  has *different* primary and secondary propositions, and assume that  $P \& \sim Q$  is *not* possible, Russellian Monism follows. Premise 2, accordingly, should be modified to read

2’) If  $P \& \sim Q$  is conceivable then  $P \& \sim Q$  is metaphysically possible or Russellian monism is true,

and the conclusion to

4’) Physicalism is false or Russellian monism is true.

Since my response to the Zombie Argument can be shown to block either premise, or either conclusion, I will ignore this issue and stick with the simplified version of the argument. The Zombie Argument is valid. Premise (3) is entailed by the proposition that Phys is a necessary condition for Physicalism.<sup>20</sup> Philosophers who think that there is a functional or representational analysis of phenomenal consciousness reject (1).<sup>21</sup> But I agree with Chalmers

---

<sup>18</sup> Chalmers introduces a battery of conceivability concepts (2002a). For my purposes I focus only on this one and a development of what he calls “positive conceivability” I introduce later in this paper.

<sup>19</sup> For an elaboration, see Chalmers 2002a.

<sup>20</sup> *Phys* states that for all true positive statements  $T$ ,  $\Box (P \supset T)$ , so if  $P \& \sim Q$  is metaphysically possible then (Phys) is false and therefore physicalism is false.

<sup>21</sup> E.g. Lewis 1966 and Jackson 2003. Analytic functionalism or representationalism concerns the *meaning* of phenomenal terms; it says that such meanings can be analyzed in functional or

that there is no functional or representational analysis of phenomenal consciousness and that no physical description *a priori* entails any positive phenomenal description. Later I will offer some considerations based on the nature of phenomenal concepts for why this is so but for now I will just assume that (1) is true. So for both Chalmers and myself the crucial premise in the argument is (2).<sup>22</sup>

My first argument against (2) – “the Zombie Refutation” – has the form of a reductio.<sup>23</sup> It is based on the claim that *if* the Zombie Argument is sound then there is a possible zombie world – a world physically exactly like our world but where there is no phenomenal experience – in which Zombie Chalmers gives an analogue of the Zombie Argument and concludes that physicalism is false.

Zombie-Chalmers’ argument goes like this:

*The Zombie\* Argument*

- 1\*)  $P \& \sim Q^*$  is conceivable.
  - 2\*) If  $P \& \sim Q^*$  is conceivable then  $P \& \sim Q^*$  is metaphysically possible.
  - 3\*) If  $P \& \sim Q^*$  is metaphysically possible then physicalism is false.
- 
- 4\*) Physicalism is false.

representational terms. *Pain*, e.g., according the analytic functionalism, has a conceptual role that connects it (in the meaning-constituting way) with complex concepts like *typically caused by injury*, *typically causes avoidance behavior*, *typically cases saying “ouch”*, etc. Analytic functionalism or representationalism rebuts the conceivability arguments by denying the conceptual, epistemic, and explanatory gaps between physical and phenomenal descriptions. Analytic functionalism/representationalism, of course, has to explain why there *seem* to be such epistemic gaps when in reality there aren’t. See also Kirk (2005) for an interesting argument against the conceivability of zombies whose grounds go beyond analytic functionalism.

<sup>22</sup> Chalmers’ defense and development of the two-dimensional framework and of the conceivability-possibility link can be found in Chalmers and Jackson 2001, Chalmers 2002a, and Chalmers 2004. There are important discussions in Block and Stalnaker (1999), Yablo (1993, 2002), and Soames (2004). I briefly discuss what I think goes wrong with a related argument by Frank Jackson (Balog 2001).

<sup>23</sup> See Balog 1999. I briefly recap that argument here to illustrate its scope and – what I later realized as – its limitations.

Here is how the Zombie\* Argument helps refuting Chalmers' Zombie Argument. The Zombie\* Argument is valid but its conclusion is false (by assumption, physicalism is true in Zombie Chalmers' world). It follows that at least one of the argument's three premises is false. I will argue that 1\* and 3\* are true and so 2\* is false. But if 2\* is false then 2 of Chalmers' Zombie Argument is false as well so Chalmers' Zombie Argument is unsound.

First, Premise 1\*.  $Q^*$  is a truth in the conceivable zombie world that corresponds to  $Q$  in our world. For example, if  $Q$  is the truth that JUDIT IS HAVING A RED SENSATION AT T, then  $Q^*$  is the truth that ZOMBIE-JUDIT IS HAVING A RED SENSATION\* AT T.  $Q^*$  is a truth since, given that the zombie-world is physically identical to ours, whenever Judit tokens RED SENSATION, zombie-Judit tokens an analogous – though different – concept RED SENSATION\*. Assuming that the physical is causally closed, and, as is plausible, that phenomenal states have neural correlates, RED SENSATION\* will be correlated with the same brain state B that RED SENSATION is correlated with and – as I argue in more detail later – RED SENSATION\* will *refer* to B. This means that if  $Q$  is true,  $Q^*$  is true as well, even though Judit and zombie-Judit refer to different states: whenever Judit has a red sensation – thereby rendering  $Q$  true – she will also be in brain-state B from which it follows that zombie-Judit will have a red sensation\* (i.e., will be in brain state B) as well.<sup>24</sup>

Assuming that i) that zombies have intentional states, largely coinciding with ours (the only differences involve phenomenal\* concepts), and ii) that – as mentioned above – phenomenal\* concepts refer to brain states of the zombies, what can we say about the conceivability of  $P \& \sim Q^*$ ? I suggest that zombies can conceive of a statement just in case we can conceive of the analogous statement.

Shelly Kagan and others have suggested in conversation that this is not correct since, even though we can conceive of  $P \& \sim Q$ , for any phenomenal statement  $Q$ , zombies cannot conceive of  $P \& \sim Q^*$  if  $Q^*$ , as it has been assumed, expresses a *physical* state of affairs true in the zombie world which therefore is already *contained* in P. An ideal (zombie) conceiver would figure it out that  $Q^*$  is already *contained* in P. However, even if the *state of affairs* that  $Q^*$  expresses is physical, zombies don't conceptualize it *as* physical. Moreover, zombies' phenomenal\* concepts have parallel conceptual roles to our phenomenal concepts in every relevant respect so they can no more rule out  $P \& \sim Q^*$  a priori than we can rule out  $P \& \sim Q$  a

---

<sup>24</sup> Of course, I do not here *endorse* the possibility of a zombie world and so the possibility of Zombie-Chalmers referring to non-phenomenal brain states with his phenomenal\* concepts. According to physicalism, those brain states in question *are* phenomenal states, and so zombies are mere “conceivabilia”. I am merely claiming that the proponent of the Zombie Argument who *endorses* the possibility of a zombie world must describe this zombie world as one where Zombie Chalmers refers to non-phenomenal brain states by his phenomenal\* concepts.

priori. If, as is plausible, phenomenal concepts pick out their referent *directly*, and not via physical, functional, etc. modes of presentation, zombies' phenomenal\* concepts also refer directly, and as a result, just as we can conceive of zombies, conceivable zombies can conceive of zombies\* as well, *in spite of the fact that*  $Q^*$  expresses a physical state of affairs already included in  $P$  and so  $P \& \sim Q^*$  is impossible.

I suggest that premise 1\* of the Zombie\* Argument is true. 3\* follows straightforwardly from the formulation of physicalism that provides the framework for the debates so it is true by definition. Since the argument is valid, and its conclusion is false, premise 2\* must be the culprit. 2\*, as I mentioned earlier, follows from Chalmers Master Principle (MP): that conceivability implies primary possibility. Both (2\*) and (2) follow from it equally, on the assumption that  $P \& \sim Q^*$  as well as  $P \& \sim Q$  has identical primary and secondary intensions. This is reasonable since the conceptual roles of, e.g., RED SENSATION, and RED SENSATION\* are parallel in all the relevant aspects so if  $Q$  has identical primary and secondary intensions so does  $Q^*$ .<sup>25</sup> Consequently, if 2\* is false then MP is false, but if MP is false, 2 loses its plausibility. Though my argument is not – as I earlier thought (Balog 1999) – strictly speaking a reductio, it is close enough. If Chalmers formulated the argument in terms of MP instead of in terms of one of its implications, i.e., premise 2 – the argument could indeed be turned into a strict reductio.<sup>26</sup>

The main objection to this argument comes from Chalmers (2009). He argues that there are no phenomenal\* statements  $Q^*$  that are true in the zombie-world, since the phenomenal\* concepts of zombies don't refer. If this is so, the argument cannot even get off the ground; premise (3\*) of the Zombie\* Argument would be either false or non-sensical.

To support his point, Chalmers argues that while eliminativists in the actual world are wrong, zombie-eliminativists are right.<sup>27</sup> This appears convincing at first: after all, by

---

<sup>25</sup> For reasons I mentioned before, I am putting aside the issue of whether the primary and secondary intension of  $P$  coincides.

<sup>26</sup> There is another way to run the Zombie Refutation I have just laid out. Instead of assuming the soundness of the zombie argument in the service of a reductio type argument, one can simply *conceive* of a purely physical world where zombies find the statement ( $P \& \sim Q^*$ ) conceivable even though  $P \& \sim Q^*$  is not really possible. On this reading, the Zombie Refutation shows that MP is conceivably false. But since MP is supposed to be a priori, it means that it *is* false. The rest of the argument goes the same way.

<sup>27</sup> Chalmers' way of arguing seems to indicate that his proposal concerns the lack of reference for phenomenal\* concept, not a lack of *meaning*. If they lacked meaning, there would be no Zombie\* Argument; just noises on the part of Zombie Chalmers; but neither would there be zombie eliminativists. On the other hand, it is hard to see how phenomenal\* concepts *could* be meaningful while lacking reference, given that they don't seem to have any mode of presentation at all. In any case, I will respond to the no reference view; but my response is equally applicable

assumption, there are no phenomenal states in the zombie-world, so eliminativism must be true there. But eliminativism about phenomenal states in the zombie-world is beside the point. Zombie eliminativism and eliminativism are different. When a zombie eliminativist says “there are no phenomenal\* states”, she doesn’t mean there are no *phenomenal* states<sup>28</sup> – by assumption, those states do not occur in her world and there is no way she could refer to *them*. What she denies is the existence of *phenomenal\** states; and it is far from obvious that eliminativism about *those* is true. Once you realize this, the appeal of zombie-eliminativism largely evaporates.

It is true, my argument presupposes that Zombie-Chalmers’ views about pain\* (and ontology in general) are gravely mistaken – that he thinks that pain\* is a *non-physical/non-functional state* whereas it *is* a physical state after all. This, however, doesn’t make zombie-Dennett right. It is not implausible that people can be *this* wrong about the nature of their own states. If the physicalist is right, Chalmers is mistaken in the very same way. Paul Bloom (2004) has argued that dualism is part of our innate theory of mind and it certainly is a natural and intuitive view. If so it will be natural for zombies as well. Physicalism cannot be proven wrong just because it implies that we tend to have some false views about consciousness; similarly, zombie-realism cannot be proven wrong just because it implies that zombies tend to be mistaken about consciousness\*.

So Chalmers’ appeal to the plausibility of zombie-eliminativism doesn’t work. To the contrary, it is quite plausible that the zombies’ concept PAIN\* refers to the brain state with which it is reliably correlated. This is true on most accounts of reference that have been taken seriously.<sup>29</sup> But before we go on, let me briefly mention two accounts that seem to support the

---

to the no meaning view.

<sup>28</sup> This interpretation of what Chalmers means is ruled out immediately by his views on phenomenal concepts (see Chalmers 2003); he thinks that (introspective) tokenings of a phenomenal concept involve tokens of the phenomenal property the concept refers to. By assumption, zombies cannot possess such concepts.

<sup>29</sup> On interpretationist accounts, like Davidson’s (1973) and Lewis’s (1983b), zombie-Chalmers will have phenomenal\* thoughts: his utterances about phenomenal\* states are just as interpretable as Chalmers’ utterances about phenomenal states. Similarly with other theories, like the informational account (see, e.g., Dretske 1988), the causal-historical account (Kripke 1972), the counterfactual account (Fodor 1990), or the teleosemantic account (Millikan 1989, Papineau 1993): Zombie-Chalmers’ concept PAIN\* carries information about brain state B; the right counterfactuals hold about the causal dependence of PAIN\* and brain state B, etc. A good case can be made on all these accounts that phenomenal\* concepts refer to zombies’ brain states. Conversely, the alternative – that phenomenal\* concepts are empty of reference – sits very uneasily with all these accounts.

idea that phenomenal\* concepts are empty of reference. One account appeals to the idea that phenomenal consciousness is essential for intentionality (see Searle 1992). Although the idea is not absurd, the argument for it does not seem to be very strong, and the contrary assumption seems far more intuitive. The other idea is that even if the reference of phenomenal concepts is not fixed descriptively, there is a *descriptive element* to them, one that entails that they cannot refer to anything physical (see Rey 1988 for a discussion of this idea). It would follow then that phenomenal\* concepts, e.g., PAIN\* could not refer to anything physical either.

I do not think that our concept has this descriptive commitment; and moreover, Chalmers doesn't seem to think that either. He assumes that we have epistemic warrant to attribute phenomenal states to ourselves on introspective evidence. If our very concept PAIN entailed that its reference is nonphysical, one wouldn't need not bother with the Conceivability Argument in the first place; one could simply argue for dualism on the basis of the claim that phenomenal states exists. But dualism doesn't come so cheap.

So I think my earlier argument stands. However, the Zombie Refutation points the way towards a different, stronger version of the Zombie Argument that cannot be tackled by the Zombie Refutation and so calls for a different response from the physicalist.

### III. Yogis and the Zombie Argument

I previously spelled out, in Balog (1999), an analogue of the Zombie Refutation called the Yogi Refutation. In the original paper I simply intended it as another refutation of MP, the core principle connecting conceivability and possibility that underlies the Zombie Argument. Yet now it seems to me that a more important lesson of the Yogi Refutation is that the analogue of premise 2, that is, the specific application of MP to the yogi case, is not even intuitively compelling. MP is stronger than Chalmers needs to get his conclusion. If it is weakened in the way I will propose my reductio no longer applies. Here is the Yogi thought experiment that explains why.

In a possible yogi world there are yogis who are like normal humans but have a few odd concepts, FLURG, FLORG, etc., that refer to brain states of theirs in an absolutely direct, unmediated way, unmediated even by phenomenal modes of presentation. They are recognitional concepts of an odd sort; yogis don't associate any criteria whatsoever with their concepts FLURG, FLORG, etc., they just pick out the corresponding brain states via some reliable mechanism that is nevertheless completely opaque to them. As a consequence, just like  $P \& \sim Q$  cannot be ruled out a priori,  $P \& \sim Q^{flurg}$  cannot be ruled out a priori either.

The Yogis can formulate the following argument:

*The Yogi Argument*

---

1<sup>yogi</sup>)  $P \& \sim Q^{flurg}$  is conceivable.<sup>30</sup>

2<sup>yogi</sup>) If  $P \& \sim Q^{flurg}$  is conceivable then  $P \& \sim Q^{flurg}$  is metaphysically possible.

Lemma<sup>yogi</sup>)  $P \& \sim Q^{flurg}$  is metaphysically possible.

3<sup>yogi</sup>) If  $P \& \sim Q^{flurg}$  is metaphysically possible then physicalism is false.

4<sup>yogi</sup>) Physicalism is false.

The conclusion of the Yogi Argument, i.e., that physicalism is false, is based on Lemma<sup>yogi</sup>, i.e. that  $P \& \sim Q^{flurg}$  is metaphysically possible which, by assumption, is false in the yogi world. Instantiating a flurg *is* instantiating a (physical) brain state and so  $Q^{flurg}$  – a true statement in the yogi world – is metaphysically necessitated by  $P$ . Though physicalism might be false in the yogi world for other reasons – e.g. because dualism about phenomenal experience is true – the fact that Lemma<sup>yogi</sup> is false shows that the argument is unsound. Since 1<sup>yogi</sup> is true, 2<sup>yogi</sup> must be false. But since 2<sup>yogi</sup> follows from Chalmers’ MP connecting conceivability and possibility in the same way as 2 does, if 2<sup>yogi</sup> is false, MP is false, and so Chalmers’ reason for 2 is undermined.

I now think, however, that the more interesting lesson of the yogi case is that yogis just *wouldn’t take this argument seriously!* Suppose that Yogi-Joe is undergoing a flurg at  $t$  and he believes this to be the case. Still, he would not conclude on the basis of the conceivability of  $P \& \sim (Yogi-Joe \text{ is undergoing a flurg at } t)$  that flurg is not physical. I propose that there is no intuitive appeal to this argument for a yogi. The reason is that yogis just doesn’t have enough of a grasp of flurgs to draw such far reaching conclusions from the conceivability of  $P \& \sim Q^{flurg}$ .

I think this shows that what makes the conclusion of the Zombie Argument – that consciousness is non-physical – so intuitively appealing is *not* the mere conceivability of zombies in the sense that their existence cannot be ruled out a priori. It is the conceivability of zombies *together* with the fact that we have a *grasp* of phenomenal states via phenomenal concepts.<sup>31</sup> By “grasp” of a property or state I simply mean that the concept in question is not “blind”, that it *presents* a property in some way that is cognitively significant. It is the same idea as is usually meant by a concept’s having a *mode of presentation* though some accounts of modes of presentation, e.g. their being merely syntactic properties of mental representations

---

<sup>30</sup>  $P$  is the full fundamental physical description of the yogi world, and  $Q^{flurg}$  is a true statement in the yogi world to the effect that somebody is undergoing a flurg.

<sup>31</sup> Nida-Rümelin (2007) uses the technical term “grasp” in a very different way. I do not mean to follow her usage here.

don't capture what I have in mind.<sup>32</sup>

I don't think a full account of what concepts are and how they operate is necessary to explain the idea. It is basically this. For most concepts, it is their conceptual role that is responsible for their cognitive significance. In the case of phenomenal concepts, it is not conceptual role but rather a characteristic phenomenal feel that guides the concept to its referent in a cognitively significant way. Thus when one refers to a toothache that one is having one does so via its awful throbbing. The tooth-ache, e.g., is grasped via that very throbbing sensation that makes it the kind of sensation it is. What is missing in yogis' concepts of FLURG, and what ultimately makes the Yogi Argument so unpersuasive is that yogis – while they refer to flurges – do not have a grasp of flurges in either of these ways.

#### IV. Positive and negative conceivability

At the time when Chalmers first proposed the Zombie Argument (Chalmers 1996), he appealed to a notion of conceivability according to which *S* is conceivable iff *S* cannot be ruled out a priori. Since then, he (Chalmers 2002) has come up with a different notion of conceivability that can be understood (although I am not saying that Chalmers so understood it) as taking into account the lesson of the Yogis. Chalmers (2002) distinguishes between two different notions of conceivability. He defines them as follows:

*S* is *negatively conceivable* when *S* cannot be *ruled out* through a priori reasoning. (p. 147)<sup>33</sup>

*S* is *positively conceivable* when one can imagine that *S*: that is, when one can imagine a situation that verifies *S*. (p. 148)

The notion of conceivability that is at play in the original Zombie Argument is negative conceivability (*conceivability<sup>neg</sup>*). As we have just seen, the Zombie Refutation refutes the Master Principle (MP) *understood as appealing to conceivability<sup>neg</sup>*, which in turn undermines Chalmers' reason for premise 2, *understood as appealing to conceivability<sup>neg</sup>*.<sup>34</sup> Let's call

---

<sup>32</sup> See, e.g., Fodor 1998.

<sup>33</sup> I take Chalmers to intend this as a definition not a conditional.

<sup>34</sup> This refutation applies equally to an argument against physicalism that can be mounted on the basis of Jackson's A Priori Entailment Thesis (Jackson 1993). The A Priori Entailment Thesis plays the same role in such an argument as premise 2 does in the original Zombie Argument by linking negative conceivability and possibility.

premise 2, understood this way, the *CP<sup>neg</sup> Principle*, or *CP<sup>neg</sup>* for short.<sup>35</sup> The upshot is, the Zombie Refutation undermines the *Zombie<sup>neg</sup> Argument*,<sup>36</sup> and in particular, it undermines *CP<sup>neg</sup>*. As we will see, it doesn't undermine either the *Zombie<sup>pos</sup> Argument*, or *CP<sup>pos</sup>*.

Whereas Chalmers thinks that the notion of negative conceivability is straightforward, he admits that positive conceivability involves something murkier. Chalmers 2002a characterizes it in terms of “modal imagination”, where

to positively conceive of a situation is to in some sense imagine a specific configuration of objects and properties. It is common to imagine situations in considerable detail, and this imagination is often accompanied by interpretation and reasoning. When one imagines a situation and reasons about it, the object of one's imagination is often revealed as a situation in which *S* is the case, for some *S*. When this is the case, we can say that the imagined situation *verifies S*, and that one has *imagined that S*. Overall, we can say that *S* is positively conceivable when one can imagine that *S*: that is, when one can imagine a situation that verifies *S*. (p. 148)<sup>37</sup>

Chalmers (2009) also says that

Where negative conceivability requires merely entertaining a hypothesis and being unable to rule it out, positive conceivability involves being able to form some sort of clear and distinct conception of a situation in which the hypothesis is true. (p. 150)

I am not going to try to give a systematic reading of Chalmers' idea here; rather, I will suggest a sense in which zombies are conceivable that is different from mere negative conceivability. On my distinction, there will be cases where a statement is negatively, but not – in the sense I am suggesting – positively conceivable.<sup>38</sup> I think it might be possible to link this to

---

<sup>35</sup> This name corresponds to Chalmers' terminology (Chalmers 2007).

<sup>36</sup> I will distinguish between versions of the Zombie Argument depending on whether they appeal to a notion of negative vs. positive conceivability; i.e., I will refer to them as “Zombie<sup>neg</sup> Argument” vs. “Zombie<sup>pos</sup> argument”.

<sup>37</sup> This is not intended as a definition. As Chalmers puts it, ““Modal imagination” [i.e., the mental act that underlies positive conceivability] is used here as a label for a certain sort of familiar mental act, and like other such categories, it resists straightforward definition.” (p. 150)

<sup>38</sup> Chalmers himself seems to think that exactly the same statements are positively conceivable as are negatively conceivable, though he adds that in case this is not so, it is positive conceivability

Chalmers' notion of positive conceivability; but I don't want to rest my case on the latter claim. In any case, the notion of conceivability I am suggesting stands up independently of Chalmers' framework and forms the basis of a conceivability argument that is more plausible than the Zombie<sup>neg</sup> Argument.

"Positive conceivability", as I understand it, is more restrictive than negative conceivability. For a situation to be positively conceivable by a person it has to be negatively conceivable and also that person has to have what I previously called a "grasp" of all the properties involved in the imagined situation.<sup>39</sup> The reason that yogis can conceive of physically identical duplicates absent flurges is that their concept FLURG refers directly to flurges. Similarly, zombies are negatively conceivable (that there is no *a priori* contradiction in a physicalist duplicate not being mental duplicate) because phenomenal concepts refer to their referent directly without the mediation of physical or functional modes of presentation. Without such modes of presentation there are no conceptual hooks to connect phenomenal with physical concepts. It is this feature that is parlayed by the Zombie<sup>neg</sup> Argument to the conclusion that the referent is not physical. This directness is shared by the zombies' phenomenal\* concepts and the yogis' concept FLURG. That is what enables the Zombie Refutation and the Yogi Argument. The positive conceivability of zombies, on the other hand, depends *also* on the *grasp* that phenomenal concepts provide of their referents. The formulation of the Zombie Argument that appeals to positive conceivability – the Zombie<sup>pos</sup> Argument – is immune to the Zombie Refutation.

Here is why. Zombies in conceivable zombie-worlds do not have a – clear and distinct – grasp of phenomenal\* states and so they cannot conceive<sup>pos</sup> of zombies\*, i.e., creatures that are physically identical to them but lack phenomenal\* states. For example, PAIN\* – like other zombie phenomenal\* concepts – refers to a brain state of zombies, but zombies do not *grasp* this state via their concept PAIN\*. If this is so, it follows that premise 1<sup>pos\*</sup> of the zombies' Zombie<sup>pos\*</sup> Argument,

(1<sup>pos</sup>)  $P \& \sim Q^*$  is conceivable<sup>pos</sup>,

(unlike (1<sup>neg</sup>) of the zombies' Zombie<sup>neg\*</sup> Argument) is false, hence the Zombie Refutation doesn't work against the Zombie<sup>pos</sup> Argument.<sup>40</sup>

---

we have to go for.

<sup>39</sup> I think it is plausible that Chalmers would agree that a grasp of the properties involved is *necessary* – but not that it is *sufficient* – for positive conceivability. So what I call positive conceivability might not be not quite as restrictive as his notion of positive conceivability. But this issue will not play any role until later; I will take up the difference between our notions then.

<sup>40</sup> Many of the dualist conceivability arguments appeal to the fact that phenomenal concepts

As we will see this is not as bad as it sounds for the physicalist camp. There is a physicalist master argument that can answer all the conceivability arguments.

## V. The Counter Conceivability Argument

The conceivability arguments are based on a premise concerning the epistemic status of phenomenal experience with respect to the physical (e.g., that zombies are *prima facie* possible (Kripke 1972); that zombies are conceivable (Chalmers 1996, 2009); that there is an explanatory gap (Levine 2001); that our grasp of phenomenal properties is unconnected to our grasp of physical properties (Nida-Rümelin 2007); or that phenomenal and physical concepts are independent yet connote essential properties of the referent (White 2007)), and another premise linking this epistemic status to metaphysical status. Both kinds of premise, if true, are a priori true.<sup>41</sup> The conceivability arguments all conclude that physicalism is false. This conclusion is not a priori knowable, even according to dualists; the conclusion depends on the – plausibly a posteriori – assumption that phenomenal experience exists.<sup>42</sup> However, I claim that each of the dualist conceivability arguments is only as good another one that wears the a priori of its premises on its sleeves. One can make some modifications in the formulation of the arguments – none that the dualist couldn't accept – to make explicit the a priori elements of the arguments. It will follow that if these arguments are successful then what I call “illuminati” are *inconceivable*.

---

afford us a grasp of their referent even as they lack any descriptive mode of presentation. For example, this notion plays an important role in Martine Nida-Rümelin's (2007) Grasping Phenomenal Properties Argument, in Levine's Explanatory Gap Argument (Levine 2001, ch. 3 formulates, though doesn't endorse, this argument), and White's Property Dualism Argument (2007). Consequently a counterargument on the analogue of the Zombie Refutation would not be effective against them.

<sup>41</sup> One might question the a priori status of either of these premises; however, the proponents of these arguments do all seem to regard their premises as a priori so I am not going to pursue this complication here.

<sup>42</sup> Perhaps one can argue to the contrary, claiming that there is a sense in which we can know a priori in the first person that we have phenomenal experience - but it is not quite the same sense as we know a priori, e.g., that zombies are conceivable. The latter is a matter of understanding the concepts involved; whereas the former depends on the implicit understanding that entertaining a first person thought of an experience as it is occurring involves the experience itself. I am not going to dwell on these issues since I can sidestep them by modifying the conceivability arguments slightly.

Illuminati are purely physical creatures that are our physical duplicates *and* enjoy phenomenal experiences.<sup>43</sup> I will use this conclusion to structure the debate between the anti-physicalist and the physicalist: the anti-physicalist denies, and the physicalist asserts the conceivability of illuminati.

The Counter Conceivability Argument is an argument that presents the case for the *conceivability* of illuminati;<sup>44</sup> it is a master argument that provides a rebuttal of *all* the conceivability arguments via an argument for the conceivability of illuminati. This argument has the advantage that it doesn't rely on assumptions about the mental states of conceivable zombies like my earlier argument does; and so it can be accepted by even those who oppose my earlier Zombie Refutation.

In what follows, I consider the particular case of the Zombie<sup>pos</sup> Argument.<sup>45</sup>

*The Zombie<sup>pos</sup> Argument:*

1<sup>pos</sup>)  $P \& \sim Q$  is conceivable<sup>pos</sup>.<sup>46</sup>

2<sup>pos</sup>) If  $P \& \sim Q$  is conceivable<sup>pos</sup> then  $P \& \sim Q$  is metaphysically possible (CP<sup>pos</sup> principle)

3<sup>pos</sup>) If  $P \& \sim Q$  is metaphysically possible then physicalism is false.

4<sup>pos</sup>) Physicalism is false.

---

<sup>43</sup> Note that “purely physical” is not meant to exclude the presence of mental properties; only the presence of *fundamental* mental properties; so illuminati are not *trivially* inconceivable.

<sup>44</sup> Physicalists, of course, are committed to the *actual truth* of illuminati. But the argument will only require their conceivability.

<sup>45</sup> It can be shown that if the argumentative strategy works against the Zombie<sup>pos</sup> Argument, it works against the other conceivability arguments as well – but I will not spell out the details. The basic idea is that these arguments can all be formulated to employ supposedly a priori premises for the (supposedly a priori) lemma that if phenomenal experience exists then physicalism is false.

<sup>46</sup>  $P$ , again, is the full physical description of the world, including the fundamental physical laws, and  $Q$  is a positive phenomenal truth.

I now construct the *Zombie<sup>pos</sup><sub>a priori</sub>* Argument whose conclusion is that phenomenal experience in a purely physical world is inconceivable. I argue that if the premises of the *Zombie<sup>pos</sup>* Argument are true then the premises of the *Zombie<sup>pos</sup><sub>a priori</sub>* Argument are true. Along the way, I make some of the implicit assumptions of the *Zombie<sup>pos</sup>* Argument explicit. The argument presented involves some technical detail; readers can skip to the next section without losing the main drift of my argument in the rest of the paper. I will use the conclusion of this argument – that illuminati are inconceivable – to clarify the dialectic a hand; but it is mainly an expository device. We could stick to the original form of the *Zombie* argument and make the same points.

### *The Zombie<sup>pos</sup><sub>a priori</sub> Argument*

Our first premise is derived from  $1^{\text{pos}}$  of the *Zombie<sup>pos</sup>* Argument:

$1^{\text{pos}}_{a\text{ priori}}$ ) It is a priori that  $\forall P (P \& \sim Q \text{ is conceivable}^{\text{pos}})$ .<sup>47</sup>

$1^{\text{pos}}_{a\text{ priori}}$  involves a generalization over  $1^{\text{pos}}$ ; this generalization is – as we will see shortly – necessary to construct the argument for the inconceivability of physicalism we are aiming for.  $P \& \sim Q$  is conceivable<sup>pos</sup> for *any physical statement* – and not just the full physical description of our world – and for the same reason, having to do with the conceptual independence of phenomenal and physical descriptions in general. Because of the connection between a priority and conceivability, this means that for any  $P$ , it is a priori that  $P \& \sim Q$  is conceivable<sup>pos</sup>. However, though it is a further claim, I think it is also knowable a priori that for any  $P$ ,  $P \& \sim Q$  is conceivable<sup>pos</sup>, and this is because the general reason for the truth of individual conceivability claims – the conceptual independence of phenomenal and physical descriptions – is knowable a priori.

The next premise is a generalization of  $2^{\text{pos}}$ :

$2^{\text{pos}}_{a\text{ priori}}$ ) It is a priori that  $\forall P (\text{if } P \& \sim Q \text{ is conceivable}^{\text{pos}} \text{ then } P \& \sim Q \text{ is metaphysically possible})$ .

$(2^{\text{pos}}_{a\text{ priori}})$  is an extension of  $(2^{\text{pos}})$  for all physical statements  $P$ ,<sup>48</sup> making explicit that  $(2^{\text{pos}})$ , if true, is true a priori.

---

<sup>47</sup>  $\forall$  is a substitutional quantifier,  $P$  is a statement variable for (logically consistent) physical statements, and  $Q$  is a phenomenal statement.

<sup>48</sup> The general formulation – that  $\forall P (\text{if } P \& \sim Q \text{ is conceivable}^{\text{pos}} \text{ then } P \& \sim Q \text{ is metaphysically possible})$  – follows from Chalmers' Master Principle<sup>pos</sup>, i.e., that conceivability<sup>pos</sup> implies primary possibility, in the same way as  $(2^{\text{pos}})$  – which is a particular application of it – does.

$3^{\text{pos}}$  claims that if  $P \& \sim Q$  is metaphysically possible then physicalism is false. This cannot be known a priori for the simple reason that the premise is only – non-trivially – true if  $Q$  is true and  $P$  is the full, true physical description of the world – which is not knowable a priori. For the purposes of the a priori version of the argument I need to accommodate these assumptions.

$3^{\text{pos}}_{\text{apriori}}$  It is a priori that if  $Q$  is true and  $\forall P$  ( $P \& \sim Q$  is metaphysically possible) then physicalism is false.

$3^{\text{pos}}_{\text{apriori}}$  follows from an a priori consequence of Phys, i.e., that if physicalism is true then  $\exists P \forall T \Box (P \supset T)$ .<sup>49</sup> It is a priori that if  $Q$  is true and  $\forall P$  ( $P \& \sim Q$  is metaphysically possible) then there is no physical description  $P$  such that it necessitates  $Q$ , hence there is no physical description  $P$  that necessitates all truths hence physicalism is false.

From ( $1^{\text{pos}}_{\text{apriori}}$ ) and ( $2^{\text{pos}}_{\text{apriori}}$ ) it follows – by plausible principles of the logic of a priority – that it is a priori that  $\forall P$  ( $P \& \sim Q$ ) is metaphysically possible, and that, together with ( $3^{\text{pos}}_{\text{apriori}}$ ) implies:

LEMMA: It is a priori that if  $Q$  is true then physicalism is false.

Here is the argument again:

### *The Zombie<sup>pos</sup> a priori Argument*

$1^{\text{pos}}_{\text{a priori}}$  It is a priori that  $\forall P$  ( $P \& \sim Q$  is conceivable<sup>pos</sup>).

$2^{\text{pos}}_{\text{apriori}}$  It is a priori that  $\forall P$  (if  $P \& \sim Q$  is conceivable<sup>pos</sup> then  $P \& \sim Q$  is metaphysically possible).

$3^{\text{pos}}_{\text{apriori}}$  It is a priori that if  $Q$  is true and  $\forall P$  ( $P \& \sim Q$  is metaphysically possible) then physicalism is false.

LEMMA: It is a priori that if  $Q$  is true then physicalism is false.

## **VI. The conceivability of illuminati**

---

<sup>49</sup>  $\forall$  and  $\exists$  is a substitutional quantifier,  $P$  is a statement variable for physical statements,  $T$  is a statement variable for true positive statements, and  $\Box$  is the metaphysical necessity operator.

What does LEMMA of the Zombie<sup>pos</sup><sub>a priori</sub> Argument say? Because of the connection between a priority and conceivability, LEMMA is tantamount to the claim that a purely physical world where phenomenal experience occurs is *inconceivable*:

LEMMA: It is inconceivable that  $Q$  is the case and physicalism is true.<sup>50</sup>

In other words, the conclusion of the a priori version of the Zombie Argument is that illuminati are inconceivable.<sup>51</sup> It turns out then that it is possible to respond to the anti-physicalist arguments by giving reasons for the conceivability of illuminati. ( $1^{\text{pos}}_{\text{a priori}}$ ) and ( $3^{\text{pos}}_{\text{a priori}}$ ) are quite clearly true, so if the physicalist could give reasons for the conceivability of illuminati it would amount to giving reasons for the falsity of ( $2^{\text{pos}}_{\text{a priori}}$ ). And because if ( $2^{\text{pos}}$ ) is true ( $2^{\text{pos}}_{\text{a priori}}$ ) is true as well, an argument for the falsity of ( $2^{\text{pos}}_{\text{a priori}}$ ) would amount to an argument for the falsity of ( $2^{\text{pos}}$ ). In other words, by arguing for the conceivability of illuminati, one can show that the crucial premise of the Zombie<sup>pos</sup> Argument is unsound.

What notion of conceivability is involved in LEMMA? Clearly, it is *negative* conceivability:  $Q$  is the case and physicalism is true is supposed to be ruled out on a priori grounds which means that illuminati are supposed to be inconceivable in the negative sense.<sup>52</sup> So to argue against LEMMA, the physicalist only needs to argue for the conceivability<sup>neg</sup> – as opposed to the conceivability<sup>pos</sup> – of illuminati.<sup>53</sup>

---

<sup>50</sup> Marton (1998) expresses a similar idea to the effect that the Zombie argument can be understood to imply the inconceivability of the supervenience of the phenomenal on the physical.

<sup>51</sup> Illuminati, in this general sense, are purely physical creatures that enjoy phenomenal experiences.

<sup>52</sup> This is true in spite of the fact that some of the supposedly a priori truth involved in ruling out  $Q$  and physicalism is true themselves employ the notion of *positive*, rather than *negative* conceivability.

<sup>53</sup> The physicalist might want to argue for the *positive* conceivability of illuminati as well; but this is not strictly necessary for the success of the Counter Conceivability Argument. On my construal of positive conceivability, it seems clear that if illuminati are conceivable<sup>neg</sup> they are conceivable<sup>pos</sup> as well. But my argument for the conceivability of illuminati doesn't have to turn on my particular understanding of conceivability<sup>pos</sup>. Conceivability<sup>pos</sup> might be construed in a way that makes it contentious whether illuminati are conceivable<sup>pos</sup>. The point here is that to counter the conceivability arguments the physicalist only needs to show that illuminati are conceivable<sup>neg</sup>.

But how can the physicalist argue for the conceivability<sup>neg</sup> of illuminati?<sup>54</sup> It will be instructive to compare what is involved in the conceivability<sup>neg</sup> of zombies as opposed to the conceivability<sup>neg</sup> of illuminati. One can see that zombies are conceivable<sup>neg</sup> immediately, without having to consider anything else, e.g., facts about how the brain – or anything else in the physical world – works. Phenomenal concepts refer to phenomenal experiences in a substantial<sup>55</sup> yet direct way, i.e., they do not have analytically sufficient conditions in physical, functional, or behavioral terms; this results in enough conceptual independence of phenomenal and physical concepts to make zombies conceivable<sup>neg</sup>.

How about the conceivability<sup>neg</sup> of illuminati? The basic idea is this. Zombies are conceivable<sup>neg</sup> because we grasp phenomenal properties<sup>56</sup> in a substantial yet *direct* way; but illuminati are also prima facie conceivable<sup>neg</sup>, and for the *very same reasons*. There is no conceptual contradiction in the claim that every instantiated phenomenal property is *identical* with some physical or (physically realized) functional property or other. Such identifications do not seem to be incoherent. Perhaps even stars or rocks are conceivably<sup>neg</sup> conscious in virtue of their physical properties. One might object to this on the grounds that only entities with some minimal cognitive organization can be conceived to have phenomenal states. While no functional or physical description is analytically sufficient for a state to be phenomenal perhaps there are some analytically necessary conditions for phenomenality, requiring that subjects have at least a rudimentary cognitive structure. If these analytically necessary conditions exist, they plausibly rule out stars and rocks as bearers of phenomenal experiences. Whether or not this is right there is certainly no conceptual contradiction in identifying qualia with physical states playing whatever functional roles are necessary to satisfy those conditions.

---

<sup>54</sup> It is worth pointing out that – due to the general nature of LEMMA – to rebut the Zombie<sup>pos</sup> Argument, the physicalist merely has to argue for the conceivability<sup>neg</sup> illuminati in *some* physical world – and not for the conceivability<sup>neg</sup> illuminati in a world *physically exactly like our world*. Doing the former might be less demanding than doing the latter.

<sup>55</sup> Levine (2007) and Chalmers (2007) both talk about the “substantiality” of our conception of phenomenal states and of our phenomenal knowledge. Later I will explain in more detail what I mean by “substantiality”. The main idea is that we grasp phenomenal states in a way that seems to reveal their essence.

<sup>56</sup> At least in the phenomenal way of grasping them; if physicalism is right, then, of course we can grasp them in a physical way as well. This doesn’t affect the first point, though.

Strictly speaking, it is not true that conceiving<sup>neg</sup> of a world where such property identities hold is thereby conceiving<sup>neg</sup> of a world where physicalism is true. Perhaps there are metaphysically possible worlds where physicalism is false for reasons having nothing to do with phenomenal properties (e.g. because of the instantiation of some non-phenomenal non-physical property). However, I don't see any reason why, in conceiving of a world where psycho-physical identities hold, one cannot just stipulate that there are no further non-physical, or non-physically realized properties. In any case, the conceivability arguments purport to rule out physicalism by ruling it out on a priori grounds that physical facts necessitate phenomenal facts, so arguing that such necessitation is conceivable<sup>neg</sup> answers those arguments.

So far so good. But there seems to be a problem. Even if – as I argued – illuminati cannot be ruled out on conceptual grounds, can't they be ruled out via the putatively a priori CP<sup>pos</sup> Principle? Wouldn't it be question-begging for the physicalist to rest her case about the conceivability<sup>neg</sup> of illuminati merely on the basis of their conceptual coherence? The physicalist clearly needs to do more. She needs to respond in some stronger fashion to the CP<sup>pos</sup> Principle. That is what my earlier Zombie Refutation did with regard to the CP<sup>neg</sup> Principle by showing how, on some plausible assumptions, it leads to a contradiction. Showing that the CP<sup>pos</sup> Principle leads to contradiction would be the most effective response to the dualist. Let me digress briefly to consider it.

The physicalist might try to argue that illuminati are not only conceivable<sup>neg</sup>, but also conceivable<sup>pos</sup>.<sup>57</sup> In effect, there is reason to think that the conceivability<sup>pos</sup> of zombies and the conceivability<sup>pos</sup> of illuminati are on a par. Both are equally prima facie conceivable<sup>pos</sup>, due precisely to the direct and substantial grasp of phenomenal properties that phenomenal concepts afford us. Moreover, and this is the key, the dualist *cannot use* the CP<sup>pos</sup> Principle against the conceivability<sup>pos</sup> of illuminati (in the way she might try to use it against the conceivability<sup>neg</sup> of illuminati). In particular, when the dualist argues that the conceivability<sup>pos</sup> of illuminati can be ruled out a priori on the basis of the conceivability<sup>pos</sup> of zombies and the CP<sup>pos</sup> Principle, the physicalist can counter that this is just special pleading. There is an argument of the *exact same form*, using the conceivability<sup>pos</sup> of illuminati and the appropriate analogue of the CP<sup>pos</sup> Principle to the effect that if illuminati are conceivable<sup>pos</sup> then they are possible, that leads to the conclusion that *zombies* are not really conceivable<sup>pos</sup>. This is an absurd conclusion; so the point is not to take the argument seriously as a positive argument. Rather, if this works, it is a *reductio*

---

<sup>57</sup> Frankish (2007) attacks the Zombie Argument on the grounds that “anti-zombies” are conceivable. Though he doesn't discuss the distinction between positive and negative conceivability, I think his claim that “anti-zombies” are conceivable is best interpreted as the conceivability<sup>pos</sup> of illuminati. Sturgeon (2000) also gives an argument similar in form to show that conceivability cuts both ways and so that the Zombie Argument doesn't have force. These both differ from my Counter Conceivability Argument.

of Chalmers' Master Principle connecting conceivability and modality.

Though this is an interesting argument I am not going to pursue its details here.<sup>58</sup> For one thing, it applies only to the Zombie<sup>pos</sup> Argument and not to the other anti-physicalist arguments. Furthermore, the anti-physicalist might question the account of positive conceivability it rests on; she might argue that illuminati are not positively conceivable after all. For example, Chalmers' notion of positive conceivability might require more than a mere grasp of the properties involved in the imagined situation.<sup>59</sup>

What if a knock-down argument is not possible with regard to the CP<sup>pos</sup> Principle? How is the physicalist going to deal with the Zombie<sup>pos</sup> Argument? In particular, how does the focus on the conceivability<sup>neg</sup> of illuminati help? So far what we have seen is that the physicalist can show that illuminati are conceivable<sup>neg</sup> *modulo* the CP<sup>pos</sup> Principle. But the CP<sup>pos</sup> Principle still needs to be answered. Here is the idea. The physicalist might not be able to outright refute the CP<sup>pos</sup> Principle. She might, however, be able to *extend* her argument for the conceivability of illuminati by not only considering the conceptual coherence of psycho-physical identities, but considering it in the context of a holistic account of laws, modality, mental causation, concepts in general and phenomenal concepts in particular, all of which fits together to provide – among other things – a physicalist account of the epistemic gaps between phenomenal and physical descriptions from which the anti-physicalist arguments take off.<sup>60</sup> If such an account is possible it would count as an argument for the conceivability<sup>neg</sup> of illuminati and so indirectly against the CP<sup>pos</sup> Principle.

A look at the CP<sup>pos</sup> Principle itself supports this diagnosis. Where does the supposed a priori status of the CP<sup>pos</sup> Principle come from? It certainly doesn't seem to be a conceptual truth. Besides citing metaphysical intuition, the dualist has other options as well. Chalmers (2002), e.g.

---

<sup>58</sup> An argument with a similar structure is suggested in Loewer (1978) with respect to Leibniz's Ontological Argument.

<sup>59</sup> In addition to the requirement concerning the grasp of relevant properties, Chalmers' distinction between negative and positive conceivability seems to have another aspect that is reminiscent of the distinction between a non-constructive proof and a constructive proof. In the former one merely shows that a certain kind of mathematical entity (e.g. transcendental numbers) is consistent while in the latter one constructs the entity and thereby shows its existence. Whether or not, for example, the statement *Space is 11 dimensional* is positively conceivable depends on whether we can provide a constructive account in terms of the concepts *space*, *dimensionality*, and so on of how it can be that space is 11 dimensional (such an account has been given by physicists). I think illuminati might not be conceivable<sup>pos</sup> in this sense even if physicalism is true.

<sup>60</sup> Hill and McLaughlin (1999) makes a similar proposal in the context of an argument for physicalism.

argues that the CP<sup>pos</sup> Principle provides the simplest and explanatorily most satisfying account of modality. It also best explains the puzzling epistemic gaps the conceivability arguments exploit. Even if it did – which the physicalist surely questions – this makes the CP<sup>pos</sup> Principle importantly different from stock examples of a priori truths, e.g.,  $2+2=4$ , or *Bachelors are unmarried*, whose falsity can be ruled out by a priori methods independent of holistic considerations about how well one’s overall metaphysical framework accounts for modality, mental causation, laws, meaning, etc. The a priori status of the CP<sup>pos</sup> Principle, in contrast, depends on these same holistic considerations.<sup>61</sup> Consequently the Zombie<sup>pos</sup> Argument, based on the CP<sup>pos</sup> Principle, cannot demand assent from all parties to the debate – like valid arguments based on premises like  $2+2=4$ , or *Bachelors are unmarried* can – independent of their metaphysical outlook.<sup>62</sup> There is a physicalist argument to be made on these general grounds against the apriority of the CP<sup>pos</sup> Principle, and for the conceivability of illuminati.<sup>63</sup> In the rest of the paper I will follow out this dialectic in broad strokes.

## VII. The Phenomenal Concept Strategy

As part of an overall physicalist strategy to answer the anti-physicalist arguments, I will now invoke an approach to the mind-body problem which has become known as the “Phenomenal Concept Strategy”. The idea is to think about our epistemic/conceptual relation to consciousness – the conceivability<sup>pos</sup> of zombies, the explanatory gap, our substantial grasp of phenomenal experiences, etc. – in terms of the peculiar nature of *phenomenal concepts*, rather than in terms of the peculiar nature of phenomenal experience itself. The key factor is that the explanation on offer is *compatible with physicalism*.

### *The core idea*

To explain the core idea I will consider “bright illuminati”, i.e., illuminati that have the same epistemic relation to their phenomenal states as we do. Conceiving<sup>neg</sup> of bright illuminati requires – among other things – that we have a clear and distinct physicalistic explanation of phenomenal states being conceived of and known by illuminati in a direct and substantial way –

---

<sup>61</sup> I am not going to decide whether the terminology *a priori* properly describes such considerations. Nothing rides on the terminology.

<sup>62</sup> I am indebted to Troy cross for discussion on this issue.

<sup>63</sup> An argument of this sort for the *conceivability* of illuminati is at the same time an argument for the *actual truth* of physicalism. Such a holistic argument for the coherence of physicalism – which removes the main reason against it – counts as reason to believe in the actual truth of physicalism.

an explanation that is conceptually coherent. I think there is such a conceivable (and probably true) explanatory account: the constitutional account.<sup>64</sup> On this account,<sup>65</sup> there is an intimate relation between phenomenal concepts and their referents; token experiences serve as modes of presentation of the phenomenal properties they instantiate.<sup>66</sup> In the case of most concepts, e.g., the concept WATER it doesn't matter exactly what neural configurations constitute a particular token of WATER as long as the requisite causal/informational relations between it and water hold. But in the case of phenomenal concepts, e.g., the concept PAIN, constitution matters for reference, both in terms of how the reference is determined, and in terms of how the concept cognitively presents its reference. More precisely, on this view, every token of a phenomenal concept applied to current experience is constituted by *that token experience*, and this fact is crucial in determining the reference of the concept. Not only is it the case that a token experience that realizes a token concept instantiates a token of the referent of the concept, but it is *because* the concept is so constituted that it so refers. There are, of course, applications of phenomenal concepts that are, on this theory, *not* constituted by token experiences; e.g., applications of phenomenal concepts to one's past or future experience, to other peoples' experiences, etc. But the canonical, first person, present tense applications are always so constituted and the other applications are dependent on the first person applications.<sup>67</sup>

This account explains the puzzling aspects of our epistemic/conceptual relation to our conscious states in a manner consistent with physicalism. First of all, the constitutional account explains how we can have a *substantial grasp* of phenomenal properties even while this grasp is

---

<sup>64</sup> I defend a version of that account in Balog 2012a. Other versions of the constitutional account have been proposed, on the physicalist side, by Hill and McLaughlin (1999), Block (2007), and Papineau (2002, 2007); David Chalmers (2003) also put forward a version of this account on the dualist side.

<sup>65</sup> There are other proposals that fall under the general umbrella of the Phenomenal Concept Strategy. Not all of these proposals acknowledge the substantial manner in which we are aware of our own phenomenal states. They include recognitional accounts (Tye 2003), demonstrative accounts (Levine 2007, Perry 2001), and information-theoretical accounts (Aydede & Güzeldere 2005).

<sup>66</sup> I consider experiences as mental states that have their phenomenal properties essentially, but instantiate other (physical, functional) properties. Also, in what follows I concentrate on phenomenal concepts that refer to phenomenal *properties*; but the account can be easily modified to apply to concepts that refer to *particular instances* of phenomenal concepts.

<sup>67</sup> Such "indirect" applications of phenomenal concepts stand in an intricate conceptual relation with the "direct", first person present tense applications. For an account, see (Balog 2012a).

direct, and unmediated by physical or functional modes of presentation. Because, on this account, in the canonical, first person, present tense applications of a phenomenal concept an instance of the property it refers to is *literally* (physically) present in the concept, there will be always something it is like to token the concept in those canonical applications.<sup>68</sup> Undergoing a token of the phenomenal property reveals something essential about that property, namely, it reveals *what it is like to have it*. This means that phenomenal concepts provide a substantial grasp of the phenomenal properties they refer to. And because, according to the theory, tokens of phenomenal concepts present their referent as the property whose token they incorporate – and not via any functional or physical description – they will refer to phenomenal properties *directly*, as well as substantively. The constitutional account can now be marshaled to account for the epistemic/conceptual gaps that drive the conceivability arguments.<sup>69</sup>

*The conceivability<sup>pos</sup> of zombies* is explained by by the *directness* and *substantiality* of our direct phenomenal concepts which, under the constitutional account, is compatible with physicalism. The directness of phenomenal concepts follows from the fact that the reference of a direct phenomenal concept is determined by how it is constituted and *not* by any description that is associated *a priori* with the concept. Phenomenal concepts are supposed to be different in this way from concepts like WATER and even name concepts like CICERO. Chalmers and Jackson (2001) claim that these concepts are associated *a priori* with descriptions (e.g. “the transparent potable liquid...”, “the Roman orator who is at the origin of a causal chain culminating in this token”) and these connections are *sufficient* to obtain *a priori* entailments from the full fundamental description to all positive statements of fact, e.g., they are sufficient to rule out a priori a scenario where everything is physically the same but yet there is no water. In other words, a situation where everything is physically the same but yet there is no water is inconceivable. One doesn’t have to commit to this to see that zombies are conceivable; however, the conceivability of zombies can only have significance if this is the case.<sup>70</sup> I am going along with the idea of a priori entailment – except, of course, for phenomenal statements – for the sake of argument; Chalmers’ argument can be rebutted in a way that doesn’t depend on rejecting its semantic framework wholesale.<sup>71</sup> The point is that even if one accepts that all true non-

---

<sup>68</sup> Levine (2006, 2007) is critical of this approach. He argues that it is impossible to explain *cognitive presence* by *physical presence*.

<sup>69</sup> For a detailed account of how this works see Balog 2012a.

<sup>70</sup> This is because the conceivability of zombies only has significance if one also accepts CP<sup>pos</sup>. For an argument for the equivalence of the *a priori entailment thesis* and the Master Principle that lies behind the CP Principle, see my doctoral thesis (Balog 1998), p 124-5.

<sup>71</sup> But see, e.g., Block and Stalnaker (1999) and McLaughlin (2007) for an argument that these

phenomenal statements are a priori entailed by the full fundamental description of the world, one can still claim that it is not so with respect to phenomenal concepts. Because of the fundamentally different cognitive architecture of phenomenal concepts, there are no a priori connections between phenomenal and physical/functional/structural concepts that are sufficient to rule out a priori the zombie scenario – and this is perfectly compatible with physicalism. *Nota bene*: I am not denying that there are inferential links between thoughts involving direct phenomenal concepts that are individuating of them. I think it is quite plausible that there are conceptual links, even perhaps concept individuating conceptual links between direct phenomenal concepts such as we apply our own occurrent phenomenal experience on the one hand, and other direct phenomenal concepts, or even indirect phenomenal concepts such as we apply to other people's phenomenal experiences on the other. My point is that to the extent that these are a priori they do not add up to conceptually sufficient conditions in terms of other mental concepts, functional, or behavioral concepts, etc. in other words, they are not of the sort that enables one to rule out a priori the zombie-scenario.

Because phenomenal concepts are direct, we can conceive<sup>neg</sup> of zombies, and because they are substantial, we can conceive<sup>pos</sup> of zombies. This explanation is perfectly compatible with a physicalist – as well as a dualist – metaphysics and leaves the a priori entailment of truths and CP<sup>pos</sup> in place for all except phenomenal statements.

*The explanatory gap* problem is that no amount of knowledge about the physical facts (brain functioning and so on) is able to explain why a particular brain state/process has a particular feel, e.g., feels giddy. This contrasts with the way the fact that water is composed of H<sub>2</sub>O molecules together with physical and chemical laws explains why water is potable, transparent and so on. Once we have an explanation of why H<sub>2</sub>O behaves in watery ways (and that it is the only substance that does so) we have an explanation of why water is H<sub>2</sub>O. Since we can't explain why a brain state feels giddy in neurophysiological terms, we can't close the physical-phenomenal gap. You can see why this is in the following way. In the case of water and H<sub>2</sub>O, the hypothesis that water=H<sub>2</sub>O is quite natural in the light of all we know about H<sub>2</sub>O and the laws that govern the behavior of H<sub>2</sub>O – indeed, the opposite hypothesis doesn't even make sense. The hypothesis that the processes involving H<sub>2</sub>O molecules are only nomologically correlated to the non-physical and non-chemical processes involving water is non-sensical.<sup>72</sup> On

---

entailments are not a priori even for positive non-phenomenal statements.

<sup>72</sup> (Block and Stalnaker (1999) discuss the possibility of 'ghost water' – a non-physical kind that exists side by side with being composed of hydrogen and oxygen atoms and has all the same causal roles as the latter. Even if that is a coherent possibility, it would be the case that "water" refers to both H<sub>2</sub>O and ghost water and *not* that water refers to ghost water alone. So even in that possibility it wouldn't be the case that H<sub>2</sub>O is merely nomologically connected to water.

the other hand, the hypothesis that a phenomenal state is *identical* with a certain neurophysiological/functional state of the brain is just as compatible with our evidence as the opposing view. The hypothesis – endorsed by certain dualists – that phenomenal states and brain states are merely nomologically correlated makes perfect sense.<sup>73</sup>

The difference is that while in the case of water we do not have any special access to its nature and properties that is not based on physical or functional information,<sup>74</sup> in the case of phenomenality we do. We do seem to have a special insight into the ultimate nature of phenomenal experience; and that nature doesn't seem captured or exhausted by any physical or functional description. As far as we know, that nature might elude any physical understanding. Notice that I stated the problem of the explanatory gap in a way that is independent of whether one subscribes to the semantic thesis discussed above that *all* but phenomenal terms have physical/functional analyses. It is significant that this can be done since it demonstrates that not all of the puzzles of consciousness will go away if we simply deny the semantic framework of the Zombie Argument. However, the constitutional account can explain why the explanatory gap arises, and it does so again in a way that is compatible with physicalism.

The constitutional account explains the gap by appealing to the direct and substantial grasp phenomenal concepts afford of their referent. When I focus on the phenomenal state, I have a “substantive” grasp of its nature. I grasp it in terms of *what it's like* to be in that state. Because this grasp is substantive but at the same time independent of any causal or functional information (unlike in the case of WATER), information about the functioning of the brain simply won't explain *what its like to be in that state*.<sup>75</sup>

---

<sup>73</sup> This hypothesis is endorsed by property dualists who accept the causal closure of physics (e.g., Chalmers 1996). If the physical is causally closed, the phenomenal nomologically supervenes on the physical (see, e.g., Loewer 1995). The most natural hypothesis of how this could be the case is that neural correlates of phenomenal properties exist. On this view, there are fundamental laws that link mental properties to each other and to certain physical/functional properties of physical systems. In contrast, *interactive dualists*, who, like Descartes, deny the causal closure of physics, also deny the existence of fundamental vertical laws connecting phenomenal and physical/functional properties. Such dualists deny that phenomenal properties have neural correlates. If so, physicalism could be proven false on empirical grounds; but this is not a very likely scenario.

<sup>74</sup> Except for water's appearance properties, for example, that its surface looks shiny in a storm, that it presents itself in a particular way to the touch, etc. But I am not going to press this point here.

<sup>75</sup> The constitutional account can also explain other puzzling features of our epistemic relation to phenomenal experience, like the incorrigibility of certain of our phenomenal judgments, the

Of course, for the constitutional account to work, some of its details have to be worked out, in particular, the idea that constitution plays a role in determining reference in the case of phenomenal concepts. Whether this can be fully done still remains to be seen.<sup>76</sup> It might also be that any full fledged constitutional account would run into *empirical* difficulties as an actual theory of *our* phenomenal concepts. I will take that issue on in the next section.

### VIII. The Physicalist Master Argument Revisited

I would like now to consider again the dialectic between the physicalist and dualist proponents of the conceivability arguments. Though the Phenomenal Concept Strategy is a major part of the physicalist answer to the conceivability arguments, it must be granted that the dualist can still counter that, as opposed to zombies, bright illuminati are not conceivable<sup>neg</sup> after all due to the a priority of the CP<sup>pos</sup> Principle and its counterparts. However, the Phenomenal Concept Strategy does help to *undermine* the intuitive strength of the CP<sup>pos</sup> Principle and its counterparts. It demonstrates how those intuitions could conceivably arise in a purely physical world, thereby undercutting the rationale behind these principles. As I discussed earlier, the argument for the conceivability of illuminati is a holistic affair so the physicalist also needs to present an alternative to the CP<sup>pos</sup> Principle, and to dualist metaphysics in general, by presenting an account of fundamental properties, laws, concepts, modality, etc. compatible with physicalism. There is much that still can be said about these issues but I won't attempt this here.<sup>77</sup> Rather, I will assess the dialectic between the physicalist and the anti-physicalist on the assumption that the general metaphysical considerations are not decisive in either direction. If this is the case, then, by her own lights, the physicalist has rebutted the anti-physicalist arguments. From her point of view

---

semantic stability of phenomenal concepts, etc. For a more detailed account see Balog (2012a).

<sup>76</sup> For a proposal see Balog (2012a). Chalmers (2007) argues that the Phenomenal Concept Strategy cannot work for very general a priori reasons. For a reply see Carruthers and Veillet (2007), and Balog (2012b).

<sup>77</sup> On the physicalist side, there are arguments in favor of physicalism that appeal to mental causation and the causal closure of physics (Loewer 1995), (Papineau 1995). Loewer and Papineau argue that the anti-physicalist is forced into adopting one of these implausible positions: epiphenomenalism, causal overdetermination, or denial of the causal closure of physics. Hill and McLaughlin (1999), and McLaughlin (2007) argue more generally that there are powerful reasons to prefer physicalism as an overall explanatory metaphysics over dualism. See also Melnyk (2003) for discussion of a history of successful reduction of higher level properties to lower level ones. On the dualist side, Chalmers, e.g., argues (2002a, 2002b) that the CP Principle – and with it, dualism – is required to give a satisfying account of modality.

the CP<sup>pos</sup> Principle and its counterparts are demonstrably mistaken.

However, a parallel argument can be mounted by the dualist. According to the dualist, for all the physicalist's claims that bright illuminati are conceivable<sup>neg</sup> it is still not *irrational* to hold on to the CP<sup>pos</sup> Principle and its counterparts: all the more so since in their light any physicalist concern about it – including the conceivability<sup>neg</sup> of bright illuminati – can be ruled out a priori. From this point of view, it is the physicalist's claim that bright illuminati are conceivable<sup>neg</sup> that is mistaken.

Let's pause here for a moment. On the one hand, the anti-physicalist argues that the CP<sup>pos</sup> Principle and its counterparts are a priori true, and uses these claims to show that the physicalist efforts to undermine them are unsuccessful. On the other hand, the physicalist argues that bright illuminati are conceivable<sup>neg</sup> – based on the Phenomenal Concept Strategy and holistic considerations about metaphysics – and uses this claim to undermine the principles. By their own lights both sides seem justified to hold their key doctrines and deny the opposing views. Where you end up depends on what you take as your starting point. And, as far as I can see, neither side has a privileged start.

This is puzzling. One would have thought that when it comes to *a priorities*, like the CP<sup>pos</sup> Principle and its counterparts, and the conceivability<sup>neg</sup> of bright illuminati, there are a priori ways to justify or refute them. But it seems like there are no principles *outside* the physicalist and anti-physicalist systems that could settle this issue by showing that either physicalism or property dualism is incoherent. What we have here is a puzzling symmetry between the two positions. The situation appears to be a stalemate.

It is unlikely that this stalemate can be broken by empirical evidence either. We have good reason to think that non-interactionist property dualism and physicalism are equally compatible with all empirical evidence. Of course, for physicalism to be a credible metaphysics, phenomenal properties have to have “neural correlates”. This is the term that became widely used for the physical or functional basis of phenomenal experience in general and of specific phenomenal experiences. The correlation in question is supposed to be (at least) nomological: a “neural correlate” of some phenomenal property is a physical or functional property that is coextensive with a phenomenal property in all creatures both natural and artificial in (at least) all nomologically possible worlds.<sup>78</sup> But because the non-interactionist property dualist – like the physicalist – believes that the physical is causally closed, she is equally committed to the existence of such neural correlates. If the physical is causally closed, the phenomenal nomologically supervenes on the physical.<sup>79</sup> The most natural hypothesis of how this could be

---

<sup>78</sup> If the phenomenal property and the correlated physical property are the same property then the correlation holds in all metaphysically possible worlds but if dualism is true the correlation may hold only in worlds in which there are laws linking the distinct phenomenal and physical properties. The term “correlation” is neutral between metaphysical and nomological connection.

<sup>79</sup> See, e.g., Loewer (1995).

the case is that neural correlates of phenomenal properties exist. So both physicalism and epiphenomenal property dualism is committed to the existence of neural correlates of consciousness. Empirical evidence so far seems to bear this out. Of course, it might turn out that the physical is not causally closed and they are both wrong on this issue. But at present there is no reason to think so.

For physicalism to be a credible metaphysics it has to be the case that the puzzling features of consciousness - the conceivability of zombies, the explanatory gap, and all the other epistemic gaps – like everything else, are *actually* physically explicable. Here is where the constitutional account comes in as an explanation of how phenomenal concepts work. The essence of the Phenomenal Concept Strategy is to explain the epistemic gaps in terms of the structure of phenomenal concepts in a way that is compatible with physicalism. Of course, the proposal that the constitutional account of phenomenal concepts – or some other version of the Phenomenal Concept Strategy – is the true account of phenomenal concepts has empirical consequences; and its verification is contingent on the findings of future psychology and neuroscience. But I believe that we have strong reason to think that some such account *will* be compatible with the empirical evidence. So it seems plausible that the stalemate between physicalism and dualism cannot be broken by future empirical evidence.

In closing, I would like to suggest another way of looking at the debate. Let's suppose for the sake of argument that non-interactive property dualism and physicalism are two metaphysical frameworks that can be developed so as to be both empirically and also philosophically equivalent. By philosophically equivalent I mean that they can each be developed so as to account equally well (or badly) for our philosophical intuitions and that there is no external, that is, non question-begging, philosophical principle that can settle the apparent dispute.

So far I have assumed that there is a matter of fact that the dualist and the physicalist disagree about. There are some other instances of statements which seem to be either true or false but where seem to be neither a priori nor a posteriori ways of deciding the issue. For example, mathematical realists typically think that Cantor's Continuum Hypothesis is an instance. But the case at hand seems different since it doesn't involve the complexity of the mathematical case. Another possibility is that the dispute between the dualist and the physicalist is a case of the underdetermination of theory by evidence of which there are examples in the sciences.<sup>80</sup>

An altogether different response in the face of this kind of stand off is to wonder whether there is a genuine dispute at all. It could be that the relevant notions that figure in explicating these different metaphysical positions (concept, property, law, necessity, etc.) are indeterminate in a way that prevents the two positions to express a real difference. In other words, the different positions on, e.g., concerning the relationship between concepts and properties entailed by the

---

<sup>80</sup> A case in point is Bohmian mechanics versus the GRW theory in quantum-mechanics.

CP<sup>pos</sup> Principle and its denial might encode a difference in the very *conception* of concepts and properties preventing the two positions to express a genuine factual difference. This position might strike one as very counterintuitive; of all metaphysical disputes the difference between physicalism and dualism seems as factual as any; yet non-factualism might be the most satisfactory solution of the metaphysical gridlock discussed in this paper.

## References:

- Aydede, M. & Güzeldere, G., 2005. "Cognitive architecture, concepts, and introspection: An information-theoretic solution to the problem of phenomenal consciousness", *Noûs*, 39 (2): 197-255.
- Balog, K., 1998. *Conceivability and Consciousness*. Dissertation, Rutgers University.
- Balog, K., 1999. "Conceivability, Possibility, and the Mind-Body Problem", *The Philosophical Review*, Vol 108, No 4 (October 1999). Reprinted in Volume XXIII of *The Philosopher's Annual*.
- Balog, K., 2001. "The A Priori Entailment Thesis", Commentary on Frank Jackson's *From Metaphysics to Ethics*, *Philosophy and Phenomenological Research*, Vol. LXII, No. 3.
- Balog, K., (2012a). "Acquaintance and the mind-body problem", forthcoming in a volume on Identity Theory, ed. by Christopher Hill and Simone Gozzano, Cambridge University Press.
- Balog, K. (2012b). "In Defense of the Phenomenal Concept Strategy", *Philosophy and Phenomenological Research*.
- Bealer, G. (1994). "Mental Properties," *Journal of Philosophy* 91:185-208.
- Block, Ned & Stalnaker, Robert, 1999. "Conceptual Analysis, dualism, and the explanatory gap", *Philosophical Review* 108:1-46.
- Block, N., 2007. "Max Black's Objection to Mind-Body Identity", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Bloom, Paul, 2004. *Descartes' Baby*, Basic Books.
- Carruthers, P., Veillet, B, 2007. The Phenomenal Concept Strategy, *Journal of Consciousness Studies* 14.

- Chalmers, D., 1996. *The Conscious Mind*, New York: Oxford University Press.
- Chalmers, D., and Jackson, F., 2001. "Conceptual Analysis and Reductive Explanation", *Phil Review* 110:315-61.
- Chalmers, D., 2002a. "Does Conceivability Entail Possibility?" Published in (T. Gendler & J. Hawthorne, eds) *Conceivability and Possibility*, Oxford University Press, pp.145-200.
- Chalmers, D., 2002b. Consciousness and its place in nature. In (D. Chalmers, ed) *Philosophy of Mind: Classical and Contemporary Readings* (Oxford University Press, 2002). Also in (S. Stich & F., Warfield, eds) *Blackwell Guide to Philosophy of Mind*. Blackwell, 2003. Reprinted (abridged) in( W. Lycan & J. Prinz, eds) *Mind and Cognition: A Reader* (Blackwell, 2007).
- Chalmers, D., 2003. The Content and Epistemology of Phenomenal Belief, in Q. Smith and A. Jokic (eds), *Consciousness: New Philosophical Perspectives*, Oxford University Press.
- Chalmers, D., 2004. "The foundation of two-dimensional semantics", in *Two-Dimensional Semantics: Foundations and Applications*, Oxford University Press.
- Chalmers, D., 2006. "Two-Dimensional Semantics", in *Oxford Handbook to the Philosophy of Language*. (Lepore and Smith, eds.), Oxford University Press.
- Chalmers, D., 2007. "Phenomenal Concepts and the Explanatory Gap." in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Chalmers, D., 2009. "The Two-Dimensional Argument Against Materialism", in B. McLaughlin (ed.) *The Oxford Handbook of the Philosophy of Mind*.
- Davidson, D., 1973. "Radical Interpretation", in *Inquiries into Truth and Interpretation*, Oxford University Press, 1984.
- Dretske, F., 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.
- Fodor, J., 1998. *Concepts: where cognitive science went wrong*, Oxford University Press.
- Fodor, J., 1990. *A Theory of Content and Other Essays*. Cambridge: MIT Press.

- Frankish, K., 2007. "The anti-zombie argument". *Philosophical Quarterly*, 57 (229). pp. 650-666.
- Hill, C. and McLaughlin, B., 1999. There are Fewer Things Than Are Dreamt of in Chalmers's Philosophy, *Philosophy and Phenomenological Research* 59(2), 445-454.
- Jackson, F., 1982. "Epiphenomenal Qualia", *Philosophical Quarterly* 32: 127-36.
- Jackson, F., 1993. Armchair Metaphysics. In *Philosophy in Mind*, ed. M. Michael and John O'Leary-Hawthorne. Dordrecht: Kluwer.
- Jackson, F., 2003. "Mind and illusion". In (A. O'Hear, ed) *Minds and Persons*. Cambridge University Press.
- Jackson, F., 2007. "A Priori Physicalism", in *Contemporary Debates in Philosophy of Mind*, Blackwell Publishing.
- Kaplan, D, 1968-1969. "Quantifying In," *Synthese*, 19.1-2 (Dec.): 178-214.
- Kirk, R. (2005). *Zombies and Consciousness*, Oxford University Press.
- Kriegel, U., 2004. "Consciousness and Self-Consciousness". *Monist*: 182-205.
- Kripke, S., 1972. *Naming and Necessity*. Cambridge: Harvard University Press.
- Levin, J., 2007. "What is a phenomenal concept?", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press
- Levine, J, 2001. *Purple Haze*, Oxford University Press.
- Levine, J. 2006. "Conscious awareness and (self)representation", in *Self-Representational Approaches to Consciousness*, Uriah Kriegel and Kenneth Williford (eds.), MIT/Bradford.
- Levine, J., 2007. "Phenomenal concepts and the materialist constraint", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Lewis, D., 1966. "An argument for the Identity Theory", *Journal of Philosophy*, LXIII. 1 (January 6, 1966): 17-25.

- Lewis, D., 1983a. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61:343-77.
- Lewis, D., 1983b. "Radical Interpretation", in *Philosophical Papers*, vol 1. Oxford University Press.
- Loewer, B., 1995. "An Argument for Strong Supervenience" In (E. Savellos & U. Yalcin, eds.) *Supervenience: New Essays*. Cambridge University Press.
- Loewer, B., 1978. "Leibniz and the Ontological Argument", *Philosophical Studies* 34 (1978) 105-109.
- Marton, Peter, 1998. "Zombies versus Materialists: The Battle for Conceivability", *Southwest Philosophy Review*, **14**: 131-138.
- McLaughlin, B., 2007. "On the limits of A Priori Physicalism", *Contemporary Debates in Philosophy of Mind*, Blackwell Publishing.
- Melnyk, A., 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge University Press, Cambridge.
- Millikan, R., 1989. "Biosemantics." *Journal of Philosophy* 86: 281-97.
- Nagel, T., 1974, "What Is It Like to Be a Bat?", *Philosophical Review* 83:435-50.
- Nida-Rümelin, M., 2007. "Grasping Phenomenal Properties", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Papineau, David. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, D., 1995. "Arguments for Supervenience and Physical Realization." In (E. Savellos & U. Yalcin, eds.) *Supervenience: New Essays*. Cambridge University Press.
- Papineau, D., 2002. *Thinking about consciousness*, Oxford UP.
- Papineau, D., 2007. "Phenomenal and Perceptual Concepts", in T Alter and S Walter (eds)

- Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Perry, J., 2001. *Knowledge, Possibility and Consciousness*. MIT Press.
- Putnam, H., 1982. *Reason, Truth, and History*, Cambridge University Press.
- Rey, G., 2007. "Phenomenal Content and the Richness and Determinacy of Color Experience", *Journal of Consciousness Studies*, 14.
- Robinson, H., 1993. "The Anti-materialist Strategy and the Knowledge Argument", in *Objections to physicalism*, ed. Howard Robinson. Oxford: Oxford University Press.
- Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Sturgeon, S., 2000. *Matters of Mind: Consciousness, Reason and Nature*. Routledge.
- Tye, M., 2003. "A theory of phenomenal concepts", *Philosophy*.
- Soames, S, 2004. *Reference and Description: The Case against Two-Dimensionalism*, Princeton University Press.
- Stoljar, D., 2005. "Physicalism and phenomenal concepts", *Mind and Language*.
- White, S., 2007. "The Argument for the Semantic Premise", in T Alter and S Walter (eds) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and physicalism*, Oxford University Press.
- Yablo, S., 1993. "Is Conceivability a Guide to Possibility?" *Philosophy & Phenomenological Research* 53 pp. 1-42
- Yablo, S., 2002. "Textbook Kripkeanism and the Open Texture of Concepts", *Pacific Philosophical Quarterly*, Vol. 81: 1, pp. 98-122.