1. Introduction

In his recent book, John Perry (2001) observes that the recent dialectic in the debate on the status of phenomenal consciousness is very similar to the one described by Hume about the status of evil and the existence of God. Hume claims that if the existence of God — as traditionally defined as the transcendent creator of the world who is omnipotent, omniscient, and omnibenevolent — were to be entertained as a hypothesis by a neutral and naïve thinker, faced with the actual abundance of evil on Earth, the thinker would conclude that the hypothesis is most probably false. The claim is meant to underscore that it is epistemically more warranted to believe that the hypothesis is false than to believe that it is true, and to adopt this as part of the modus operandi in one’s metaphysical commitments and methodological considerations.

Perry likens Hume’s discussion to what is now happening in the debate on consciousness between materialists and anti-materialists. If no one position has a knockdown argument against the other, the disagreement ought to be settled by evaluating the overall plausibility of each position in the face of all the known empirical evidence along with the theoretical and methodological implications that can be gleaned from the generally accepted theories already in place.

In accordance with Perry’s assessment of this dialectic, we will start with what seems to us the most plausible supposition:

(M) In the debate between materialists and anti-materialists, if no one has any knockdown argument against the other, it is epistemically more warranted to accept materialism as true and reject anti-materialism.
(M) is a defeasible dictum like Occam’s Razor. There are clear-cut theoretical reasons for insisting that materialism must be the correct hypothesis. For it seems to us that the only plausible form of anti-materialism is epiphenomenalism, and epiphenomenalism renders mental states causally inefficacious. This consequence is not only contrary to what appears to be the case, that is, the reality of mental causation and agency, but it also threatens the integrity and coherence of our common sense as well scientific world views at their most fundamental levels. The point is not just that methodological principles such as simplicity and conservatism dictate that we should avoid such radical commitments if we can; rather, giving up the causal efficacy of the mental qua mental seems to go against the grain of a very fundamental conception of ourselves as agents.

For instance, if the ontology of our phenomenology were real but causally idle, this would mean that we have been systematically and radically wrong in our first-person warrants to believe almost anything. Could it be just luck that our warranted beliefs tend to be true? Again, if the phenomenal properties of our experiences are causally inefficacious, how could we ever come to know about them, let alone scientifically investigate them? As Kim (1996: 130) observes, under this conception, “each mental event is a solitary island unto itself”, and that fundamentally there is not much difference between epiphenomenalism and eliminativism, “for a plausible criterion for distinguishing what is real from what is not real is the possession of causal power” (Kim 1998: 119).

These consequences of epiphenomenalism seem to us too big a price to pay.

We know that there are attempted answers designed to absorb the shock of these consequences on the part of epiphenomenalists, and we do not mean to suggest here that these few observations about epiphenomenalism are conclusive. But we do want to highlight the enormous difficulties that its proponents face. Materialists don’t have any of these difficulties. But, are the challenges that they face strong enough to motivate the abandonment of materialism as well?

We think not. But it has been claimed that there are knockdown arguments against materialism. The conceivability arguments, we are told, are such: they purportedly show that materialism is false. Conceivability arguments are not new they go back at least to Descartes — but they have been revived lately by Kripke (1980), Nagel (1974), Jackson (1982, 1986, 1994), Levine (1983, 1993, 2001), Hart (1988), McGinn (1991), and Chalmers (1996), among others, and generated a substantial amount of discussion.

Our primary aim in this article is to lay out, at the most fundamental level, the conceptual terrain of the debate, as we see it, between the anti-materialists who make their case by conceivability arguments and their materialist respondents. We want to explore the dialectics of this debate with an eye towards understanding what
it would take to settle the debate. Indeed we will try to make a case for a possible strategy to resolve the issue. In our opinion, if this strategy doesn’t work, there is not much hope of going beyond what seems to be a standoff between the parties.

Relying on the specifics of two particular versions, we will present the outlines of a generic argument against materialism, which we believe is at the core of all conceivability arguments. We will focus on one of the standard and increasingly more popular materialist replies, generally called perspectivalism. We will point out the strengths and weaknesses of this reply, and discuss why present versions of perspectivalism are less than convincing. We will then explain what would constitute a better perspectivalist strategy — one that has a chance to settle the dispute.

2. Jackson's Knowledge Argument: First Round

(M) is worthless if there are indeed knockdown arguments against materialism as claimed by the defenders of conceivability arguments. We will start with Jackson’s influential “Knowledge Argument”, because this argument brings out very clearly and intuitively what is at issue in the debate. We will use it as a step toward revealing the common denominator of conceivability arguments. Here is the set-up for Jackson's thought-experiment:

Mary is confined to a black-and-white room, is educated through black-and-white books, and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. (1986: 567)

Mary is released and sees for the first time a ripe tomato in good light, and comes to know what it is like to see red, something she allegedly did not know before despite her omniscience about physical facts. Jackson runs his argument thus (1986: 568):


(1)’ Mary (before her release) knows everything physical there is to know about other people.
(2)’ Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release).
Therefore,
(3)’ There are truths about other people (and herself) which escape the physical story.
According to Jackson, materialism is the doctrine that the world consists entirely of physical facts. Hence if someone knows all the physical facts, then one knows all there is to know. According to Jackson, Mary comes to know a new fact, a fact which she did not know before, but since, by hypothesis, she knew all the physical facts, the fact she comes to know upon seeing red for the first time, cannot be a physical fact. Hence, Jackson concludes, there are non-physical facts, and therefore materialism is false.

There have been many materialist responses to this argument in the past. As one of the most compelling and influential materialist responses, perspectivalism acknowledges that Mary indeed comes to learn something new and factual in character, but not a new fact, rather a conceptually (epistemically) new way of relating to an “old” fact which she already knew under its physical conception consequent upon her omniscience with respect to the physical world. Mary already knew what experiencing red is under its scientific description: she knew how red objects strike the retina, how the brain processes the retina’s output in different areas of the visual cortex, etc. So let us say that Mary knew that

(1) experiencing red is SD,

where ‘SD’ stands for the complete scientific description of experiencing red. In this sense, Mary already had the necessary concept(s) and knowledge expressed by ‘SD’. Upon looking at a ripe tomato for the first time after her release, she comes to occupy a certain experiential/brain state for the first time, to which she knows the description ‘SD’ applies. But now, consequent upon her experiencing red for the first time, she also comes to acquire a new concept, which she did not have previously. She is now capable of representing her experience to herself thus:

(2) experiencing red is like this, (or, this is experiencing red)

where ‘this’ expresses the mental tokening of a certain perspectival concept she has just acquired that in turn expresses (the instantiation of) the same property expressed by ‘SD’. It is important to be clear about what is new in Mary when she first experiences red. First, there is the objective property of redness (redness) physical objects possess. Second, there is the visual experience of red (call it, “exp”), which we will treat as non-conceptual (sensory) representation of redness. Finally, there is the phenomenal concept that Mary acquires consequent upon having the experience of red and applies it (call this concept, EXP_{red}) to her experience.

Note that Mary can come to know about her experiences only through the exercise of her concepts applying to them. This is required by Jackson’s argument.
which is about factual/propositional knowledge (knowledge-that as opposed to
knowledge-how). Now the perspectivist reply to Jackson can be stated more
explicitly and clearly. The extensions of Mary’s two concepts, her concept of ex-
periencing red (EXP_red) and her (complex) concept expressed by ‘SD’ are numeri-
cally identical; they denote one and the same property. Assuming that Mary can
project her essentially perspectival concept, EXP_red, to other people’s experiences of
red, we can then respond to Jackson’s second premise in two ways. We can grant it
under one reading that takes the novelty involved not as a novelty in facts, but as a
novelty in representing identical facts. But this is harmless for materialism: Jac-
kson’s conclusion does not follow. Or, we can read (2)’ as claiming that Mary comes
to discover a new fact which she did not represent (let alone, know) in any way be-
fore. But then the premise is false, according to perspectivist materialism.

This materialist strategy should be familiar from one standard way of dealing
with “Frege cases”: one can denote one and the same thing by using different repr,
resentations without knowing that they refer to one and the same thing. Unlike many
concepts in Frege cases, however, EXP_red is supposed to be “perspectival” in a spe-
cial way. Part of what makes it perspectival is this: necessarily, (i) Mary could not
have acquired it had she not had the experience of red; and (ii) EXP_red acquires
(could only have acquired) its extension partly in virtue of standing in a special (di-
rect causal/nomological) informational relation to the experience of red.

We think that this materialist response to the Knowledge Argument is sati-
sfactory from a technical viewpoint: it indeed blocks the argument against materi-
alism. But this is only the first round. The action starts after this round.

3. Conceivability Arguments: Second Round

Jackson claims that Mary learns a new fact upon seeing red for the first time. Facts
can be expressed by statements if we have the relevant concepts. This is common
ground. So, Jackson must be agreeing with the materialist that upon seeing red
Mary comes to have a new concept. But why is he not worried that this new con-
cept may, for all we know (for all Mary knows), denote the same property that the
scientific conception (of experience) denotes? Indeed, this would be the obvious
concern given that Jackson is well aware of the epistemic nature of his argument
and the standard philosophical problems surrounding them. On what basis, then,
does he think that it must be obvious that what the new concept denotes cannot be
entirely physical?

Here things get complicated and obscure, but at least part of the answer is that
it is just intuitively obvious that they don’t denote the same thing. In other words,
when Mary experiences red for the first time, she is confronted with something, call
it the qualitative character of her experience of red (or, red qualia), that is intuitively hard to conceptualize as entirely physical; at least there is nothing in her grasp of this new element that suggests to her that it is physical or, for that matter, functional. When Mary comes to have the experience for the first time, intuitively she comes into direct and unmediated cognitive contact with the qualitative character of her experience, so that her concept formed on this basis directly and immediately picks out that peculiar character of her experience. But this direct cognitive grasp does not present the phenomenal character of her experience as physical or functional, or for that matter, as anything else. Relatedly, our most specific phenomenal concepts seem to apply to particular qualia in a way that do not present them as having internal constituents. In fact, it is this latter fact that seems to generate the intuition that qualia cannot be physical/functional — for how could anything (at this level) that appear to be metaphysically simple be physical? It seems to follow from this that phenomenal concepts are semantically primitive, simple, and unanalyzable.

As will be clearer in a moment, we are prepared to grant all this. Of course, this is not to say that phenomenal concepts do not have conceptual roles, in the sense that they are embedded in a tightly connected belief system, and thus connected to other concepts. To the extent that phenomenal concepts are systematically related to sensory concepts such as RED, SWEET, and LOUD, and are directly acquired from experiences representing the so-called secondary qualities like red, sweet, and loud, they will reflect the interconnections that exist between sensory concepts. Some of these links may be conceptual or analytic: e.g., that red is a color, that red is different from green or sweet. These would yield beliefs such as the belief that an experience of red is an experience of color, that an experience of red is visual. But many of them appear to be contingent: e.g., the belief that the experience of red is typically caused by seeing red things under certain conditions, and is apt to cause other beliefs about red things and their experiences.

There does not seem to be any theoretical obstacle to capturing the conceptual roles of phenomenal concepts by standard Ramsey-Lewis techniques. Interestingly, however, their conceptual roles do not seem essential to establishing the semantic contact between the phenomenal concepts and what they denote; nor for that matter are these roles part of their semantics, as indeed emphasized by Jackson (1994) and Chalmers (1996). Their semantics seems to be exhausted by their reference, which is fixed independently of their conceptual roles, rather by somehow coming into direct and nonmediated contact with the qualitative character which they denote. Indeed, the persistence of intuitions about inverted spectrum and absent qualia seems to show this.

As it has been getting increasingly clear, perspectivalist materialists can agree to all of this — or so we claim. So far there is nothing that shows that what is being
confronted, and what the phenomenal concepts directly pick out, is non-physical. Even if what is confronted in experience does not present itself as physical/functional or as having internal constituents, in other words, even if the phenomenal concepts directly pick out qualities of experiences without presenting them as physical/functional and complex, still, for all we know, what is thus confronted or picked out may be entirely physical/functional and complex. What the anti-materialist needs is a plausible story that would turn these observations into an anti-materialist argument. Indeed, what makes recent anti-materialist arguments based on conceivability more interesting than their historical ancestors is precisely that they attempt to supply such stories.

In the hands of Jackson and Chalmers, this story turns on the nature of reductive explanations, which is closely tied to considerations of logical supervenience. We will not examine their arguments here in any detail. Suffice it to say that there are two key elements in their argument. We have just described one, namely that phenomenal concepts are not analyzable: (a) they pick out their referents directly and immediately, and (b) they do not conceive of their denotations as physical/functional and complex. The other is that nothing can logically supervene on the physical, or nothing can be reductively explained — where the reduction base is physical — unless there is a derivation of (a statement of) the fact to be reductively explained from the physical facts. On their account, it so turns out that this kind of reduction (and logical supervenience) requires that the concepts used in the expressions of such facts, i.e., facts to be reduced, are always analyzable in one way or other; or else, they pick out what they do as already essentially physical/functional. Given that this is so, Chalmers and Jackson claim that they can show how almost everything can in this sense be reductively explained, i.e., how almost any fact can be derived from physical facts, except the phenomenal facts.

We think that they are probably mistaken in their claim that reductive explanations require derivations in the sense they intend. But we will not pursue this here. Instead what we want to highlight is that their argument turns on the fact that our phenomenal concepts are unique in that they pick out their denotation directly and immediately, and while doing that they do not conceive of their referent as physical/functional and complex — thus they are semantically unanalyzable. The result is that there is nothing to stop us from conceiving the physical and the phenomenal as existing independently of each other, for there is nothing that conceptually ties phenomenal concepts to the physical/functional.

We submit that this is in fact the basis of all conceivability arguments: they harbor a premise to the effect that it is genuinely conceivable that there could be a world physically/functionally identical to ours that lacks phenomenal qualities and consciousness. And what underlies this claim is, to repeat, that phenomenal con-
cepts pick out their referents directly and immediately and they do not present (conceive of) their referent as physical/functional and complex, which yields the conceptual independence of phenomenal concepts from physical/functional ones. When this is combined with what is required by reductive explanations, it follows that qualia cannot be reductively explained. Given the tight connection between reductive explanation and metaphysical supervenience that most defenders of conceivability arguments envision, this yields, in turn, the metaphysical result that qualia do not “logically” supervene on the physical. We will come back to this last move below.

It is important to point out that the sort of anti-materialist argument that Jackson and Chalmers defend can be run even without any commitment to conceptual analysis and the analyticities generated by such, any form of which is bound to be controversial these days. So suppose, instead of talking about analyses of concepts, we can talk about how the denotation of any concept is established (how they do acquire their referential semantics). Call any mechanism that underlies this referential contact, a reference fixing mechanism. Then the relevant assumption in the Jackson/Chalmers argument can be recast accordingly: namely, that all concepts except the phenomenal (and, sensory — see below) ones have reference fixing mechanisms that deploy other concepts. One can then run, *mutatis mutandis*, the same sort of conceivability argument against materialism without committing oneself to any view about whether these concepts deployed in fixing the reference is part of the semantics of the target concept. These mechanisms may even be captured by standard Ramsefication methods: the functional roles thus generated may then be taken as merely reference fixers, or as mechanisms that sustain the informational/referential contact between the concepts and their denotations. Put differently, these roles may be what determine the reference of a concept by determining what they will hook up with in a given context.

If our phenomenal concepts pick out their referent directly and do not present them as physical/functional and complex, it will always be possible to conceive of a physically identical world without phenomenal properties. This seems to be the way Kripke runs his modal argument against materialism. As you will recall, the reason why psychophysical identities (e.g., that pain = C-fiber-firing) are special, on Kripke’s view, is because their apparent contingency cannot be explained away in the way in which the apparent contingency of other so-called scientific identities (e.g., that water = H₂O) can be explained away. And the reason for this turns out to be that whereas our folk natural kind concepts pick out their referent by some reference fixing mechanisms that deploy other concepts — especially the concepts we use in describing how natural kinds sensorially/perceptually seem to us, all contingently tied to the target concept — phenomenal concepts pick out their referent di-
rectly/immediately, or as Kripke sometimes puts it, by their essential properties, namely on the basis of how they feel. But then, our phenomenal concepts (at least the most specific ones) do not conceive of these feels as physical/functional and complex. So it seems that for Kripke, too, it is genuinely conceivable that C-fiberfirings exist without pains (and vice versa).

In fact, we can generalize the main point that seems to drive all conceivability arguments as follows:

\[
\text{(PC)} \quad \text{Necessarily, the acquisition of phenomenal concepts (from experiences) and their informational deployment (in direct consequence of having the relevant range of experiences — call such deployments vertical) are such that:}
\]

\(\begin{align*}
(a) & \quad \text{there are no intervening mental states — intervening between the experiences and their vertical tokenings — that carry further information about the object/condition they denote (or, are about) such that this information has to be used in their acquisition and vertical deployment, and} \\
(b) & \quad \text{is consciously available to the same person for further conceptualization, whether or not this information is actually conceptualized.}
\end{align*}\)

Condition (a) implies that the abstraction distance between the relevant feature of an experience (say, its being a “reddish” experience) and its conceptualization (\(\text{EXP}_{\text{red}}\)) is maximally short, i.e., there is (almost) no abstraction going on. The effect of condition (b) is to ensure that if there is in fact information of the relevant sort that is used in fixing the reference, it is not consciously available — assuming (as we do) that conscious availability requires availability to a conceptual system. To illustrate the importance of this last condition, consider the limiting case: an experience of red carries information about whatever objective physical property of a surface is detected as red (which we identify as \text{redness}). If this property is a complex physical property like a surface spectral reflectance of a certain sort, then the experience carries information about this complex property. But certainly this information, even if carried by the experience itself, is of no use for conceptualizing the constituents of the objective property (\text{redness}) detected. The relevant feature of the experience that carries information about \text{redness} does not present this information to the conceptual system as if \text{redness} had constituent structure. But then the relevant feature of the experience does not present itself as if it had a constituent structure. Thus the phenomenal concept, \(\text{EXP}_{\text{red}}\) directly and immediately acquired from this feature, does not present it as having a complex structure either. Indeed, (PC) describes the special nature of the direct and immediate link between phenomenal concepts and the qualitative character of experiences they denote. No other concepts seem to work this way.
We think that a materialist can — and, given the persistence of qualia intuitions, ought to — agree to all of this: that phenomenal concepts are unique in just this way, and any statement denying the non-identity of qualia with the physical/functional is genuinely conceivable, insofar as conceivability is a matter of concept use and the nature of phenomenal concepts is as stated. To prevent potential misunderstandings, it is important to emphasize that we use ‘genuinely conceivable’ in a technical sense that we can explain by comparing the case at hand to the standard ways of explaining the intuitions about the conceivability of the denials of true a posteriori necessities (e.g., denials of true scientific identities or supervenience claims) like ‘water is not H$_2$O’.

We will say that these denials are not genuinely, but only apparently, conceivable just insofar as there are statements that are C-related to these denials such that what is actually and genuinely conceived are these C-related statements. C-relations are explained in terms of the connotations of commonsense kind concepts used in the denials. Depending on what kind of semantic theory one assumes, connotations can be explicates either as semantic analyses of these concepts, or as merely the reference fixing mechanisms that happen to deploy other concepts (or, the information carrying intervening states of the sort we have introduced above) — or, as a combination of both. So the claim would be that what is claimed to be conceived is not what is truly expressed by

(3) water is not H$_2$O,

but rather a statement which is C-related to (3) such as

(4) the substance that is colorless, tasteless, odorless, etc., and that typically falls on earth in rain, that we bathe in, drink, …, is not H$_2$O.

As stated, the concepts used in the definite description may function as only reference fixers for the concept of water (WATER) or as somehow giving (part of) its semantic content (or, as both).

So when we grant that statements like (Z) below are genuinely conceivable — statements containing phenomenal concepts and denying the identity (or supervenience, as the case may be) of their denotations to anything physical/functional — all we are claiming is that there are no statements that are relevantly C-related to them.

(Z) There exists an exact physical/functional duplicate of me at a time when I know I am having an experience with quale Q, without this replica’s having an ex-
If phenomenal concepts are unique in the way described by (PC), then there will not be any relevantly C-related statement available for explaining the apparent conceivability of statements like (Z). So, obviously, to the extent to which the standard strategy of explaining the apparent conceivability of statements like (3) crucially relies on there being statements relevantly C-related to them, and to the extent that the existence of C-related statements minimally require semantically relevant (or, informationally loaded) reference fixing mechanisms, to that extent the standard strategy will not work for Z-like statements involving, as they do, concepts of which (PC) is true.

Indeed, (PC) is precisely what ultimately underlies Kripke’s strategy for blocking the physicalist’s obvious reply who insists that scientific identities (supervenience claims) are necessary but a posteriori (so that a Fregean co-denotational-but-distinct-concepts-strategy can apply): such claims come with an obligation to explain away the intuition that they are contingent, or as Kripke likes to put it, the intuition that “there is an apparent air of contingency” about such statements. But it is precisely the lack of C-related statements that makes the case for the genuine conceivability of Z-like statements (e.g., ‘pain is not C-fiber-firing’), and thus blocking one obvious materialist strategy to explain the relevant intuitions away.

We submit that the lack of C-related statements is due to the semantic/referential structure of phenomenal concepts and the unique nature of their vertical acquisition and deployment. And the non-availability of this strategy (to arguing for the claim that Z-like statements are only apparently conceivable) is a consequence of this, and a very significant result of recent philosophical developments.

4. The Unargued Premise

This result about conceivability, however, in and of itself, does not seem to have any consequence for or against materialism unless conceivability is somehow connected to ontology. From what has been said thus far, a materialist can agree that no true statement about phenomenal qualities can be derived from purely physical/functional premises, and all Z-like statements are genuinely conceivable. But with no further premise or argument, there is nothing in any of this that threatens materialism: conceivability is purely a matter of psychology and epistemology. So, how the psychology of conceivability turns out, in and of itself, has no tendency to show anything metaphysical. What is needed is a further argument that would take us from genuine conceivability to genuine metaphysical possibility. What is needed, in other words, is an argument for what we might call the Bridging Premise:
(BP) For any proposition \( P \), if \( P \) is genuinely conceivable, then \( P \) is metaphysically possible.

Clearly, both Jackson and Chalmers, as well as Kripke, accept something like this premise, although it is suppressed in their arguments.

Notice that to say that \( Z \)-like statements are genuinely possible is to say that phenomenal qualities cannot be reductively explained, as argued by Jackson and Chalmers. So, if we are right, the failure of reductive explanation of the envisioned sort, in and of itself, does not threaten materialism. This might at first seem odd. But remember, we are operating with a notion of reductive explanation that requires that the concepts whose denotations are to be reduced have either semantic analyses or else fairly rich reference fixing mechanisms involving other concepts or containing information available for further conceptualization. We did not explain the operative notion of reductive explanation assumed by Jackson and Chalmers in any detail because we believe that the main element of such a notion which makes it both relevant and crucial to anti-materialist conceivability arguments can be brought out clearly without going into the details of what reductive explanation is: this element is the semantic and acquisitional structure of concepts involved as described above, i.e., (PC).

Levine (2001) also assumes a similar notion of reductive explanation except that he does not think that the target concepts to be reduced have, or ought to have, semantic analyses. But it is clear from his discussion that successful reductive explanations all turn out to have concepts that are embedded in fairly rich systems of beliefs that can then be used at least as reference fixing mechanisms for those concepts. For Levine, the structure of the reductive explanations of qualia is the same as the structure of other standard reductive explanations: both are committed to identity statements at some stage of reduction to connect the two disparate vocabularies so that the reduction exhibits a deductive structure. Levine notes, however, that there is still an explanatory gap in the case of envisioned reduction of qualia, because the identities in such reductions (e.g., ‘pain = C-fiber-firings’) are “gappy”, quite unlike the identities (e.g., ‘water = \( H_2O \)’) involved in standard successful reductions. Levine (2001) does not quite say what makes the identities involved in phenomenal reductions gappy, although he clearly thinks that this is the source of the explanatory gap, and rightly suspects that it stems from the structure of our concepts. 22

We submit that the identities like ‘water = \( H_2O \)’ are not gappy because WATER is embedded in a fairly rich web of information/beliefs that functions at least as a reference fixer for WATER. Our point is that what is distinctive about phenomenal concepts and why they yield gappy identities stem from their seman-
tic/referential structure and from the way they are acquired and deployed as we explained it above: their acquisition and vertical deployment are direct and immediate with no intervening concepts or states that carry consciously available information used in reference fixing. Indeed this is what sets phenomenal concepts apart from all others.

So if we are right about the nature of phenomenal concepts, the failure of reductive explanation should not come as a big surprise carrying with itself a metaphysical tag. What is needed to turn the radical difference between phenomenal concepts and all others into something metaphysically significant is a good argument for (BP).

What is the status (BP)? There are basically two options. One is to read it as a logical entailment. Indeed, this seems to be the preferred reading for Chalmers. But conceivability is purely a matter of epistemology or psychology — i.e., the capacity of cognitive organisms to represent metaphysical reality one way or other, and this depends on what concepts one has and what the nature of these concepts is. As such any reflection on what is conceivable and what is not ought to have no logical bearing on the constitution of the metaphysical reality (necessity/possibility). To think otherwise is to risk, we claim, an indefensible form of verificationism. So we reject (BP) as implausible on this strong reading. There are good — and, as far as we can see, conclusive — arguments in the literature against this reading. So we will not dwell on this too much. The only remotely plausible arguments in favor of reading it as an entailment seem to come from considerations about the epistemology of modality and rationality. The claim is that if we do not read (BP) as entailment, we will not have a satisfactory epistemology of modality. We think this conditional is false. But even if we had to choose between having an unsatisfactory epistemology of modality and an unsatisfactory metaphysics by giving up materialism, we — probably along with most others — would opt for the former instead of giving up materialism. The lesson, in brief, is that if the only arguments for reading (BP) as entailment are (often philosophically obscure) considerations about the epistemology of modality, and the cost of reading (BP) as an entailment is the denial of materialism, then good philosophical judgment dictates rethinking about the epistemology of modality. We (again, along with many others) are ready to do whatever adjustments are necessary to the epistemology of modality to accommodate materialism — if this turns out to be necessary.

We do not think, however, that our epistemology of modality is in need of radical revisions. It can be saved by a weaker reading of (BP). There is no doubt that conceivability is often our only, and often quite reliable, guide to possibility. As such it would be foolish to reject it wholesale. We take it that (BP) can and should be read as stating a reliable but defeasible rule of inference. Alternatively, it
could be read as a ceteris paribus generalization. In this weaker sense, (BP) can readily be accepted by materialists.

But if so, materialists have an obligation to tell why (BP) needs to be suspended when conceivability involves phenomenal concepts; they need to find principled reasons to block the licensing force of the rule embodied there in — that is, they need to provide legitimate defeaters in its application to the case at hand where genuinely conceivable Z-like statements involve phenomenal concepts. In the absence of providing a principled and naturalistic defeater, the materialist strategy of insisting that conceivability, in and of itself, does not yield a metaphysical result is quite unsatisfactory, and can rightly be accused of special pleading. All perspectivalist materialists have a debt to discharge in this regard. This is not an easy job to do, and it is also not clear whether it is purely a philosopher’s job.

We should note that until very recently (BP) has been assumed without any serious argument by anti-materialists, or without any serious argument to show that the particular use into which they put it in their argument is unproblematic. Saying this and leaving the matter here, however, is not enough for materialists. As we said, what needs to be provided is an account of why (BP) should be suspended in the case of consciousness, since rejecting (BP) wholesale, as far as we can see, is not open to materialists either. The dialectics of this situation puts materialists in a particularly tough position susceptible to accusations of ad hocness and unfairness, because perspectivalist materialists who accept the dialectics described so far are in effect saying that they accept (BP) generally except when applied to consciousness. The burden of proof here is indeed on the shoulders of materialists.

Notice how close such a perspectivalist materialist comes to embracing all the intuitions and premises of the anti-materialist who argues on the basis of conceivability arguments — assuming that the materialist has a principled defeater preventing the last coup de grace. Drawing on the acquisitional nature and semantic structure of phenomenal concepts, the perspectivalist can grant — indeed, insists on — the genuine conceivability of Z-like statements, and even accepts (BP), if only with a weaker reading. This is all good news for materialism provided that there is indeed a defeater for the last move. For the intuitions driving the anti-materialist arguments have proved to be very resilient and powerful, and we do not think that this is an accident or some sort of intellectual blindness. The strength of these intuitions should be acknowledged and accommodated by materialism. Indeed, it seems that the success of the materialist’s pending account of why (BP) does not apply in the case of consciousness depends pretty much on its ability to explain and predict why we have these intuitions in the first place.
5. Strategic Meta-Reflections

So, can the perspectivalist physicalist successfully discharge the burden of proof and provide a principled defeater that would make the use of (BP) illegitimate in conceivability arguments? What would such an account look like?

As we have seen, perspectivalism is typically advanced in three stages. First, it diagnoses the puzzle, involved in attempting to conceive of the phenomenal in terms of the physical, as a Frege case, namely, as one arising from co-denotational but distinct concepts. Secondly, it points out that the Frege case at hand is nevertheless quite special in a way that marks it off from all other Frege cases we know of, and this specialness needs accounting for. Thirdly, it postulates a group of concepts, typically called “phenomenal concepts”, whose nature is said to be perspectival. As far as we can see, all extant perspectivalist accounts fail to completely discharge the aforementioned burden by failing to integrate the account of phenomenal concepts into a general naturalistic account of concepts at large. Thus, they are rightly open to the charge of being *ad hoc*: it is only when the nature of phenomenal concepts is accounted for on the basis of a general naturalistic semantics for concepts is it reasonable to allow for a special pleading for them on the part of naturalists.

Let us, however, that there is such an account, and it is thoroughly naturalistic (say, information-theoretic à la Dretske (1981) or Fodor (1987, 1990)) and successful (or, at least, very promising). Suppose that this general account explains and predicts a certain set of interesting features of a special class of concepts, say the *sensory* concepts that apply to sensible qualities of objects, to so-called secondary qualities. In fact, suppose that it follows from the basic information-theoretic principles of this general theory that necessarily there are certain concepts whose acquisition and informational deployment are direct and immediate in that there is no abstraction distance of the relevant sort between these concepts and the experiences from which they are acquired, which is to say that there are no intervening states in between them that carry information about the objects of experiences available for further conceptualization. Suppose further that an information-theoretic account of these sensory concepts entails that these concepts are bound to (re)present their denotations as having no internal constituents, i.e., as not complex. The intuitive principle behind this would be that simple signals can carry complex information, i.e., information about the complex structure of events to which they are informationally connected.

Now surely there is some very close internal connection between *sensory* concepts and *phenomenal* concepts. As far as we can tell, no perspectivalist has so far provided any account of what that might be: sometimes they are even conflated. So suppose that it follows again from the information-theoretic principles that sen-
sory concepts carry information about their denotations out in the world by carrying information about experiences, about their “proper” sensory bases from which they are directly and immediately acquired. This would mean that there is information about these experiences in the proper vertical tokenings of sensory concepts that is not conceptually recoverable or extractable. Now suppose that what makes a concept phenomenal is a kind of re-utilization of a sensory concept to denote the most immediate object/event it carries information about, i.e., experiences from which it is immediately acquired or their relevant features. This would mean that phenomenal concepts do not (re)present the relevant features of these experiences as having a complex structure.

Now as naturalists, materialists would want to identify experiences with physical/functional events in the brain. Informational semantics requires this anyway on the grounds of methodological considerations about causality and nomic relations of the sort we have pointed out at the beginning. Then we would have in our hands a general naturalistic account of concepts that would yield exactly what is claimed to be so special about these perspectival phenomenal concepts: namely, they are nothing but re-deployments of sensory concepts acquired immediately and directly from experiences such that they do not present them as physical/functional and complex. This would account for why introspection does not reveal any “brainish” quality if we can introspect the qualities of our experiences. In other words, we would have a completely justified and principled account of what makes these phenomenal concepts so special and unique that would explain and predict why acquisition and informational deployment of these concepts would be puzzling vis-à-vis the conceivability of zombies and reductive explanations.

Now finally, suppose that all these “suppose”s can be successfully sustained and cashed out: can the perspectivalist materialist then claim to have discharged the burden in a principled way, and shown why using (BP) in conceivability arguments against materialism is illegitimate? We think that the answer is a resounding “yes”. If our diagnosis about what makes phenomenal concepts resistant to reductive explanation is right, and genuine conceivability is a matter of concept use of this nature, then any successful naturalistic account of concepts that has the consequence of showing why phenomenal concepts are the way they are ought to preempt the use of (BP) in conceivability arguments against materialism. We do not know what more can reasonably be asked of a perspectivalist materialist to rest their case.

6. Conclusion
In this essay, we have tried to present the flavor of how a general naturalistic account of phenomenal concepts would look like. In fact, we were more specific than required: we drew on a specific information-theoretic psychosemantics. However, any naturalistic theory that succeeds in showing in a principled way how there could (even, should) be sensory and phenomenal concepts whose acquisition (and informational deployment) meet the directness and immediacy requirements, as well as how there could be such concepts that do not conceive of their referent as having a complex nature and thus as being something physical/functional — i.e., any naturalistic theory that succeeds in showing how (PC) is satisfied in the case of phenomenal concepts — will do to block the use of (BP) in conceivability arguments.

It is important to remember where we are in the dialectics of the debate. Successfully discharging the burden by providing a naturalistic and principled account of phenomenal concepts, in and of itself, does not constitute an argument against anti-materialism. It does not show that epiphenomenalism is false; accordingly, it does not demonstrate that materialism is correct. What it does show is that conceivability arguments against materialism are unsound, and to that extent no knockdown argument against materialism has yet been offered. This result yields a powerful argument for materialism only when combined with the methodological assumption (M) we have made at the beginning. But having a general naturalistic account of sensory/phenomenal concepts of the sort we have outlined will considerably strengthen this case for materialism by showing that materialism can in fact explain and predict the intuitions powering the conceivability arguments. In fact, if such an account could be given, it would also constitute a strong argument against reading (BP) as a logical entailment, which would indeed force us to reconsider what the epistemology/psychology of modality and rationality ought to look like.

What is most interesting to note about this dialectic is that a naturalistic story of the sort that would constitute a defeater for (BP) and ultimately argue against the a prioristic conceivability arguments would likely be empirical in nature. For we take it that a naturalistic account of phenomenal concepts would ultimately be a form of philosophically motivated but empirically informed theoretical psychology to be vindicated by findings from empirical science.

Indeed, consider Jackson’s thought experiment again about the omniscient color scientist, Mary. Before her release, suppose that Mary knows all there is physical to know not only about colors and color vision but also all about introspection and concept formation. Then, supposing that there is a naturalistic account of these of the sort that would prevent the use of (BP), the details of this account are what she would know exhaustively. But then she would automatically be in a position to know about the curious asymmetry involved in the epistemic access to phe-
nominal/physical facts as claimed by perspectivist materialists. This body of knowledge she has before her release would not of course remove her curiosity (the surprise element) about coming to know in a first-person way about facts she already knew under their scientific description. On the contrary, we would expect that she would be even more curious and intrigued to instantiate those phenomenal/physical states herself, which are necessary for her to acquire the peculiar perspectival concepts, and thus first-person knowledge. Knowing all the scientific facts would make her also know that she lacks a certain class of concepts necessary to know facts in a perspectival way different from the way she already knew in a third-person way. We would expect her not to be moved by the familiar conceivability arguments at all. Given her scientific omniscience, she would be in a position to know better.

All this is as it should be: if we are right about the dialectic of the debate and the importance of providing a general account of phenomenal consciousness, it is clear that the strategy to pursue in order to go beyond what seems to be a standoff between the parties and settle the debate is to develop empirically informed and ultimately scientific theories of sensory/phenomenal concepts — and how they interface with sensory experiences — to reveal their peculiar nature that gives rise to puzzling intuitions and conceivability arguments in the first place. If successful, such an endeavor would be a perfect example of an achievement in one of the most important intellectual roles philosophy has assumed historically: philosophy as a proto-science and/or a theoretical endeavor interacting with sciences but working in their foundations.29

NOTES

1In the literature, there are anti-materialists who claim to hold positions different from epiphenomenalism, such as the “naturalistic dualism” of Chalmers (1996). We believe that these variations of anti-materialism are all committed to some form of epiphenomenalism in the end, but we will not argue for that here.

2Indeed even Jackson, one of the most articulate and influential advocates of epiphenomenalism, has recently given up his position in favor of materialism mainly on the basis of these kinds of considerations about causality. See his “Postscript on Qualia” in his (1998).


4For an excellent review, see van Gulick (1993).

5Or, conception — most of the times we will use these interchangeably. What
is meant is the mental representation (conception) whose linguistic expression is a description.

We will assume that there is an objective property that color experiences and our color concepts represent. This may be a relational or dispositional property, but for convenience we will assume a primary quality view of colors (and other so-called secondary qualities in general) à la Armstrong (1968) and Hilbert (1987), but nothing crucial hangs on this assumption. Also, we will use words in caps as referring to concepts as mental representations, while using italics to refer to the (instantiations of) properties these concepts refer to. So “red” refers to red (or, redness) which in turn is expressed by RED. We will be relaxed in talking about a property as the reference of a concept.

This is, in essence, the account that Loar (1997) and Lycan (1996) present, although Lycan prefers to express the main point in terms of “phenomenal information” rather than “phenomenal concepts”. See also Papineau, Sturgeon, and Hill, op cit.

We are ignoring Lewis/Nerimow “ability” reply to Jackson here to confront Jackson in his own terms. But we also find it hard to believe that ability is all there is to what Mary gains without acquiring new representational/conceptual capacities (see Loar 1997 and Lycan 1996 for criticisms of the “ability” reply).

‘SD’ stands in lieu of a scientific description of experiencing red, but when Mary uses this linguistic expression in speaking English she expresses her thoughts that are made up of its conceptual counterpart. This is a mental representation realized in her brain whose reference (i.e., experience of red) is also realized in her brain upon actually seeing red for the first time.

Where the modal force of “necessarily” is nomological. Here we will bypass the complications that arise depending on whether one holds an actualist or counterfactualist theory of psychosemantics. Note that nothing we have said so far implies that experiencing red is, by itself, sufficient for EXPred. More needs to be said about the nature of this sort of perspectival concepts — in particular, one needs to account for the fact that they seem to be both indexical and predicative at the same time. But spelling out of this ought to be part of a naturalistic story, which, as we will urge later on, must be developed in empirically sensitive ways if we want to settle the impasse in the debate between materialism and anti-materialism.

This is not to say that they and the experiences from which they are acquired represent their denotations as simple or as having no constituents. They are essentially topic-neutral as Smart (1959) emphasized long time ago.

Lewis (1972), Block (1980a).

This is not to say that they are independent of their functional roles if these roles include causal links between concepts and the experiences from which they
acquired. We mean by conceptual roles only those generated by their being embedded in propositional attitudes.

14 Here we will forgo discussing the early Australasian attempts to define sensation concepts and their qualities by their commonsense functional roles. We think that absent qualia arguments and spectrum inversion thought experiments à la Block (1980b) show that qualia concepts cannot be defined by such roles — see also Block (1995).

15 For arguments to that effect, see Block and Stalnaker (1999); cf. Levine (2001).

16 Kripke writes (1980: 152–3): “To be in the same epistemic situation that would obtain if one had a pain is not to have a pain… Pain… is not picked out by one of its accidental properties; rather it is picked out by its immediate phenomenological quality… If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon is pain.”

17 The parenthetical remark is in fact unnecessary: what makes a concept phenomenal is not what it denotes, but rather how it is acquired. Phenomenal concepts are those that are acquired from their “proper” sensory bases — this is partly why they are perspectival. But see fn. 10 above.

18 Excepting sensory concepts such as RED, SWEET, LOUD, etc. which are acquired in fact prior to phenomenal concepts like EXP \textsubscript{red}, EXP \textsubscript{swe}, etc. See below. We describe the relation between them in detail in Aydede and Güzeldere (ms.); there we also explain the special nature of the acquisition and vertical deployment of such concepts in purely information-theoretic terms à la Dretske (1981).

19 For a detailed elaboration of this line, see Loar (1997) and Levine (1993).

20 The point is not just that we may be ignorant about the relevant C-related statements. It is rather that they do not exist.

21 (Z) in effect describes a phenomenal zombie. There are various non-equivalent ways of describing zombies (e.g., on the basis of local vs. global supervenience; token vs. type identities; state-vs. individual-vs. species-based scope; by including propositional attitudes vs. not, etc.). Some of these involve important nuances. We will not bother to be more specific here as our main point can be made without such detail.

22 Levine (1993: 134) puts the point succinctly: “to the extent that there is an element in our concept of qualitative character that is not captured by the features of its causal role, to that extent it will escape the explanatory net of a physicalistic reduction”.

23 Among others, see Yablo (1993), Levine (1993, 1998, 2001), and Byrne (ms.).

24 See Chalmers (1996) and (ms.).
Again see Yablo and Levin op cit.

Loar (1997) may be an exception to this sweeping claim, but we think that even Loar fails to make a case for what is so special about phenomenal concepts by not giving an account of the relationship between sensory concepts (like RED, SWEET, etc.) that apply to qualities of objects in the first instance, and phenomenal concepts (like \( \text{EXPR}_\text{red} \)) that apply to experiences of these qualities. However, for Loar’s polemical purposes in his (1997), this was perhaps not needed.

See, for instance, Churchland (1985, 1989), and to some extent, Tye (1999).

See Aydede and Güzeldere (ms.) for a detailed information-theoretic account of sensory and phenomenal concepts along the lines outlined. By giving such a general account, we claim to have demonstrated that all these “suppose”s can be successfully cashed out.

We are in agreement with Fodor (1981) when he says, “Philosophy is what you do to a problem until it’s clear enough to solve it by doing science”. We would like to read this not necessarily as expressing a linear process, but rather as expressing a continuous interaction between sciences and philosophy, influencing and enriching each other in complex ways.

REFERENCES


AYDEDE, Murat, and GÜVEN GÜZELDERE (ms.). Information, introspection, and phenomenal consciousness. Available at http://web.clan.ufl.edu/users/maydede/introspection.pdf


BYRNE, Alex (ms.). Chalmers’ two-dimensional argument against physicalism, MIT.

CHALMERS, David (ms). Does conceivability entail possibility? Available at http://www.u.arizona.edu/~chalmers/papers/conceivability.html


