

TERRY HORGAN

MATERIALISM: MATTERS OF DEFINITION, DEFENSE, AND DECONSTRUCTION

ABSTRACT. How should the metaphysical hypothesis of materialism be formulated? What strategies look promising for defending this hypothesis? How good are the prospects for its successful defense, especially in light of the infamous “hard problem” of phenomenal consciousness? I will say something about each of these questions.

1. FORMULATING MATERIALISM¹

In seeking a satisfactory formulation of materialism, it helps to employ the notion of a *possible world*. Possible worlds are plausibly construed not literally as universes other than the single real universe (i.e. not as *cosmoi*), but rather as *total ways the cosmos might be* – i.e. maximal *properties* instantiable by the single real world (the single cosmos). On this usage, the item designated as the actual world – considered as one among the various possible worlds – is not itself the cosmos either, but rather is the total cosmos-instantiable property that is *actually* instantiated by the cosmos. But it will be convenient in practice to speak as though the actual world is the cosmos and as though other possible worlds are other such *cosmoi* – a harmless enough manner of speaking, as long as one bears in mind that it is not intended literally. (The various claims made below in terms of possible worlds can all be reformulated in terms of the language of maximal cosmos-instantiable properties, but I will not bother to do so.)

The possible worlds we are interested in are *physically* possible worlds – that is, worlds in which obtain all the same physical laws that actually obtain in the cosmos.² Moreover, they do not contain – alongside various physical entities like quarks, electrons, and composites entirely composed of such physical parts – any such entities as immaterial Cartesian souls. In order to hone in on the physically possible worlds we are interested in, it is useful to borrow from Jackson (1998) the idea of a *minimal duplicate* of a physically possible world: if w is a physically possible world, then w^* is a minimal physical duplicate of w just in case (i) w^* is *physically* just like w , and (ii) w^* contains no entities other than those required for it to meet condition (i). Here then is an initial first pass at characterizing the metaphysical hypothesis of materialism:

(M₁) The actual world w is a minimal physical duplicate of itself; and for any physically possible worlds w_1 and w_2 , (i) if w_1^* is a minimal physical duplicate of w_1 , and (ii) w_2^* is a minimal physical duplicate of w_2 , then $w_1^* = w_2^*$.

A problem with M₁, however, is that it does not preclude a putative possible world that (1) is physically possible, (2) is a minimal physical duplicate of itself, and (3) contains two spatio-temporal regions that are just alike in all intrinsic physical respects but are not just alike in *all* intrinsic respects. The metaphysical hypothesis of materialism, pretheoretically understood, surely *should* preclude such putative worlds, however. It also should preclude the possibility of such pairs of spatiotemporal regions in *distinct* physically possible worlds (provided that each of the distinct worlds is a minimal physical duplicate of itself). i.e. regardless whether the two spatio-temporal regions are within the same physically possible world or two different ones, if these two regions are intrinsically just alike in all physical respects then they should be intrinsically just alike simpliciter (provided that each region belongs to a world that is a minimal physical duplicate of itself). So here is a second formulation that eliminates this loophole in (M₁):

(M₂) The actual world is a minimal physical duplicate of itself; and for any physically possible worlds w_1 and w_2 , if (i) w_1^* is a minimal physical duplicate of w_1 , (ii) w_2^* is a minimal physical duplicate of w_2 , (iii) r is a spatiotemporal region of w_1^* , (iv) s is a spatiotemporal region of w_2^* , and (v) r and s are intrinsically just alike in all physical respects, then r and s are just alike in *all* intrinsic respects.

Notice that (M₁) is a special case of (M₂) – viz., the case where r comprises all of the spatiotemporal extent of w_1 rather than a proper part of it, and s comprises all of the spatiotemporal extent of w_2 . Notice too that (M₂) also covers special cases where r and s are spatio-temporal parts of a *single* world – viz., cases where $w_1^* = w_2^*$.

But even (M₂) is still too weak to explicate the doctrine of materialism. Consider, for instance, Moore's metaethical position (Moore, 1903, 1922). Moore held that *intrinsic goodness* is an objective, non-natural, property. He held that its instantiation is supervenient on the instantiation of certain natural properties (although he did not use the term 'supervenient'), in a modally very strong way: in *any* possible world in which thus-and-such natural property is instantiated by an individual i at a time t , the non-natural property of intrinsic goodness is thereby instantiated by i at t . He also held that the necessary connection linking the pertinent natural property to intrinsic goodness is itself metaphysically fundamental and *sui generis* – rather than being derivative from any other facts. (He held that we know such metaphysically fundamental, synthetic, necessary truths by a special faculty of moral intuition.)

Moorean non-naturalism in metaethics surely should not be considered consistent with metaphysical materialism. Yet, because of the modal strength of the metaphysically fundamental necessary connections that supposedly obtain between certain natural properties and intrinsic goodness – connections that obtain in *all* possible worlds, even though they are synthetic – Moorean non-naturalism is indeed consistent with thesis (M₂), and hence with (M₁) too. Even a *minimal* physical duplicate of any physically possible world, or of any spatiotemporal region of such a world, will be just like that world (or region) with respect to how the non-natural property of

intrinsic goodness is instantiated. So (M₂) is too weak, as it stands.

Here is another way to make the point. *Standard emergentism*, as I will call it, is a non-materialist metaphysical position asserting that certain properties – for instance, *phenomenal-consciousness* properties of one kind or another – supervene with *nomic* necessity upon certain physical properties. (Advocates of standard emergentism have often held too that these emergent properties generate additional fundamental causal forces, over and above physical causal forces. But that is a separate doctrine that we can set to one side here.) But in principle, someone could advocate a non-standard form of emergentism – call it *Moorean* emergentism – asserting (1) that certain properties (e.g., phenomenal-consciousness properties) supervene with *metaphysical* necessity on underlying physical properties (rather than supervening with mere nomic necessity), and (2) that the metaphysically necessary connections that obtain between the physical supervenience-base properties and the emergent supervenient properties are themselves metaphysically fundamental and *sui generis* – rather than being derivative from any other facts.

Moorean emergentism surely should not be considered consistent with metaphysical materialism. Just as standard emergentism is at odds with materialism because it posits metaphysically fundamental *laws of nature* other than the metaphysically fundamental laws of physics (viz., inter-level laws linking physical properties to emergent properties), likewise Moorean emergentism is at odds with materialism because it posits *sui generis principles of metaphysical necessitation* (viz., inter-level principles expressing brute, fundamental, metaphysical-necessitation relations between physical properties and emergent properties). A doctrine worthy of the label ‘materialism’ should not countenance brute inter-level relations of metaphysical necessitation, any more than it should countenance brute inter-level laws of nature. And it should not countenance Moorean-emergent properties, any more than it should countenance standard-emergent properties. Yet

Moorean emergentism is consistent with thesis (M₂) above, and hence with thesis (M₁) too – for the same reason that Moorean non-naturalism in metaethics is consistent with these theses. This reinforces the moral that thesis (M₂) is too weak to be an adequate explication of materialism.

The moral can be further reinforced in yet another way. In principle, someone could advocate a form of Moorean emergentism involving non-physical *particulars*. For instance, one could hold that in every metaphysically possible world, living organisms with the physical composition of human beings *have immaterial Cartesian souls, in which inhere mental properties*. That is, one could hold that full-fledged Cartesian dualism obtains as a matter of metaphysical necessity. Even *this* view is consistent with thesis (M₂) – since the view entails that for any physically possible world *w* containing living organisms with the physical composition of human beings, any minimal physical duplicate of *w* will be a world in which those organisms have immaterial souls. *If there are brute relations of metaphysical necessity linking human physical composition to possession of a Cartesian soul, then there is no metaphysical peeling away of the immaterial souls by appeal to the idea of minimal physical-duplicate world*. So: since there is a form of Cartesian substance-dualism that is consistent with thesis (M₂), and since no form of Cartesian substance-dualism is consistent with materialism, thesis (M₂) is not an adequate explication of materialism.

How then should (M₂) be strengthened? I propose the following thesis, which incorporates the content of (M₂) but adds to it:

- (M₃) (1) The actual world is a minimal physical duplicate of itself,
 (2) for any physically possible worlds *w*₁ and *w*₂, if
 (i) *w*₁* is a minimal physical duplicate of *w*₁,
 (ii) *w*₂* is a minimal physical duplicate of *w*₂,
 (iii) *r* is a spatiotemporal region of *w*₁*,
 (iv) *s* is a spatiotemporal region of *w*₂*, and
 (v) *r* and *s* are intrinsically just alike in all physical respects,
 then *r* and *s* are just alike in *all* intrinsic respects, and
 (3) there are no brute inter-level relations of metaphysical necessitation linking physical particulars or properties to non-physical particulars or properties.

Perhaps clause (3) could itself be further explicated; I leave open whether this is so, and also whether it would be worthwhile seeking such an explication. Meanwhile, I propose (M₃) as an explication of the thesis of materialism. Since (M₃) entails both (M₂) and (M₁), my proposal treats each of the latter theses as a condition that is *necessary but not sufficient* for the truth of materialism.

2. MATTERS OF DEFENSE

I take it that the metaphysical doctrine of materialism is a broadly empirical hypothesis. Accordingly, its defense must be partly an empirical matter, involving the explanatory range and explanatory capacities of the various sciences – and physics in particular. Things look good in that respect. They already looked good when Oppenheim and Putnam’s classic paper on the unity of science was published (Oppenheim and Putnam, 1958), and they look better now.

But I take it that an adequate defense of materialism should incorporate another dimension too, beyond an appeal to the large-scale explanatory capacities of physics. This dimension may be called *ideological* – where by ‘ideology’ I mean theoretical inquiry into the workings of human concepts and human language. Although I regard ideology as a broadly empirical enterprise itself, much of the pertinent empirical data for it is ready to hand in the armchair – for instance, data about one’s own conceptual/semantic intuitions about how to correctly describe various thought-experimental scenarios (e.g., Putnam’s Twin Earth scenario, Gettier cases in epistemology, etc.). Ideological inquiry thus falls within the purview of philosophy (cf. Horgan, 1993a; Graham and Horgan, 1994; Henderson and Horgan, 2000, 2002). Philosophers, by dint of training and inclination, are good at theorizing about the workings of human concepts and language, in a way that draws heavily on pertinent armchair-accessible data.

How might matters of ideology, of a kind amenable to philosophical theorizing, figure in a defense of materialism?

2.1. *Ideology and cosmic hermeneutics*

When asking about the role that ideological factors should play in full-scale defense of materialism, it is useful to invoke a philosophical trope I introduced in Horgan (1984): the idea of a LaPlacean Demon of unbounded cognitive capacity who holds before her mind the complete physical history of the cosmos (as described in the language of a correct and completed physics), and whose task is to ascertain, on the basis of this physical history, all facts about the cosmos as described in any vocabulary whatever (or at any rate, any vocabulary employable by humans). The demon's task of reading off the cosmos's entire history from its physical history is what I called *cosmic hermeneutics*.

We should suppose that our LaPlacean Demon possesses a very rich conceptual/linguistic repertoire – at least as rich as the conceptual/linguistic repertoire of humans. Otherwise, she could hardly be expected to do cosmic hermeneutics; for, she would lack the capacity to deploy the very concepts and language in terms of which the pertinent facts about the cosmos are to be understood and expressed.

Thinking about how the Demon could accomplish her cosmic-hermeneutic task brings into focus the role of ideology vis-à-vis the metaphysical doctrine of materialism. If this doctrine is correct, then presumably the Demon already has all the resources she needs to carry out her task. Given her mastery of the non-physics-level concepts and linguistic expressions in her conceptual/linguistic repertoire, including her mastery of any pertinent facts about the workings of those concepts and linguistic expressions themselves (e.g., the fact that concept POISON is a functional concept, the concept of a kind of stuff that tends to cause illness and death), she can *logically deduce* the facts about the cosmos that are expressible in non-physics-level vocabulary from the conjunction of (1) the specification (in physics-level vocabulary) of the cosmos's physical history, and (2) the pertinent facts about the semantic workings of non-physics-level concepts and terms.

Suppose, however, that Moorean emergentism is true – say, with respect to the relations between physical properties and phenomenal-consciousness properties. And suppose that the LaPlacean Demon is ignorant of the pertinent, metaphysically brute, principles of metaphysical necessitation linking the two kinds of properties. Then she simply will not be able to tell, on the basis of her exhaustive knowledge of the cosmos's physical history together with her full mastery of the semantical workings of her own concepts and language (including her own *phenomenal* concepts and language), when and where (if ever) various phenomenal-consciousness properties are instantiated in the cosmos she is contemplating. She will not be able to tell, for instance, whether the humans in that cosmos are phenomenally conscious or instead are mere zombies – creatures whose behavioral-control mechanisms comprise physical-functional properties that lack phenomenal character altogether. The reason she will not be able to tell is that the instantiation of phenomenal properties depends upon certain facts beyond those that materialism countenances – viz., brute modal facts linking physical properties to phenomenal ones. These brute inter-level modal facts are neither (1) physics-level facts, nor (2) facts about the semantical workings of higher-level concepts and language, nor (3) facts that are logically/conceptually supervenient on facts of type (1) – that is, facts that follow logically from facts of types (1) and (2). But if materialism is right, then facts of types (1)–(3) should be all that are needed for the LaPlacean demon to be able to carry out the project of cosmic hermeneutics. (As I argued in section 1, materialism should eschew brute modal inter-level facts, and thus should eschew any putative Moorean-emergent properties.)

In short, if materialism is true, then all the facts about the cosmos expressible via non-physics-level vocabulary should follow logically from the conjunction of (1) the facts expressible in physics-level vocabulary, and (2) pertinent facts about the ideological workings of non-physics-level concepts and vocabulary. This, in turn, provides a general schematic template for the philosophical component of an adequate

overall defense of materialism. The philosophical task is this: Articulate and defend, at least in broad outline, a *materialism-friendly* theoretical account of the ideological workings of non-physics-level concepts and vocabulary – i.e., a theoretical account under which the conjunction of the physics-level facts and the ideological facts really does logically entail all the facts expressible in non-physics-level vocabulary.

How might the template be filled out?

2.2. *Location via conceptual analysis*

One potentially promising implementation of the template is familiar in recent philosophy, especially in philosophy of mind. The idea is that non-physics-level concepts have “conceptual analyses” in terms of concepts that are deemed physicalistically kosher – where the notion of *causation* is counted as physicalistically kosher. Functionalism in philosophy of mind is often so presented: mentalistic concepts are said to be analyzable as causal-role concepts – i.e., concepts of properties whose essence is their typical-cause relations to sensory inputs, behavioral outputs, and other causal-role concepts with the same kind of essence. An influential articulation of this idea, deriving from Frank Ramsey and developed by Lewis (1970, 1972), is to construe mentalistic concepts as functionally definable simultaneously, via the Ramsey sentence of a pertinent psychological theory. (On one version, that theory is common-sense psychology; this yields so-called analytical functionalism. On another version, it is the ideally complete empirical psychological theory true of humans; this yields so-called psychofunctionalism; cf. Block 1980.)

A natural idea is to generalize this approach beyond philosophy of mind. The thought is that *all* non-physics-level concepts in terms of which true claims can be made are functional-role concepts, possessing physicalistically kosher conceptual analyses. Two philosophers who have recently articulated and defended such a picture are Jackson (1998) and Melnyk (2003). Two philosophers who have articulated and defended a *qualified* version of the picture, in which phenomenal consciousness

is left to one side, are Chalmers (1996) and Kim (1998, 2005). (Chalmers and Kim both think that the recalcitrance of phenomenal consciousness makes trouble for full-fledged metaphysical materialism – a matter to which I return in section 3.)

Such conceptual analyses, perhaps couched in terms of Ramsey sentences of various common-sense and/or special-science theories, would provide the ideological underpinnings for empirical solutions to what Jackson (1998) calls *location problems*. The task is to locate, in the cosmos, materialistically kosher *referents* for the non-physics-level concepts in terms of which we form judgments and statements that we hold true. In the case of mentalistic concepts, for instance, the referent-properties would be either certain physical brain-state properties (if one embraces a first-order functionalist account of mentalistic concepts, as did Lewis, 1966, 1972, 1980 and as does Armstrong, 1968, 1970), or else certain functional properties (if one embraces an account of mentalistic concepts under which they denote second-order role properties rather than first-order filler properties). The facts of physics combine with the (functionalist) ideological workings of non-physics-level concepts, to secure certain physicalistically kosher properties as the referents of those concepts.

I myself have serious doubts about the viability of this conceptual-analysis strategy as a basis for defending materialism. One worry can be put in terms of a dilemma. A conceptual analysis should be framed either (1) solely in terms of common-sense platitudes (by analogy with analytical functionalism in philosophy of mind), or else (2) in a way that incorporates substantive empirical theories in addition to common-sense platitudes (by analogy with psychofunctionalism in philosophy of mind). If one sticks to mere common-sense platitudes, however, then it is unlikely that there will be conceptual analyses that are rich enough and specific enough to “pin down,” in conjunction with pertinent physical facts, unique referent-items for non-physics-level concepts. On the other hand, if one incorporates substantive empirical theories into one’s putative conceptual analyses, then one will be guilty of *conceptual chauvinism* – ruling out certain potential

referent-items as not falling under the scope of one's concepts that really *do* fall under their scope.

The second horn, in my view, already arises in philosophy of mind itself. Psychofunctionalism is guilty of conceptual chauvinism. Suppose, for instance, that there are non-human creatures elsewhere in the cosmos who undergo internal states that (a) conform to the principles of common-sense psychology, but (b) do not conform to laws of the ideally complete *empirical* psychological theory that is true of humans. According to psychofunctionalism, these creatures do not instantiate beliefs, desires, intentions, or any other states of the kind picked out by human mentalistic concepts. But surely the right thing to say – the conceptually non-chauvinistic thing to say – is that the creatures *do* undergo such mental states (even though their cognitive architecture conforms to empirical psychological laws somewhat different than the psychological laws that govern human cognitive architecture).

But the conceptual-analysis strategy may not be the only game in town anyway, insofar as one is seeking out a way to implement the template described in section 2.1 for materialism-friendly ideology. Let me set forth some ideas that could figure, at least partially, in an alternative potential implementation.

2.3. *Ontological austerity via indirect correspondence*³

Thinking about materialism-friendly ideology with an eye toward generating solutions to location problems rests on a strong presupposition: viz., that singular, predicative, and quantificational constituents of true thoughts and statements *denote items in the correct ontology*. If one is also a materialist, then one also claims that the denoted items are materialistically kosher items.

But another way to approach the ideology of non-physics-level concepts and language is to call this presupposition into question – at least for many, if not all, of the pertinent concepts and discourse. This is a strategy of *ontological austerity*: the idea is that numerous thoughts/statements employing non-physics-level concepts and terms possess both of these

two features: (a) they are indeed true, and yet (b) they are not ontologically committed to individuals, properties, or kinds that would answer to their singular, predicative, or quantificational constituents.

Consider, for instance, claims like the following, about various kinds of socially constructed entities:

- (1) The State of Arizona has three public universities.
- (2) The University of Arizona has a medical school.
- (3) NATO conducted a massive bombing campaign against Serbia.

All three claims are certainly true. But does the right ontology contain items – materialistically respectable or otherwise – answering to singular expressions like ‘The State of Arizona’, ‘The University of Arizona’, ‘NATO’, or ‘Serbia’? Does it contain properties answering to predicative expressions like ‘is a public university’, ‘is a nation’, or ‘is a multi-national organization’? On just a moment’s reflection, it is very plausible that the answer to these questions is ‘No’. There is no unique physical object – or set of physical objects, or mereological sum of physical objects – that is a single best candidate for being the putative referent of the expression ‘The University of Arizona’. Rather, numerous candidate-items are equally qualified, with none of them being *best* qualified, and none of them really even being *well* qualified (since it would be a Rylean “category mistake” to try identifying a university with some mereological sum of concrete physical objects anyway). Nor is it plausible that there is some spooky *non-physical* object denoted by ‘The University of Arizona’ – not plausible from either a materialistic perspective or even from the perspective of some non-materialist metaphysical position such as Cartesian dualism. (The University of Arizona, after all, is not a conscious being – a minded entity.) Likewise, *mutatis mutandis*, for the other singular and predicative constituents in the above three statements. Although we routinely take it in stride that such statements are often true, we do not find ourselves taking seriously an ontology containing items denoted by the

singular, predicative, and quantificational constituents of such statements.

Fans of the conceptual-analysis strategy might be tempted at this point to propose “eliminative” conceptual analyses, at the level of whole statements – analyses that effectively provide paraphrases (“regimentations,” in Quine’s terminology) that construe the original singular, predicative, and quantificational idioms as dispensable shorthand. (An example might be a regimentation of ‘She has a charming smile’ as ‘She smiles charmingly’ – where the paraphrase eliminates any mention of the putative entity *her smile*). But the trouble is that for numerous true thoughts and statements, such eliminative conceptual analyses are not even remotely in sight. (The reader is invited to try doing it for statements (1)–(3) above, for example.)

I recommend a different implementation of the ontological-austerity strategy, an implementation that does not require eliminative conceptual analyses. One key aspect of the alternative implementation is a familiar conception of *truth conditions* for a thought/statement: take a given claim’s truth conditions to be constituted by a *range of possible worlds* – or, for at least some thoughts/statements, a range of *centered* possible worlds, where a centering is a designated location (construed as a location that might be occupied by the thinker/utterer of the given claim). Truth, for the given claim, is constituted by the actual world’s *belonging* to this range of worlds.

The key point to appreciate about this familiar possible-worlds approach to truth conditions is that it says nothing whatever about whether the singular, predicative, or quantificational constituents of a given claim need to designate any genuine objects or properties within the (centered) worlds that collectively constitute the claim’s truth conditions. Thus, it is perfectly possible that numerous claims expressible in non-physics-level vocabulary possess both of the following features. (Take statements (1)–(3) above as plausible examples.) First, the given claim is true; i.e., the actual world

belongs to the range of possible worlds that constitutes the claim's truth conditions. Second, for many or all of these possible worlds – including the actual world – there is nothing in those worlds answering to certain of the claim's singular, predicative, or quantificational constituents.

The possible-worlds approach to truth conditions retains the familiar idea that truth is *correspondence* between language/thought and the world: truth is a matter of the cosmos's being one of the various potential ways that collectively constitute a given claim's truth conditions, rather than the cosmos's *not* being any of these truth-constituting ways that it might be. On the other hand, when the second condition lately mentioned obtains, truth is an *indirect* form of correspondence – in the sense that the given claim has certain singular, predicative, or quantificational constituents that do not pick out any object in the cosmos or any property instantiated in the cosmos. (Direct correspondence, by contrast, involves genuine referential linkages between such constituents on one hand, and objects or properties in the cosmos on the other hand.)

I believe that there is a strong philosophical case to be made for the contention that very large portions of human thought and discourse are normally governed by semantic standards under which truth is indirect correspondence. I have argued the case at some length elsewhere, sometimes collaboratively (cf. note 3). I will not attempt to reproduce that argumentation here, beyond what I have already said about claims like statements (1)–(3) above.

I believe too – as I think is plausible on its face – that the semantic standards that govern typical uses of non-physics-level concepts and language are largely responsive to *pragmatic* considerations, involving the point and purpose that such concepts and such language serve to further. (This too is argued at some length in the texts cited in note 3.) Thought and talk about socially constructed entities like nations and universities, and human social behavior that is influenced by such thought and talk, take place within a rich background of individual and

collective purposes and goals – purposes and goals that are furthered in myriad ways by conceptual/linguistic practices involving quantification over such socially constructed entities. Given such goals and purposes, it is very plausible that the truth conditions for claims that posit such entities – e.g., claims (1)–(3) above – are such that the none of the possible worlds (or centered possible worlds) that collectively constitute the given claim’s truth conditions contain items answering to the claim’s socially constructed singular or predicative constituents. (Likewise, none of these (centered) possible worlds contains items answering to the claim’s existential quantifiers whose substitution instances are such singular or predicative constituents.)

Does the ontological-austerity strategy, as tied to the idea of truth as indirect correspondence, provide a materialism-friendly treatment of the ideological workings of certain non-physics-level concepts and discourse, such as discourse about socially constructed entities? I.e., does it provide an implementation of the template for materialism-friendly ideology described in section 2.1 – an alternative to the implementation provided by the conceptual-analysis strategy? Well, partially. In order to harness the ontological-austerity strategy in defense of materialism, it seems we need the the following assumption: certain facts about human mentality – in particular, facts about people’s goals and purposes – are themselves materialistically kosher. This assumption needs to be in place because of the pragmatist element involved in indirect-correspondence semantic standards: since those standards arise within a rich background of individual and collective purposes and goals, the possible worlds in which such standards are operative had better be ones in which there are creatures who *possess* purposes and goals. Moreover, since the ontological-austerity strategy evidently *presupposes* the materialistic respectability of facts about human goals and purposes, this strategy evidently cannot itself be used to *establish* their materialistic respectability. That would need to be done some other way – perhaps via the conceptual-analysis strategy (e.g., via a functionalist account of the concept of a goal-state).

Suppose the assumption is true. What, then, will the physically possible worlds be like that collectively comprise the truth conditions for a claim like statement (1) above? Well, in none of these worlds are there any identifiable particulars, concrete or otherwise, answering to the singular term 'The State of Arizona' or answering to the predicate 'is a public university'. Nevertheless, in all of these worlds, things are so arranged physically, and things so transpire physically, that claim (1) is semantically correct in these worlds, under the indirect-correspondence standards fixed by the background human goals and purposes at work when concepts like 'state', 'nation', and 'university' are deployed. Likewise, *mutatis mutandis*, for all claims governed by indirect-correspondence semantic standards.

So, to the extent that the indirect-correspondence construal of truth can be made plausible for the ideology of a given form of thought/discourse – and assuming that human possession of goals and purposes can be somehow squared with materialism, perhaps via the conceptual-analysis strategy – the upshot is a materialism-friendly philosophical account of the ideology of such thought/discourse. Think again of the LaPlacean Demon, as she sets about the project of cosmic hermeneutics. For each statement *S* that is expressible in her non-physics-level vocabulary and is governed by indirect-correspondence semantic standards, she holds before her infinite mind the range *R* of possible worlds that collectively constitute *S*'s possible-world truth conditions. Within *R* are physically possible worlds that are minimal physical duplicates of themselves, and thus are completely characterizable *in physicalistic vocabulary*. She consults the cosmos's actual physical history, which she also holds before her infinite mind (and is also physicalistically formulated). (Given materialism, the actual world is a minimal physical duplicate of itself.) If the actual world is one of the worlds in *R*, the Demon assigns the semantic value *true* to *S*. Otherwise she assigns it the semantic value *not true*. She thereby accomplishes the task of cosmic hermeneutics for claims couched in vocabulary governed by indirect-correspondence

semantic standards – even though these claims do not, in general, have conceptual analyses via vocabulary that quantifies only over materialistically kosher entities.

2.4. *Other modes of defense?*

Might there be other materialism-friendly approaches to the ideology of non-physics-level concepts and vocabulary, other than the conceptual-analysis strategy and the ontological-austerity strategy? This is an important question. I know of no other approach that currently strikes me as both clearly coherent and seriously viable, but perhaps there is one waiting to be uncovered and articulated.⁴ (In section 3 I discuss a putative alternative approach advocated by some philosophers, which seems to me *not* coherent.)

Needless to say, mixing and matching of any available materialism-friendly approaches is a possibility. I have already suggested that the ontological-austerity strategy evidently needs supplementation by some other ideological approach to mentalistic notions like goal-possession.

Might it be possible to mount an adequate defense of materialism without any appeal to materialism-friendly treatments of the ideology of non-physics-level concepts and vocabulary? Perhaps so, but it seems to me that nobody has yet offered a cogent picture of how such a defense would go. After all, what seems wanted is a *unified* metaphysical picture, in which the truth of claims couched in non-physics-level vocabulary is shown to *fit* with a materialist ontology. It is hard to see how to accomplish the needed unification without appeal to some form of materialism-friendly ideology for higher-level concepts and discourse. Perhaps there are ways to suppress this itch for ideological unification, rather than trying to scratch it – but if so, they are yet to be articulated.

3. MATTERS OF DECONSTRUCTION

One especially prominent and persistent challenge to materialism concerns the phenomenal features of mentality – those

aspects of the mental such that there is something it is like, for the experiencing subject, to undergo them. One can put the worries in terms of a *prima facie* challenge to thesis (M₁) above – a thesis which I have claimed constitutes a necessary, but not sufficient, condition for the truth of materialism. There are a number of familiar ways to formulate the challenge, all of which are similar in spirit. For instance, it seems that no matter how much detail one might acquire about the physical workings and the functional architecture of the human brain and central nervous system, it would still be possible to coherently conceive a scenario in which there are creatures who are physically exactly like humans (undergoing internal states and processes that are physically and functionally exactly like those that occur in humans), but who nonetheless lack phenomenal consciousness altogether – i.e., they are *zombie duplicates* of actual humans. (Likewise, it would still be possible to coherently conceive a scenario in which human-like creatures are exactly like ourselves physically functionally, and yet undergo phenomenal states that are systematically inverted, along some dimension, relative to ours – e.g., color-experience-wise inverted.)

These conceivability facts are dialectically very powerful, it seems to me, because they call into question various materialism-friendly approaches to the ideological workings of mental concepts – specifically, phenomenal concepts. The point is familiar. If functionalism is the correct account of phenomenal concepts, for instance, then zombie-duplicate scenarios and phenomenal-inversion scenarios are not really conceptually possible at all: if one fully grasps the phenomenal concepts, one will appreciate that a physical-functional duplicate of oneself could not fail to be a *phenomenal* duplicate of oneself. But zombies and inverts are indeed robustly conceivable.⁵ So much the worse, therefore, for a functionalist account of the ideology of phenomenal concepts.

It is worth stressing that this challenge remains in force even against psychofunctionalist accounts of the ideology of phenomenal concepts (as opposed to analytic-functionalist

accounts). Admittedly, on the psychofunctionalist picture it is an empirical matter, not an a priori matter, what the correct functionalist characterization of phenomenal concepts is – since this characterization depends on what the correct empirical psychological theory of human cognition turns out to be. Nevertheless, the problem remains. For, since that empirical psychological theory is supposed to be articulable in terms of the cognitive-architectural *roles* played by various phenomenal mental states, it seems intuitively clear that no matter how detailed an empirical theory one might come to acquire about these cognitive architectural roles, one would still be able to robustly conceive a zombie-duplicate of oneself. Such a zombie would have internal physical states and processes that are physically/functionally just like one's own, and thus these internal physical states would entirely conform to the relevant, empirical, psychological theory. Yet the zombie would lack phenomenal consciousness altogether. In short: if a psychofunctionalist account of the ideology of phenomenal concepts is correct, then zombies would not be robustly inconceivable *given* the correct empirical psychological theory of human cognitive architecture; but zombies would still be robustly conceivable, even given such a theory; so a psychofunctionalist account of phenomenal concepts is mistaken.⁶

I myself find the robust conceivability of zombies and inverts a compelling and decisive consideration against functionalist accounts of the nature of phenomenal concepts. I think the materialist needs some other materialism-treatment of the ideological workings of these concepts.

Could one perhaps adopt the ontological-austerity strategy, and claim that phenomenal concepts do not pick out any real properties at all? To me this idea seems a non-starter, because nothing is more indubitably real than one's own phenomenal consciousness. The idea is a non-starter even though there are philosophers around (e.g., Dennett, 1988, 1991) who maintain, in effect, that we are all of us really zombies. Denying phenomenal consciousness does make it easier to be a materialist, but it does so at the cost of denying what is undeniable.

So what the materialist needs, it seems to me, is a different kind of materialism-friendly ideological treatment of phenomenal concepts. Phenomenal *properties* may yet turn out to be functional properties – or, instead, first-order physical properties. But phenomenal *concepts* should not be construed as functional concepts. Instead they should be construed in a way that renders them conceptually *independent* of physical and functional concepts. They also should be construed in a way that fully accommodates the robust conceivability of zombies and inverts, and does so in a way that is consistent with the metaphysical impossibility of such scenarios. And they should be construed in such a way that zombies – who on this approach are indeed robustly conceivable even though not metaphysically possible – would not possess, and self-apply, genuine phenomenal concepts.

This is a very tall order. Nonetheless, there are a number of contemporary philosophers – notably Hill (1997), Loar (1997), Hill and McLaughlin (1999), and McLaughlin (2001) who claim to be able to fill the order. Tienson and I have dubbed this approach “new wave materialism” in Horgan and Tienson (2001). Three key ideas are involved. First, phenomenal concepts are a species of so-called “recognitional” concepts: their functional role in human cognitive economy is to enable the cognitive subject to self-ascribe certain internal states on the basis of undergoing those states.⁷ Second, genuine phenomenal concepts are not presentationally blank – as would be the corresponding recognitional concepts of zombies, whose self-ascriptions would be like those of a self-ascribing super-blindsighter. (Blindsighters lack visual experience but process retinal information subliminally, and thus score better than chance when they are asked to say what kinds of objects are in front of them.) Third, genuine phenomenal concepts operate via phenomenal “modes of presentation”: the modes of presentation are the phenomenal properties themselves, as currently instantiated in the experiencing subject.

In effect, the new wave materialists are proposing yet another implementation of the generic template for defending

materialism that was described in section 2.1 above. They are offering an ideological account of phenomenal concepts that is very different from a functionalist ideological account – even though their ideological treatment of phenomenal *concepts* is consistent with the claim that phenomenal *properties* are indeed functional properties. Their account, I take it, is not supposed to be a conceptual analysis of phenomenal concepts. Rather, it is an account that triangulates these concepts in terms of the distinctive role they play in human cognitive economy – a non-blindsighter-like recognitional role, vis-à-vis certain internal physical or functional properties. These properties themselves are identical to phenomenal properties, according to the new wavers. And, because phenomenal concepts play a role in human cognitive economy that is independent of the roles played by physical/functional concepts, zombies are robustly conceivable. Nevertheless, zombies are not metaphysically possible, because phenomenal concepts refer – and refer rigidly – to the very same properties referred to by certain physical or functional concepts.

But on close inspection, unfortunately, this story appears to be incoherent. Here I will summarize the problem very briefly. (For a more extended presentation, see Horgan and Tienson (2001); for a reply by a new waver, see McLaughlin, 2001.) Consider the following argument. (Following Loar, 1997, the term ‘physical-functional’ properties is used in order to remain neutral among different specific versions of new-wave-ism.)

3.1. *Deconstructive argument*

1. When a phenomenal property is conceived under a phenomenal concept, this property is conceived otherwise than as a physical-functional property.
2. When a phenomenal property is conceived under a phenomenal concept, this property is conceived directly, as it is in itself.
3. If (i) a property P is conceived under a concept C, otherwise than as a physical-functional property, and (ii) P is

conceived, under C, as it is in itself, then P is not a physical-functional property.

Hence,

4. Phenomenal properties are not physical-functional properties.⁸

The argument is valid, and the new wave materialists are committed to premises 1 and 2. (Note well that premise 1 does not say that phenomenal properties are conceived, under phenomenal concepts, as non-physical-functional properties. Conceiving a property *otherwise than as* a physical-functional property is different from, and weaker than, conceiving it *as otherwise than* a physical-functional property.) Yet premise 3 looks very hard to deny; on the contrary, it seems virtually tautologous, given that the pertinent form of direct conceiving is supposed to be not presentationally blank (and thus *nakedly* referential, as in the case of an introspective “super-blindsighter”) but rather is supposed to work via the phenomenal property P as a self-presenting *mode of presentation*. If indeed phenomenal properties, when conceived under phenomenal concepts, not only are conceived otherwise than as physical-functional properties, but also are conceived under a mode of presentation that self-presents them *as they are intrinsically*, then how could these properties fail to *be* otherwise than physical-functional? I.e., how could it be that they are not physical-functional properties?

Since the deconstructive argument is valid, the new wavers are obliged to reject at least one premise. And since they are committed to premises 1 and 2, they are obliged to reject premise 3. Furthermore, they acknowledge three explanatory tasks that a credible version of materialism should simultaneously accomplish:

- (A) Explain the differences between phenomenal concepts and associated physical-functional concepts in a way that renders them conceptually independent, and thereby renders separability scenarios (e.g., scenarios involving

creatures physically just like humans who are zombies or whose qualia are inverted) coherently conceivable.

- (B) Explain the differences between phenomenal concepts and associated physical-functional concepts in a way that fully respects the phenomenology of conscious experience.
- (C) Explain the differences between phenomenal concepts and associated physical-functional concepts in a way that is consistent with the claim that phenomenal properties are identical to physical-functional properties.

But the problem is to see how one could deny premise 3 while still simultaneously meeting all three explanatory tasks. The only way to clearly meet tasks A and C is to deflate the idea of conceiving a property *directly* under a phenomenal concept – and thereby to deflate the idea of a phenomenal property functioning as a self-presenting *mode of presentation* – to the point where these notions (as thus deflationally construed) would be applicable to zombies whose experience is phenomenally empty. And the cost of such deflation, of course, is a failure to meet task B.⁹

So at present I see no cogent way for new wave materialism to avoid the deconstructive argument. And, for reasons explained early in the present section, it also appears to me that there is no other presently viable materialism-friendly ideological treatment of phenomenal concepts; there is none that successfully applies the conceptual-analysis strategy, there is none that successfully applies the ontological-austerity strategy, and there are no other clearly coherent, clearly viable, strategies presently in sight for doing materialism-friendly ideology. So phenomenal consciousness remains a very big challenge to materialism.¹⁰ I myself still want to embrace materialism, but I don't know what an adequate materialist account of phenomenal consciousness would look like.¹¹

NOTES

¹ Ancestors of some ideas expressed in this section can be found in Horgan (1978, 1981, 1982, 1984, 1987, 1993b, c).

² If one construes the semantics of counterfactuals in such a way that the possible worlds pertinent to evaluating ordinary counterfactuals can be worlds in which minor deviations from the actual physical laws occur – what Lewis (1979) called “divergence miracles” – then the scope of physically possible worlds will need to include these kinds of worlds too. But I will ignore this complication, for simplicity.

³ Some of the ideas briefly sketched in this section are developed more fully in Barnard and Horgan (in press), Horgan (1986a, b, 1991, 2001), Horgan and Potrc (in press, forthcoming).

⁴ In Dowell (manuscript), an approach is proposed that differs in significant ways from any I discuss in this paper. At this writing I am just beginning to think about Dowell’s proposal.

⁵ When I say *robustly* conceivable I mean that the conceivability persists no matter how much detail one fills in with respect to matters physical-functional. In a non-robust way one can conceive someone proving the completeness of elementary number theory. But one certainly can’t conceive it robustly, since the details *cannot* be filled in.

⁶ It is worth noting how different this is from how things work with natural-kind concepts like the concept *water*. Given that water is composed of H₂O molecules, can one robustly conceive a scenario in which some stuff is not composed of H₂O molecules but nevertheless is water? Arguably, no. Cf. Kripke (1980), who argues in effect that all one can really conceive is a scenario in which there is some stuff *other than water* that has the various contingent features by which the concept of water has had its reference fixed.

⁷ It bears emphasis that individuating recognitional concepts via their functional roles in cognitive economy is a very different matter from proposing a *functionalist conceptual analysis* of these concepts. New wave materialists are not doing the latter at all.

⁸ This argument is similar in spirit to the “property dualism argument” presented in White (1986, pp. 351–353).

⁹ The new wave response to the deconstructive argument offered by McLaughlin (2001) seems to me to encounter this fate. It is telling that McLaughlin says nothing about why or how the new wave account would fail to apply to zombies who recognitionally conceive their own physical-functional states in a phenomenally empty “super-blindsighter” manner. By contrast, Loar (1997) explicitly seeks to articulate the new wave position in a way that excludes introspective “super-blindsighters”; but, as is argued in Horgan and Tienson (2001), Loar does so at the cost of losing any clear way to simultaneously meet explanatory demands A and C.

¹⁰ In my view this challenge is very broad in scope, because the most fundamental kind of mental intentionality is fully constituted by phenomenology. Cf. Horgan and Tienson (2002), Horgan et al. (2004).

¹¹ My thanks to Uriah Kriegel for helpful comments and suggestions.

REFERENCES

- Armstrong, D.M. (1968): *A Materialist Theory of Mind*, Macmillan.
- Armstrong, D.M. (1970): 'The Nature of Mind', in C.V. Borst (ed.), *The Mind/Brain Identity Theory*, Macmillan.
- Barnard, R. and Horgan, T. (in press): Truth as Mediated Correspondence, *The Monist*.
- Block, N. (1980): 'What is Functionalism?' in N. Block (ed.), *Readings in the Philosophy of Psychology*, Vol. 1, Harvard University Press.
- Chalmers, D. (1996): *The Conscious Mind*, Oxford University Press.
- Dennett, D. (1988): 'Quining Qualia', in A.J. Marcel and E. Bisiach (eds.), *Consciousness and Contemporary Science*, Oxford University Press.
- Dennett, D. (1991): *Consciousness Explained*, Little Brown.
- Dowell, J.L. (manuscript). Serious Metaphysics and the Vindication of Reductions. <http://www.bgsu.edu/departments/phil/faculty/dowell/dowell.html>.
- Graham, G. and Horgan, T. (1994): 'Southern Fundamentalism and the End of Philosophy', *Philosophical Issues* 5, 219–247. Reprinted in M. DePaul and W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophy*, Rowman and Littlefield, 1998.
- Henderson, D. and Horgan, T. (2000): 'What Is A Priori and What Is It Good For?' *Southern Journal of Philosophy* 28, Spindel Conference Supplement on the Role of the Empirical and the A Priori in Philosophy, 51–86.
- Henderson, D. and Horgan, T. (2002): 'The A Priori Isn't All That It's Cracked Up to Be, But It Is Something', *Philosophical Topics* 29, 219–250.
- Hill, C. (1997): 'Imaginability, Conceivability, Possibility, and the Mind-Body Problem', *Philosophical Studies* 87, 61–85.
- Hill, C. and McLaughlin, B. (1999): 'There are Fewer Things Than Are Dreamt of in Chalmers's Philosophy', *Philosophy and Phenomenological Research* 59, 445–454.
- Horgan, T. (1978): 'Supervenient Bridge Laws', *Philosophy of Science* 45, 227–249.
- Horgan, T. (1981): 'Token Physicalism, Supervenience, and the Generality of Physics', *Synthese* 49, 395–413.
- Horgan, T. (1982): 'Supervenience and Microphysics', *Pacific Philosophical Quarterly* 63, 29–43.
- Horgan T. (1984): 'Supervenience and Cosmic Hermeneutics', *Southern Journal of Philosophy* 22, Spindel Conference Supplement on Supervenience, 19–38.
- Horgan, T. (1986a): 'Psychologism, Semantics, and Ontology', *Nous* 20, 21–31.
- Horgan, T. (1986b): 'Truth and Ontology', *Philosophical Papers* 15, 1–21.
- Horgan, T. (1987): 'Supervenient Qualia', *Philosophical Review* 96, 491–520.

- Horgan, T. (1991): 'Metaphysical Realism and Psychologistic Semantics', *Erkenntnis* 34, 297–322.
- Horgan, T. (1993a): 'The Austere Ideology of Folk Psychology', *Mind and Language* 8, 282–297.
- Horgan, T. (1993b): 'Nonreductive Materialism and the Explanatory Autonomy of Psychology', in S. Wagner and R. Warner (eds.), *Naturalism: A Critical Appraisal*, Notre Dame University Press.
- Horgan, T. (1993c): 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World', *Mind* 102, 555–586.
- Horgan, T. (2001): 'Contextual Semantics and Metaphysical Realism: Truth as Indirect Correspondence', in M. Lynch (ed.), *The Nature of Truth: Classic and Contemporary Perspectives*, MIT Press.
- Horgan, T. and Potrc, M. (in press): 'Abundant Truth in an Austere World', in P. Greenough and M. Lynch (eds.), *Truth and Realism: New Debates*, Oxford University Press.
- Horgan, T. and Potrc, M. (forthcoming): *Austere Realism*.
- Horgan, T. and Tienson, J. (2001): 'Deconstructing New Wave Materialism', in C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*, Cambridge University Press.
- Horgan, T. and Tienson, J. (2002): 'The Intentionality of Phenomenology and the Phenomenology of Intentionality', in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford.
- Horgan, T., Tienson, J., and Graham, G. (2004): 'Phenomenal Intentionality and the Brain in a Vat', in R. Schantz (ed.), *The Externalist Challenge*, Walter de Gruyter.
- Jackson, F. (1998): *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford University Press.
- Kim, J. (1998): *Mind in a Physical World*, MIT Press.
- Kim, J. (2005): *Physicalism, or Something Near Enough*, Princeton University Press.
- Kripke, S. (1980): *Naming and Necessity*, Harvard University Press.
- Lewis, D. (1966): 'An Argument for the Identity Theory', *Journal of Philosophy* 63, 17–25.
- Lewis, D. (1970): 'How to Define Theoretical Terms', *Journal of Philosophy* 70, 427–446.
- Lewis, D. (1972): 'Psychophysical and Theoretical Identifications', *Australian Journal of Philosophy* 50, 249–258.
- Lewis, D. (1979): 'Counterfactual Dependence and Times Arrow', *Nous* 13, 455–476.
- Lewis, D. (1980): 'Mad Pain and Martian Pain', in N. Block (ed.), *Readings in the Philosophy of Psychology*, Vol. 1, Harvard University Press.
- Loar, B. (1997): 'Phenomenal States', in N. Block, O. Flanagan and G. Guzeldere (eds.), *The Nature of Consciousness*, MIT Press.

- McLaughlin, B. (2001): 'In Defense of New Wave Materialism: A Response to Horgan and Tienson', in C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*, Cambridge University Press.
- Melnyk, A. (2003): *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge University Press.
- Moore, G.E. (1903): *Principia Ethica*, Cambridge University Press.
- Moore, G.E. (1922): 'The Conception of Intrinsic Value', in *Philosophical Studies*, Harcourt, Brace, and Co.
- Oppenheim P. and Putnam, H. (1958): 'Unity of Science as a Working Hypothesis' in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Vol. 2, University of Minnesota Press.
- White, S. (1986): 'Curse of the Qualia', *Synthese* 68, 333–368.

Department of Philosophy
University of Arizona
Tucson, AZ 85721-0027
USA
E-mail: thorgan@email.arizona.edu