COSMIC HERMENEUTICS*

Alex Byrne
MIT

1 Introduction

1.1 Cosmic hermeneutics

Imagine a Laplacian demon. He knows everything that is expressed by true sentences of some appropriate physical language—every physical fact.\(^1\) He knows everything that is a priori. Let \(\psi\) be a true sentence composed from any vocabulary whatsoever.\(^2\) Using only the knowledge just stipulated as premises, can the demon deductively infer, and thereby come to know, (the proposition expressed by\(^3\)) \(\psi\)? Borrowing Terry Horgan’s delightful neologism, is cosmic hermeneutics possible?\(^4\) That is the topic of this paper.

Put this baldly, one might wonder if anyone actually thinks it is possible. In fact—waiving certain qualifications for dramatic effect—David Chalmers, Frank Jackson and David Lewis all think it is.\(^5\) Concentrating on some recent arguments of Chalmers and Jackson, I shall be arguing that there is no reason to suppose that cosmic hermeneutics is possible.

Cosmic hermeneutics is possible iff, for every true \(\psi\), there is some true physical sentence \(\phi\) such that \(\phi \supset \psi\) is knowable a priori. (Left to right: suppose that the demon can deductively infer \(\psi\) from a true physical premise \(\phi\) and an a priori premise \(\sigma\) (we can always conjoin multiple physical or a priori premises into one big physical or a priori premise). Then \(\sigma \supset (\phi \supset \psi)\) is a priori. So \(\sigma \supset (\phi \supset \psi)\) is a priori. But as \(\sigma\) is a priori, so is \(\phi \supset \psi\). Right to left: suppose \(\phi \supset \psi\) is a priori. Then the demon knows \(\phi \supset \psi\), and since he knows \(\phi\), he can deductively infer \(\psi\).\(^6\)

Setting aside the contingent a priori as a rare special case\(^7\), it follows that cosmic hermeneutics is possible only if, for every true \(\psi\), there is a true physical sentence \(\phi\) such that \(\phi \supset \psi\) is (metaphysically) necessary. (Henceforth the variable ‘\(\phi\)’ will signal that we are quantifying over (consistent) sentences in some suitable physical vocabulary.)

This necessary condition for the possibility of cosmic hermeneutics, that every truth is metaphysically determined by some physical truth, is, near enough,
the contemporary thesis of physicalism. But now we must face an irritating complication.

1.2 Physicalism

Physicalism, put loosely, is the claim that the facts are “nothing over and above” the physical facts. It is usually put more precisely as a global supervenience thesis. The simplest version that might be thought to do the trick is:

(P) Any physical duplicate of our world is a duplicate simpliciter of our world.

The more natural reading has (P) being true at a world w iff all physical duplicates of @ (the actual world) are duplicates simpliciter. This reading makes (P), if true, necessarily so. Physicalists, however, have traditionally supposed their thesis to be contingently true. We can accommodate this by taking ‘our world’, evaluated at world w, to refer to w. So interpreted, (P) is true at a world w iff all physical duplicates of w are duplicates simpliciter of w.

In fact, (P) is equivalent to the necessary condition for cosmic hermeneutics just discussed—that every truth is metaphysically determined by some physical truth. (Left to right: we may assume that there is a physical sentence \( \phi_{w^*} \) that is true at a world w iff w is a physical duplicate of world w*. Suppose (P) is true at w*. Then for every true \( \psi \), \( \phi_{w^*} \supset \psi \) is necessary. Right to left: suppose that, for every \( \psi \) true at world w* there is a \( \phi \) true at w* such that \( \phi \supset \psi \) is necessary. Now a world is a physical duplicate of w* iff the same physical sentences are true at both; and a world is a duplicate simpliciter of w* iff the same sentences are true at both.\(^8\) Let w be a physical duplicate of w* and \( \psi \) be a sentence true at w*. Then there is a \( \phi \) true at w* such that \( \phi \supset \psi \) is necessary, and so \( \psi \) is true at w. Now let \( \psi' \) be a sentence true at w; if \( \psi' \) is false at w*, it’s false at w, so it’s true at w*. Thus the same sentences are true at w and w*, as required.)

(P), unfortunately, is too strong, for many physicalists will accept that some physical duplicate of this world has items that do not interact with anything physical—inmaterial spirits, as it might be—and that our world lacks.

For illustration suppose that, at time t, one and no more than one politician is in pain, and one and no more than one White House “coffee” is occurring. Then if (P) is true, there are physical sentences \( \phi_1, \phi_2, \phi_3, \phi_4 \) such that:

(i) \( \phi_1 \supset \text{a politician is in pain at } t' \).
(ii) \( \phi_2 \supset \text{a White House coffee is occurring at } t' \).
(iii) \( \phi_3 \supset \text{exactly one politician is in pain at } t' \).
(iv) \( \phi_4 \supset \text{exactly one White House coffee is occurring at } t' \).

are all necessary.
Now the problem just raised does not affect the physicalist’s commitment to necessary truths of the form (i) and (ii). No amount of epiphenomenal ectoplasm added to matters physical will prevent a politician from being in pain, or a White House coffee occurring—at least according to physicalism. But it might create some immaterial politician in pain, or a ghostly coffee. For this reason, (many) physicalists will deny that any sentence of the form (iii) or (iv) is necessary.

A number of ways of patching this difficulty with (P) have been proposed\(^9\); let us adopt Frank Jackson’s, and formulate physicalism thus:

\[(P^-) \text{ Any minimal physical duplicate of our world is a duplicate simpliciter of our world.}\]

A minimal physical duplicate of our world “is a world that (a) is exactly like our world in every physical respect...and (b) contains nothing else in the sense of nothing more by way of kinds or particulars than it must to satisfy (a)” (Jackson 1998a, 13; see also 1994a, 28–9, 1994c, 485). For the reasons given earlier, (P\(^-\)) should be understood as contingent, which is Jackson’s intent (1998a, 12): (P\(^-\)) is true at a world \(w\) iff any minimal physical duplicate of \(w\) is a duplicate simpliciter of \(w\).

Now, if nothing stronger than (P\(^-\)) is true, cosmic hermeneutics, at any rate as originally explained, is not possible: the demon will not be able to infer, for example, that exactly one politician is in pain at \(t\), for there is no appropriate necessarily true conditional. Can this barrier be removed, with only a minor adjustment to the demon’s initial stock of knowledge?

Yes. For we can simply allow the demon to know (P\(^-\)) (assuming, of course, that (P\(^-\)) is true). Letting ‘\(\pi\)’ abbreviate the sentence (displayed above) that expresses (P\(^-\)), and \(\phi@\) be a sentence true at a world \(w\) iff \(w\) is a physical duplicate of the actual world, any physicalist will regard \(\neg(\phi@ \land \pi)\) \(\supset \) exactly one politician is in pain at \(t\) as necessarily true. For consider a world \(w\) at which \(\phi@\) is true and the consequent is false: a world just like ours physically but in which some immaterial politician is in pain at \(t\). Given that (P\(^-\)) is contingent, ‘\(\pi\)’ is false at \(w\), and so the conditional is true at \(w\).

Let us conveniently stipulate that ‘\(\pi\)’ is a physical sentence\(^{10}\), and amend the task of cosmic hermeneutics by adding (P\(^-\)) to the demon’s initial stock of knowledge (so if (P\(^-\)) is false, then cosmic hermeneutics is not possible). From this stipulation, amendment, and what we established earlier, it follows that cosmic hermeneutics is possible iff physicalism is true and, for every true \(\psi\), there is a true \(\phi\) such that \(\neg(\phi \supset \psi)\) is necessary a priori.

The conclusion of the paper can now be put more specifically: there is no reason to suppose that cosmic hermeneutics is possible, even if physicalism is true.

1.3 The significance of our question

This is best brought out by giving a series of examples.
Example 1: Black-and-white Mary

Frank Jackson’s superscientist Mary, who learns every physical fact in a monochromatic environment, is, of course, our demon in another guise. Let ‘red-feeling’ denote the phenomenal character distinctive of visual experiences of ripe tomatoes, strawberries, maraschino cherries, and the like (this is a piece of reference-fixing, not synonymy-supplying). Then ‘some visual experiences are red-feeling’ is true. And if physicalism is true, then there is some physical sentence $\phi$ such that $\phi \supset$ some visual experiences are red-feeling’ is necessary. Mary knows $\phi$. She does not want for a priori knowledge. Can she thereby know that some visual experiences are red-feeling?

According to Jackson, at any rate at the time of “Epiphenomenal Qualia” (1982), the answer is no.\textsuperscript{11} When Mary is released from her black-and-white cell, and sees a ripe tomato for the first time, she will come to learn that some visual experiences are red-feeling. That is something she could not have known beforehand.\textsuperscript{12}

Jackson drew the conclusion that physicalism was false. To this it was objected that the knowledge argument only shows (at best) that no conditional of the form $\phi \supset$ some visual experiences are red-feeling’ is a priori, and that is consistent with some such conditional being necessary, as physicalism requires.\textsuperscript{13} Although that reply is fine as an opening move, charity demands we take Jackson to be tacitly assuming that, if physicalism is true, cosmic hermeneutics is possible (which, indeed, he was: see Jackson 1994b).

Example 2: Morality

(For the purposes of this example, assume that moral sentences have truth values.) On one interpretation of ‘you can’t derive an ought from an is’, no conditional of the form $\delta \supset \mu$ is a priori, where $\delta$ and $\mu$ are, respectively, descriptive and moral (contingent a posteriori) sentences.\textsuperscript{14} Almost universally held is that the moral supervenes on the descriptive, from which it follows that some such conditionals are necessarily true (of course the necessity here is metaphysical, not merely nomological). The conjunction of ‘no ought from an is’ and supervenience would seem to be eminently defensible.\textsuperscript{15} But if cosmic hermeneutics is possible, it is mistaken.

Example 3: Kripke’s Wittgenstein

Kripke’s Wittgenstein purports to show that “[t]here can be no fact as to what I mean by ‘plus’, or any word at any time” (Kripke 1982, 21). Some have argued that the sceptical argument only shows, at best, something epistemological, namely that semantic/intentional facts—that Jones means addition by ‘+’ , that Jones intends to add, etc.—are not a priori consequences of non-semantic/non-intentional facts.\textsuperscript{16} So, this line of thought continues, the sceptical argument is no threat to the
modest claim that intentional/semantic facts supervene on the physical, and thus Kripke’s Wittgenstein’s apparent irrealism about meaning (and intentionality) may be resisted. But if cosmic hermeneutics is possible if physicalism is true, the sceptical argument is reinstated.

**Example 4: A posteriori necessity**

Suppose that physicalism is true, and that cosmic hermeneutics is not possible with respect to many facts, for example psychological ones. Then we have a class of a posteriori necessities (of the form 'ϕ ⊃ Alfred believes that snow is white’, etc.) that appear to be quite different from those that Kripke famously drew to our attention.

The second example helps bring out the point that physicalism is best thought of as being a prominent test case in the discussion to follow. Let fundamentalism be the view that everything supervenes (with metaphysical necessity) on some (interestingly proper) portion of everything. Physicalism is, of course, a variety of fundamentalism. So is descriptivism: everything supervenes on the descriptive. So is Humean supervenience: everything supervenes on “local matters of particular fact” (Lewis 1986b, ix). And so is phenomeno-physicalism: everything supervenes on the physical and the phenomenal (Chalmers 1996, 72). Fundamentalism is a highly popular position: probably the only dissenters would be those sceptical of metaphysical necessity.17 For each variety of fundamentalism, we can ask whether a demon equipped with knowledge of the appropriate supervenience base could deductively infer everything. Although the considerations pro and con will not be quite the same in each case, there will be considerable overlap. The conclusion of this paper could, I think, be readily extended to cover fundamentalism in general. But I shall not take the space to do that.

*1.4 Further clarification*

In the following section I turn to the case for cosmic hermeneutics. This section ties a few loose ends.

First, the Laplacian demon’s initial stock of a posteriori knowledge is specified in a “physical language”—but what is that, exactly? Fortunately for present purposes the details won’t much matter. It will suffice to think of it as the language of contemporary physics, or perhaps of a future physics “somewhat improved” (Lewis 1983, 361; see also Chalmers 1996, 33, and Jackson 1998a, 6–8). All physicalists would agree that the physical supervenience base can be specified in something like a language of this sort. (Having said that, for the sake of some later examples, one of which will be touched on in the next paragraph, I shall augment the physical language with expressions from chemistry and geology.)

Second, something should be said at the start about the examples Kripke gave of the necessary a posteriori, for they are the obvious first candidates for counterexamples to the claim that cosmic hermeneutics is possible. Take, for
instance, the fact that water covers most of the Earth. The following argument is
tempting. If the demon can know that water covers most of the Earth, he must be
able deductively to infer this from the fact that H$_2$O covers most of the Earth (if
he can’t do that, how else could he come to know that water covers most of the
Earth?). But if he can perform such a deduction, the necessary conditional ‘H$_2$O
covers most of the Earth $\supset$ water covers most of the Earth’ must be a priori,
whereas in fact it is a posteriori. So the demon cannot know that water covers
most of the Earth, and hence cosmic hermeneutics is not possible.

This problem will take center stage when we examine Jackson’s arguments
(2.4 below). I mention it now to stress that only needless complexity will come of
raising Kripkean examples of the necessary a posteriori before then; so until we
arrive at 2.4, forget them.$^{18}$

2 The case for cosmic hermeneutics

First, I briefly examine how far conceptual analysis might take us towards
cosmic hermeneutics. Second, I turn to two arguments for the possibility of cosmic
hermeneutics, both due to David Chalmers.$^{19}$ Finally, I investigate at some length
Frank Jackson’s argument for the (weaker) conclusion that if physicalism is true,
cosmic hermeneutics is possible.

2.1 Conceptual analysis and cosmic hermeneutics

One straightforward way of showing that $\psi$ can be deductively inferred from
physical facts is to give a conceptual analysis of that sentence using only physical
vocabulary. For if the analysans is $\phi$, then $\phi \supset \psi$ is a priori, as required.

However, although we cannot be sure of very much as far as the future of
philosophy goes, we can be quite sure that successful conceptual analyses, let
alone physicalistically acceptable ones, will remain almost as rare as an uncon-
troversial philosophical argument.$^{20}$ Jackson, discussing an example of Stephen
Stich’s, concedes that “Stich is right that we cannot write down necessary and
sufficient conditions for an animal displaying grooming behaviour in austerely
physical terms” (1998a, 62). And in general the proponents of cosmic hermeneu-
tics do not pretend to supply conceptual analyses, at least not physically accept-
able ones.

Yet all is not lost. For suppose we had a theory of analysis itself, that
told us that in so-and-so circumstances, $A$-vocabulary can be analysed in terms
of $B$-vocabulary. Then maybe we could apply the theory to particular cases of
non-physical vocabulary to show that they can be analysed in a physicalist-
ically acceptable way, whether we know how to do this or not. If so, we
would have argued “non-constructively” for certain physically acceptable
analyses.

There is only one (partial) theory of analysis that might deliver this happy
result, and that is David Lewis’s elegant account of the meanings of “theoretical
terms” (1970).
Somewhat simplified, Lewis’s central idea is as follows. Sometimes new words are introduced into the language without being explicitly defined beforehand. Suppose, for illustration, that some engineer introduced the words ‘nut’ and ‘bolt’, not by explicitly telling us what they meant, but by uttering the following sentences:

Bolts are rods with a screw thread at one end. Every bolt has a nut that screws onto it. Every nut screws onto the end of some bolt.

We can regard these sentences as comprising the “theory” of nuts-and-bolts. This theory implicitly specifies what we might call the ‘nut-role’: the property of being the first member of a unique pair of kinds \( \langle N, B \rangle \) such that if we called any instance of \( N \) ‘a nut’, and any instance of \( B \) ‘a bolt’, then (with ‘nut’ and ‘bolt’ interpreted this way) the nuts-and-bolts theory would be true. The idea is that ‘nut’ means, roughly speaking: thing that has the nut-role, and similarly for ‘bolt’. If that’s right, then the theory of nuts-and-bolts implicitly defines its theoretical term ‘nut’ using the rest of the theory’s vocabulary.

Here, more generally and accurately, is Lewis’s proposal. Let \( T \) be a theory written using two sorts of vocabulary—the \( T \)-vocabulary (the Theoretical vocabulary) and the \( O \)-vocabulary (the Old or Other vocabulary). If \( T \) implicitly defines the \( T \)-vocabulary in terms of the \( O \)-vocabulary, then the implicit definitions can be made explicit as follows. Convert the \( T \)-vocabulary to names by writing, for instance, “has the property \( F \)ness” for “is \( F \)”. Write out the theory in question as a long conjunctive sentence \( T[\tau_1, \ldots, \tau_n] \)—the postulate of \( T \)—where \( \tau_i \) is a name in the amended \( T \)-vocabulary. Replace \( \tau_i \) by the variable \( x_i \) to get the open sentence \( T[x_1, \ldots, x_n] \)—the realization formula of \( T \). Define the \( T \)-term \( \tau_i \) by:


Thus, if \( \tau_i \) refers at all, it refers to the \( i \)th member of the \( n \)-tuple that uniquely realizes the realization formula of \( T \). (We can ignore subtleties concerning multiple or partial realization of the realization formula.\(^{21}\)

Let us go through this procedure for our example. Converting the \( T \)-vocabulary to names, and writing out \( T \) (the nuts-and-bolts theory) as a long conjunctive sentence gives:

A thing having bolthood is a rod with a screw thread at one end and everything having bolthood has a thing having nuthood that screws onto it and everything having nuthood screws on to the end of something having bolthood.

This is the postulate of \( T \). Then ‘nuthood’ is defined as:

The \( y_2 \):

The \( y_2 : \exists y_1 \forall x_1 x_2 ( [A \text{ thing having } x_1 \text{ is a rod with a screw thread at one end and everything having } x_1 \text{ has a thing having } x_2 \text{ that screws onto it and} \] \)
everything having \( x_2 \) screws on to the end of something having \( x_1 \)] iff \( y_1 = x_1 \) & \( y_2 = x_2 \)^{22}

And similarly for ‘bolthood’. (Note that if this definition is correct, then ‘if nut-hood exists, the theory of nuts-and-bolts is true’ is analytic.\(^{23}\))

Let us understand a claim of the form: *so-and-so vocabulary is implicitly defined by such-and-such vocabulary*, to mean that there is some theory \( T \) (the *defining* theory) with so-and-so the \( T \)-vocabulary and such-and-such the \( O \)-vocabulary, such that applying Lewis’s method correctly defines the former vocabulary by means of the latter.

Then we can put one thesis that is at least strongly suggested by Lewis’s original (1970) paper thus: the “theoretical” vocabulary of science (‘electron’, ‘gene’, ‘mollusc’, ‘white dwarf’, etc.) is implicitly defined by the rest of the scientific vocabulary (i.e. relatively commonsense and topic neutral words plus, presumably, mathematical vocabulary).

This is important, and if it is right then the theoretical vocabulary of science can be analysed in more-or-less everyday vocabulary (plus math). But discussion of this can be dropped, because for present purposes we want an analysis in the *other* direction.

The most well-known purported example of the desired sort is the core of *commonsense* (or *analytic*) functionalism: mental vocabulary is implicitly defined by non-mental/non-semantic vocabulary, with the defining theory being folk psychology.\(^{24}\) And recently Frank Jackson and Philip Pettit have defended *moral* functionalism: moral vocabulary is implicitly defined by non-moral vocabulary, with the defining theory being folk morality (Jackson and Pettit 1995; Jackson 1992, 1998a). (This isn’t quite enough for our needs, of course: the non-mental/non-semantic and non-moral vocabulary might not be physicalistically acceptable without further analysis; but we can set this complication aside.) Let us concentrate on the former example.\(^{25}\)

In Lewis’s 1970 paper, some opening assumptions are that “the best scientific explanation we can devise for a body of data includes a new theory \( T \), formulated by means of a postulate in which there occur some new terms \( \tau_1 \ldots \tau_n \), terms we have never used before... Our only clue to their meaning is the postulate of \( T \) that introduced them” (79–80). What is relevant for present purposes is not the bit about “the best scientific explanation”, but the assumption that the “only clue” to the meaning of the “new terms” is the postulate of \( T \). The form of this crucial assumption can be set out thus:

(A) The new (meaningful) terms \( \tau_1, \ldots, \tau_n \) were introduced into the language by means of a postulate (containing other antecedently understood vocabulary), which was our only clue to the meaning of these terms.

(A) is important because, if it’s right, Lewis’s suggestion that the new terms are implicitly defined by the remainder of the vocabulary in the postulate is obvi-
ously worth taking seriously—for if they are not, how could we have learned what they mean? That question is not unanswerable, but let us grant for the sake of the argument that if (A) holds in a particular case, then Lewis’s suggestion is plausible.26

Perhaps, idealizing only slightly, (A) holds for some scientific terms like ‘electron’ and ‘gene’.27 But however that may be, it certainly does not hold for mental vocabulary. Further, no one has actually produced a reasonable candidate for the (alleged) defining theory for psychological vocabulary, for good reason. First, any such candidate must be a very rich theory—it is clear that psychological terms do not have simple reductive definitions.28 Second, generalizing a point noted at the end of the “nuts-and-bolts” example, if $T$ implicitly defines $\tau_1, \ldots, \tau_n$, then “if $\tau_1, \ldots, \tau_n$ exist, $T$ is true” is analytic. Putting these two together, we need a rich folk psychology meeting the analytic constraint. But a little experimentation will soon show that analytic psychology is a difficult business—the kinds of analytic-smelling psychological truths seem relatively few in number.

To sum up, in the case of mental vocabulary, (A) fails, and no one is likely to display a theory that stands a chance of being defining. So why suppose that we have here the basis of an argument for the existence of analyses of mental vocabulary?

Consider the following passage from Lewis:

Imagine our ancestors first speaking only of external things, stimuli, and responses ...until some genius invented the theory of mental states, with its newly introduced $T$-terms, to explain the regularities among stimuli and responses...[this] story is a myth...in fact, Sellars’ myth of our Rylean ancestors... It is a good myth if our names of mental states do in fact mean just what they would mean if the myth were true. I adopt the working hypothesis that it is a good myth. This hypothesis can be tested, in principle, in whatever way any hypothesis about the conventional meanings of our words can be tested. I have not tested it; but I offer one item of evidence... There is a strong odor of analyticity about the platitude of common-sense psychology. The myth explains the odor of analyticity and the plausibility of behaviourism. If the names of mental states are like theoretical terms, they name nothing unless the theory (the cluster of platitudes) is more or less true. Hence it is analytic that either pain, etc., do not exist or most of our platitudes about them are true. If this seems analytic to you, you should accept the myth, and be prepared for psychophysical identifications (1972, 213, footnotes omitted).

We can set out Lewis’s “working hypothesis” as follows:

(i) There is a counterfactual circumstance $C$ in which (A) holds for mental vocabulary. That is, in $C$ mental vocabulary is introduced into the language by means of a postulate, whose other vocabulary is (as Lewis plainly assumes) non-semantic, which provides our only clue to the meaning of the new vocabulary; and in $C$ the mental vocabulary means what it actually does.
And:

(ii) In C, the mental vocabulary is implicitly defined by the remainder of the vocabulary in the postulate. (And so mental vocabulary as we actually use it can be defined in non-mental/non-semantic terms.)

In the quoted passage, the gap between (i) and (ii) is elided: plainly Lewis is tacitly assuming that (ii) is a reasonable inference from (i). I have already said that I am not going to fuss over this step.

Why believe this “working hypothesis”? It is not absolutely clear in this passage just how strong Lewis takes his “item of evidence” to be, but in any case, how strong is it? First, it surely cannot be that the working hypothesis is the best explanation of psychological analyticities and the plausibility of behaviourism. As far as the analyticities go, let us suppose, with Lewis (1969), that they are to be explained by our conventions of language. True, those conventions may be such that every item of mental vocabulary is analytically equivalent to some non-mental expression. However, the analyticities might well be accounted for by conventions that do not have this strong consequence. Why think otherwise? And surely the plausibility of behaviourism can be explained in other ways.

It seems to me, then, that Lewis has not convincingly argued that mental vocabulary is implicitly defined, by physical vocabulary or anything else. And no other attempt that I am aware of—in either the mental or the moral case—comes any closer.

In any event, the entire strategy is piecemeal, with each claim that so-and-so vocabulary is implicitly defined needing to be treated individually. We now turn to a series of arguments with more global pretensions.

2.2 The argument from conceivability (Chalmers 1996)

Chalmers introduces the notion of “logical” supervenience, and tells us that if the B-facts supervene “logically” on the A-facts, then “Laplace’s demon could read off the B-facts from a specification of the A-facts, as long as it possesses the B-concepts in question” (1996, 36). In other words, logical supervenience implies a priori deducibility (and conversely: see 70, 76). With that in mind, consider the following argument.

The logical supervenience of most high-level facts is most easily seen by using conceivability as a test for logical possibility. What kind of world could be identical to ours in every last microphysical fact but be biologically distinct? Say a wombat has two children in our world. The physical facts about our world will include facts about the distribution of every particle in the spatiotemporal hunk corresponding to the wombat, and its children, and their environments, and their evolutionary histories. If a world shared those physical facts with ours, but was not a world in which the wombat had two children, what could that difference consist in? Such a world seems quite inconceivable...
The same goes for architectural facts, astronomical facts, behavioral facts, chemical facts, economic facts, meteorological facts, sociological facts, and so on. A world physically identical to ours, but in which these sorts of facts differ, is inconceivable. In conceiving of a microphysically identical world, we conceive of a world in which the location of every last particle throughout space and time is the same. It follows that the world will have the same macroscopic structure as ours, and the same macroscopic dynamics. Once all this is fixed there is simply no room for the facts in question to vary...

...Even a superbeing, or God, could not imagine such a world... Once they have imagined a world with all the physical facts, they have automatically imagined a world in which all the higher-level facts hold (1996, 73).34

Restricting attention to the supposed fact that a wombat has two offspring, there is little doubt about the conclusion of this argument. It is (modulo the assumption about wombats):

(1) There is a φ such that \( \phi \supset \text{a wombat has two offspring} \) is knowable a priori.

But how does the argument actually run? Chalmers asks us to suppose that there is a wombat with two offspring in our world, and then to imagine (conceive of) a world \( w \) that is physically just the same as our world (under the supposition), but in which the high level facts (in particular, those about wombat reproduction) differ from the facts in our world. Allegedly, we fail. Now what we were asked to imagine? On one superficial reading, this:

(2) There is a world physically just the same as the actual world in which there’s no wombat with two offspring.

If that’s right, then the argument proceeds by claiming that (2) is inconceivable.35 Assuming, with Chalmers, that P’s inconceivability implies that the negation of P is a priori, we may conclude:

(3) It is a priori that: in every world physically just the same as the actual world there is a wombat with two offspring.

The main difficulty is that (3) only gets us as far as:

(4) It is a priori that there is a φ such that \( \phi \supset \text{a wombat has two offspring} \) is necessary.

And of course (4) does not entail (1).36

On an alternative and more plausible reading of what we are supposed to imagine, it is something like:
(5) $\Phi$ and there is no wombat with two offspring,

where `$\Phi$' is replaced by a certain very complex physical sentence specifying, for example, the distribution of every particle in the wombat enclosure at Taronga Park Zoo. Running the argument as before, we get:

(6) It is a priori that if $\Phi$ then there is a wombat with two offspring,

from which (1) follows. The difficulty here is that we don’t actually know what the right replacement for `$\Phi$' is, and it seems entirely a matter for speculation whether the result would be inconceivable (remember that a physical sentence is from the language of physics: it can mention “mass, charge, spatiotemporal position; properties characterizing the distribution of various spatiotemporal fields, the exertion of various forces, and the form of various waves; and so on...[but] [s]uch “high-level properties as juiciness, lumpiness, giraffehood, and the like are excluded, even though there is a sense in which these properties are physical” (Chalmers 1996, 33)).

Another way of arguing for (1) is suggested by the very last portion of the quoted passage. First establish that a Laplacian demon who imagines the physical facts would “have automatically imagined” that a wombat has two offspring, then argue that (1) follows. Although this suggestion is not developed under the heading of ‘conceivability’, Chalmers later gives an argument that appears somewhat related, to which we now turn.

2.3 The mental simulation argument (Chalmers 1996)

...in principle one could build a big mental simulation of the world and watch it in one’s mind’s eye, so to speak. Say a man is carrying an umbrella. From the associated microphysical facts, one could straightforwardly infer facts about the distribution and chemical composition of mass in the man’s vicinity, giving a high-level characterization of the area... It would be clear that he was carrying some device that was preventing drops of water from hitting him. Doubts that this device was an umbrella could be assuaged by noting from its physical structure that it can fold and unfold; from its history that it was hanging on a stand that morning, and was originally made in a factory with others of similar kind, and so on (1996, 76).

Chalmers is trying to show that there is a $\phi$ such that `$\phi \supset$ there’s a man carrying an umbrella’ is a priori, and similarly for other “high-level facts” (76). Talk of “building a mental simulation” and of “watching it one’s mind’s eye” encourages the comparison with building a physical model using $\phi$ as a recipe, and viewing the result. In the latter case, if `$\phi \supset$ there’s a man carrying an umbrella’ is necessarily true, the model would indeed contain a man carrying an umbrella, and we could presumably see that this was so. That might be thought to lend some plausibility to the claim that we would “see” the same thing if we built a mental
simulation instead, and thus plausibility to the claim that ‘φ ⊢ there’s a man carrying an umbrella’ is a priori. But obviously this thought would be mistaken: when Chalmers says that if one produces a mental simulation according to φ, one will see in one’s mind’s eye that there’s a man carrying an umbrella, he is simply choosing a metaphorical way of saying that ‘φ ⊢ there’s a man carrying an umbrella’ is a priori. And what we might learn by building physical models is irrelevant to the a prioricity of this conditional: the only effect of the metaphor is to mislead us to think otherwise.

The mental simulation argument, then, simply boils down to Chalmers’ assertions that one could infer from the physical description that there is an object that “can fold and unfold”, that “was hanging on a stand that morning”, that “was made in a factory with others of a similar kind, and so on”, eventually arriving at the conclusion that there’s a man carrying an umbrella.

However, this attempt to show how one might infer by a series of steps that there’s a man carrying an umbrella is somewhat cosmetic, because the conclusion is supposed to apply to “almost any sort of high-level phenomena” (77), and the umbrella example is the only one discussed in any detail. Even if cosmic hermeneutics is possible here, for all Chalmers has said it might fail elsewhere, for a very different kind of sentence. And in any case the brief treatment of the example is unconvincing, for two reasons. First, even if we allow that one could infer that there is an object that can “fold and unfold”, that “was hanging on a stand that morning”, etc., this only helps if these conclusions amount to jointly a priori sufficient conditions for the existence of an umbrella. As stated, they do not: tablecloths, Panama hats and briefcases all fold and unfold, can hang on stands, are made in factories, and can be used as shelter from rain. Chalmers’ use of ‘and so on’ indicates that he recognizes this fact, but it is not obvious what else to add. Second, and more importantly, if it is initially quite unclear whether one can infer that there’s a man carrying an umbrella from the physical facts (as it surely is, especially when the physically austere nature of the premises is emphasized—see 2.2 above), it ought to be equally unclear whether one can infer, for example, that an object has been “made in a factory with others of a similar kind”. That is, factory-facts and umbrella-facts are at the same “high-level”. Thus we are no further forward.

2.4 Two-dimensionalism

As we have just seen, Chalmers tries to establish the conclusion that—a qualification about consciousness aside—cosmic hermeneutics is possible (and so physicalism is true). But Jackson attempts the more modest goal of showing that if physicalism is true, cosmic hermeneutics is possible. As will soon be apparent, the dialectic here is somewhat convoluted, so a road map will come in handy.

According to Jackson, his central argument needs the premise that a certain semantic framework—two-dimensionalism—is correct. So, the first order of busi-
ness is to explain two-dimensionalism and Jackson’s motivation for it (2.41). Next, the central argument (2.42), which in fact does not need the distinctively “two-dimensional” part of two-dimensionalism as a premise, but instead another component of it, that every necessary proposition is knowable a priori. Moreover, this component is inadequately defended. Some of Jackson’s remarks suggest another argument, which does indeed require something “two-dimensional”. However, this could at best only fend off an objection to Jackson’s conclusion (2.43).

2.41 Two-dimensionalism explained

Jackson’s motivation for two-dimensionalism begins with this passage:

Consider what happens when I utter the sentence, ‘There is a land mine two metres away’. I tell you something about how things are, and to do that is precisely to tell you which of the various possibilities concerning how things are is actual. My success in conveying this urgent bit of information depends on two things: your understanding the sentence, and your taking the sentence to be true. We have here a folk theory that ties together understanding, truth, and information about possibilities; and the obvious way to articulate this folk theory is to identify, or at least essentially connect, understanding a sentence with knowing the conditions under which it is true; that is, knowing the possible worlds in which it is true and the possible worlds in which it is false; that is, knowing the proposition it expresses on one use of the term ‘proposition’... it would, I think, be wrong to regard the folk theory as being as controversial as [its articulations in the work of David Lewis and Robert Stalnaker]. The folk theory is, it seems to me, a commonplace (Jackson 1998a, 71; see also 1994a, 37–8).

Suppose we articulate our “folk theory” in the way Jackson suggests. That is, we “identify, or at least essentially connect, understanding a sentence with... knowing the possible worlds in which it is true and the possible worlds in which it is false”. Now we have, as Jackson goes on to remark, a puzzle. For surely someone can understand, say, a necessary truth like ‘Water is H₂O’, and yet not have any inkling that it expresses a true proposition: indeed, such a person might believe that the proposition it expresses is false.³⁷ But how can that be, if understanding that sentence involves knowing that it expresses a proposition true at every world?

We can see Jackson’s puzzle as generated by the following pair of schematic claims:

(7) Understanding a sentence S = knowing which proposition S expresses.
(8) Knowing which proposition S expresses involves “knowing the possible worlds in which it is true”.

From (7) we get that someone who understands ‘Water is H₂O’ knows which proposition it expresses. From (8) we get that someone who knows which proposition ‘Water is H₂O’ expresses knows “the possible worlds in which it is true”,
which in the case at hand amounts to knowing that the sentence expresses a proposition true at every world. Putting the two together gives us the puzzle.

Importantly, Jackson’s solution is not to deny (8). Rather, he denies (7) (for sentences like ‘Water is H₂O’). Thus two dimensionalism is motivated because, as we’ll see in a moment, it provides an elegant alternative account of what it takes to understand sentences like ‘Water is H₂O’. So before turning to two dimensionalism, we should pause to consider why Jackson thinks it would be wrong to deny (8) instead. (8) says, taking ‘Water is H₂O’ as an instance, that anyone who knows which proposition it expresses knows:

(9) The proposition expressed by ‘Water is H₂O’ is true at every world.

Apparently Jackson seeks to rest the argument for this on “commonplace” premises, avoiding the “controversial articulations” of Lewis and Stalnaker. But if we take this requirement seriously, it is very hard to see what is wrong with denying (8). What is commonplace is that someone who knows which proposition ‘Water is H₂O’ expresses knows (if he’s conceptually sophisticated), not (9), but:

(10) The proposition expressed by ‘Water is H₂O’ is true at a world w iff, in w, water is H₂O.

And since (9) is not an a priori consequence of (10), there is no evident reason to suppose that anyone who knows which proposition ‘Water is H₂O’ expresses knows (9). Therefore, disallowing appeal to any “controversial articulations”, the right response to Jackson’s puzzle is to reject (8).³⁸

Jackson’s official motivation for two-dimensionalism, then, is unconvincing. But let us set this difficulty aside, and turn to two-dimensionalism itself.

Are there any uncontroversial cases where a speaker may be said to understand a sentence without knowing which proposition it expresses? Of course: if I tack a note saying ‘Back in 10 minutes’ or, less elliptically, ‘I’ll be back here in 10 minutes’, to my office door, plainly someone could be said to understand the (token) sentence without knowing which proposition it expresses. For he may have no idea when the note was placed on the door, whose office it is, or even where the office is.

In the example of the note, the speaker understands the (token) sentence because he knows how the proposition expressed by tokens of the type ‘Back in 10 minutes’ varies with arbitrary context of utterance (cf. Kaplan 1989, 520–1). (We may give exactly the same account of understanding the sentence type ‘Back in 10 minutes’.)

And similarly, Jackson thinks, with ‘Water is H₂O’ and other sentences containing ‘water’, for instance ‘Water covers most of the Earth’. According to him, “understanding ‘Water covers most of the Earth’, does not require knowing the conditions under which it is true, that is, the proposition it expresses. Rather it requires knowing how the proposition expressed depends on context of utterance” (1998a, 73). (See also Jackson 1994a, 38–9, 1994c, 489–90.)
To fill in the details of this idea, Jackson uses some formal machinery borrowed from two-dimensional modal logic, which we now need briefly to explain. (A slightly expanded version of what follows can be found in Jackson 1998a, 47–52; Chalmers 1996, 56–65; and Block and Stalnaker forthcoming; a highly compressed version is in Lewis 1994, 415.) Take the sentence ‘Water covers most of the Earth’ (understood as we English speakers understand it). According to Jackson, thought experiments of the Twin Earth variety (Putnam 1975) tell us how the proposition this sentence expresses varies with the immediate environment of the speaker. For example, the proposition expressed by an utterance of that sentence in a context in which H₂O falls as rain, flows in streams, etc., is true at a world w iff, in w, H₂O covers most of the Earth. And the proposition expressed in a context in which XYZ falls as rain, flows in streams, etc., is true at a world w iff, in w, XYZ covers most of the Earth. These facts, about how the proposition expressed depends on context, are (because they are knowable via appropriate thought experiments) knowable a priori.

Now (we are supposing, with Jackson) two tokens of ‘Water covers most of the Earth’ uttered in different possible worlds, or within the same world, may express different propositions. An example of the latter is a world containing both Earth and Twin Earth (with tokens uttered in both places). It will greatly ease exposition at no significant cost if we just ignore this second alleged aspect of context relativity. (Perhaps the easiest way of doing this is to pretend that every world has exactly one privileged context of utterance, and that the privileged context for the actual world is the one enjoyed by speakers of English here on Earth.)

So, imagine some possible world w. Ask: what proposition would ‘Water covers most of the Earth’ express (in the privileged context) if w had turned out to be actual? Here we are, in the terminology of Davies and Humberstone 1980, considering w as actual. (An analogous question is: what proposition would (my token of) ‘I am depraved’ express if Clinton had turned out to be me?, understood so that the correct answer is: the proposition that Clinton is depraved; here, we might say, we are considering Clinton as the speaker.) For example, let w* be a world where XYZ falls as rain, flows in streams, etc. Considering w* as actual and according to Jackson, ‘Water covers most of the Earth’ expresses a proposition true at a world w iff, in w, XYZ covers most of the Earth.

We can also conduct a similar exercise with the word ‘water’. We can ask, for any worlds w, w*: what is the reference of ‘water’ in w, if w’ had turned out to be actual? If w’ = w*, Jackson’s answer is, of course, XYZ. Let ‘the watery stuff’ be a non-rigid description with the same descriptive content as ‘water’, whatever that is (cf. Chalmers 1996, 57). Then ‘the watery stuff’ refers to a substance S in w iff, considering w as actual, ‘water’ refers to S in w. And, considering a world w’ as actual, ‘Water covers most of the Earth’ expresses a proposition true at a world w iff, in w, the substance that is the watery stuff in w’ covers most of the Earth.

Of course, because H₂O falls as rain, etc., ‘Water covers most of the Earth’ in fact expresses a proposition that is true at a world w iff, in w, H₂O covers most of the Earth. (We are here considering each world w as counterfactual.)
Thus ‘Water covers most of the Earth’ determines a function $F$ from worlds to propositions, $F(w)$ being the proposition that ‘Water covers most of the Earth’ expresses, considering $w$ as actual. Let us follow Jackson’s exposition of two-dimensionalism and assume that propositions are sets of possible worlds (or functions from worlds to truth values). Then we can display $F$ in the following matrix:

<table>
<thead>
<tr>
<th>counterfactual</th>
<th>actual</th>
<th>$w^*$</th>
<th>$w^\uparrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>@</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>$w^*$</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>$w^\uparrow$</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

(If propositions are more fine-grained than sets of worlds, then the matrix does not display $F$: the matrix is determined by $F$, but not conversely.)

Here @ is the actual world, $w^*$ is a Twin Earth world, and $w^\uparrow$ is a world much like ours except that the oceans have largely dried up (all the other possible worlds have been omitted for reasons of space). A horizontal row specifies the proposition expressed by ‘Water covers most of the Earth’, considering the row-world as actual. So, considering @ as actual, the proposition expressed by ‘Water covers most of the Earth’ is a set of worlds that has @, but not $w^*$ or $w^\uparrow$, as members. And that proposition (i.e. the proposition that water covers most of the Earth), given that we are now identifying propositions with sets of worlds, is the proposition that $H_2O$ covers most of the Earth.

As we’ve seen, Jackson maintains that understanding ‘Water covers most of the Earth’ does not require knowing which proposition it expresses, but rather “knowing how the proposition expressed depends on context of utterance”. We can use the two-dimensional apparatus to give a more precise formulation of the second of these claims. The proposition expressed depends on the context of utterance in this way: if the context is $w'$, (i.e. if $w'$ is considered as actual), then ‘Water covers most of the Earth’ expresses a proposition that is true at any world iff, in that world, the substance that is the watery stuff in $w'$ covers most of the Earth. So, according to Jackson, understanding ‘Water covers most of the Earth’ requires knowing:

(11) For all worlds $w'$, considering $w'$ as actual, the proposition expressed by ‘Water covers most of the Earth’ is true at a world $w$ iff, in $w$, the substance that is the watery stuff in $w'$ covers most of the Earth.

An important consequence of (11) is (setting $w'=w$):

(12) For all worlds $w$, considering $w$ as actual, the proposition expressed by ‘Water covers most of the Earth’ is true at $w$ iff, in $w$, the watery stuff covers most of the Earth.
The proposition that the watery stuff covers most of the Earth thus specifies the context in which ‘Water covers most of the Earth’ expresses a true proposition. So if someone knows that the watery stuff covers most of the Earth, and understands ‘Water covers most of the Earth’, he knows that this sentence expresses a true proposition, although he may not know which one it is. A glance at our matrix shows that the proposition that the watery stuff covers most of the Earth can be read off the diagonal from top left to bottom right; borrowing the terminology of Stalnaker 1978, say that this is the diagonal proposition associated with ‘Water covers most of the Earth’.

Turn now to the necessary a posteriori truth ‘Water is H₂O’. According to two-dimensionalism, it has the following matrix:

<table>
<thead>
<tr>
<th>counterfactual→</th>
<th>@</th>
<th>w*</th>
<th>w†</th>
</tr>
</thead>
<tbody>
<tr>
<td>actual</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>@</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>w*</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

The proposition expressed by ‘Water is H₂O’, considering @ as actual, is necessary (and so, given the identification of propositions with sets of worlds, is the necessary proposition). Hence, since the necessary proposition is knowable a priori (it is, for example, the proposition that everything is self-identical, and that is knowable a priori), it is knowable a priori that water is H₂O. But the diagonal proposition associated with ‘Water is H₂O’ (the proposition that the watery stuff is H₂O) is not necessary. Therefore the sentence is true in some contexts and false in others, and so merely knowing how the proposition expressed depends on the context is not sufficient for knowing that the sentence is true. That is why, according to Jackson, understanding ‘Water is H₂O’ is not sufficient for knowing that the sentence is true.

On this way of analysing the necessary a posteriori, as Lewis says, “there is no such thing as a necessary a posteriori proposition” (1994, 415; see also Jackson 1998a, 84–6, 1994c, 489). Instead, it is the sentence ‘Water is H₂O’ that is properly described as necessary a posteriori, understood to mean that the proposition it expresses is necessary (and a priori) and that the diagonal proposition associated with it is a posteriori (and contingent).

Turn now to ‘The watery stuff is water’. Within the two-dimensional framework, it is a contingent a priori sentence. It expresses, of course, the same (contingent) proposition as ‘The watery stuff is H₂O’. However, someone who understands ‘The watery stuff is water’ knows that it, unlike ‘The watery stuff is H₂O’, expresses a truth in every context. In other words, the diagonal proposition associated with ‘The watery stuff is water’ is necessary.

On this way of analysing the contingent a priori, there is no such thing as a contingent a priori proposition. Instead, it is the sentence ‘The watery stuff is water’ that is properly described as contingent a priori, understood to mean that
the proposition it expresses is contingent (and a posteriori) and that the diagonal proposition associated with it is necessary (and a priori) (cf. Stalnaker 1978).

It is important to realize that two-dimensionalism has two independent components. First, that the semantics of a sentence (e.g. ‘Water covers most of the Earth’) determines a matrix of the sort discussed. Second, that there are no necessary a posteriori propositions (nor contingent a priori ones). (The identification of propositions with sets of worlds is one way, but not the only way, of securing the second component.) Neither component implies the other. In particular, as we will see in the section after next, a popular view of the semantics of words like ‘water’ retains the first and rejects the second.46

After some discussion of Jackson’s argument from two-dimensionalism, in the following section, it will turn out that the only component of two-dimensionalism he needs is the second.

2.42 The argument from two-dimensionalism (Jackson 1998a, 1994c)

With the two-dimensional framework in place, Jackson considers the question of whether the “physicalists are committed to the existence of conceptual entailments from the physical to the psychological”. If, he says, “the explanation drawing on two-dimensional modal logic we gave above of the necessary a posteriori is correct”, the answer is yes (1998a, 81). The argument revolves around the following example:

**Argument A**

(13) H₂O covers most of the Earth.  
Therefore, (14) Water covers most of the Earth.47

It will become clear—eventually—how Jackson’s treatment of this example can be turned into an argument for the possibility of cosmic hermeneutics with respect to psychological truths (and more generally with respect to any truth), if physicalism is true.

Jackson first notes that A is modally valid: “every world where the...proposition expressed by (13)...is true is a world where...the proposition expressed by (14)...is true” (1998a, 81, n35). But, he continues, “the conditional with the premiss as antecedent and the conclusion as consequent is necessary a posteriori, not a priori” (81–2); and a little further on: “the passage from (13) to (14) is a posteriori” (82).

It is clear from these remarks that, on Jackson’s usage, ‘(v)’ placed to the left of sentence α names α. (Earlier in this paper, obvious exceptions aside, ‘(v)’ placed to the left of sentence α named the proposition expressed by α.) For the remainder of this section, let us adopt Jackson’s convention.

The following three consequences of applying two-dimensionalism to A are important.

First, the proposition expressed by ‘Water covers most of the Earth’ is distinct from the diagonal proposition associated with it. The former is the propo-
sition that water covers most of the Earth, the latter the proposition that the watery stuff covers most of the Earth. (We may assume, apparently with Jackson, that the proposition expressed by ‘H₂O covers most of the Earth’ is the diagonal proposition associated with it.)

Second, because we are taking propositions to be sets of possible worlds, the proposition expressed by (13) is the proposition expressed by (14). Thus, if we ask, of the propositions expressed by the sentences in A, whether the first entails the second a priori, the answer is, quite trivially, yes.

Third, although, as just pointed out, the passage from the proposition expressed by (13) to the proposition expressed by (14) is a priori, evidently even a keen logician, who understands both (13) and (14), and believes (13) to be true, will not thereby conclude that (14) is true. The reason for this is that the conditional:

\[(15) \text{H}_2\text{O covers most of the Earth} \supset \text{water covers most of the Earth.}\]

is necessary a posteriori. That is, the proposition it expresses is necessary and the diagonal proposition associated with it is contingent. So understanding (15), which requires only knowing how the proposition it expresses depends on context, is not sufficient for knowing which proposition it expresses.

Say that sentence α a priori implies β iff the diagonal proposition associated with ‘α ⊃ β’ is the necessary proposition. Now we can sum up Jackson’s comments about A as follows: the proposition expressed by (13) entails the proposition expressed by (14), but (13) does not a priori imply (14).

Let us return to Jackson’s argument. He continues:

Thus, if the two-dimensional explanation of the necessary a posteriori is correct, the appropriate supplementation of the premisses by contextual information will give a set of premisses that do lead a priori to the conclusion. We will be able to move a priori from, for example, sentences about the distribution of H₂O combined with the right context-giving statements, to the distribution of water (82).

Adding the appropriate “contextual information” to A gives:

**Argument B**

(13) H₂O covers most of the Earth.
(13a) H₂O is the watery stuff.⁴⁸

Therefore, (14) Water covers most of the Earth.

Jackson comments that “the passage from (13) together with (13a) to (14) is a priori in virtue of the a priori status of ‘Water is the watery stuff [...]’.”⁴⁹ His point is that adding (13a) to A yields an argument the conjunction of whose premisses a priori implies (14). That is, the diagonal proposition associated with the conditional:
(16) \( \text{H}_2\text{O} \) covers most of the Earth & \( \text{H}_2\text{O} \) is the watery stuff \( \supset \) water covers most of the Earth.

is necessary (in other words, the conditional expresses a truth in any context). And the diagonal proposition associated with (16) is, of course, the proposition expressed by:

(17) \( \text{H}_2\text{O} \) covers most of the Earth & \( \text{H}_2\text{O} \) is the watery stuff \( \supset \) the watery stuff covers most of the Earth.

Suppose our demon understands all the sentences in \( B \). Then because the diagonal proposition associated with (16) is necessary, he knows that it expresses a truth in any context. So he knows that (14) expresses a truth if (13) and (13a) do. Thus, if he knows that (13) and (13a) express truths he can know that (14) does. We have already assumed, with Jackson, that the diagonal proposition associated with (13) is the proposition it expresses, and we may assume the same for (13a). Therefore, understanding (13) and (13a), and knowing the propositions they express, is sufficient for knowing that they express truths. *The upshot is that if the demon understands all the sentences in \( B \), and knows that \( \text{H}_2\text{O} \) covers most of the Earth and that \( \text{H}_2\text{O} \) is the watery stuff, he may deductively infer that ‘Water covers most of the Earth’ expresses a truth.*

After discussing the example of water and \( \text{H}_2\text{O} \), Jackson finally argues for the promised conclusion, the possibility of cosmic hermeneutics (if physicalism is true), as follows:

The crucial point here is that the way that the contextual information, the relevant information about the way things actually are, by virtue of telling us in principle the propositions expressed by the various sentences...enables us to move a priori from the \( \text{H}_2\text{O} \) way things are to the water way things are. But if physicalism is true, all the information needed to yield the propositions being expressed about what the actual world is like in various physical sentences can be given in physical terms, for the actual context is given in physical terms according to physicalism. Therefore, physicalism is committed to the in principle a priori deducibility of the psychological on the physical (83).

This is perhaps a little compressed. Step back and quickly review what two-dimensionalism has told us about \( A \).

Suppose our demon knows that \( \text{H}_2\text{O} \) covers most of the Earth. Can he thereby conclude that water covers most of the Earth? Yes—he knows it already! *This is implied solely by the claim that propositions are sets of worlds.*

As a competent user of English, can he thereby conclude that the sentence ‘Water covers most of the Earth’ expresses a true proposition? No, he can’t. But if he knows, in addition, that \( \text{H}_2\text{O} \) is the watery stuff, then he can conclude that ‘Water covers most of the Earth’ expresses a true proposition. But does the demon
know that H₂O is the watery stuff? That is the question, I take it, that the above quotation from Jackson is designed to answer in the affirmative.⁵¹

So, is Jackson right? Although the expression ‘H₂O’ is (we have been granting) part of the physical lexicon, we should not grant that ‘H₂O is the watery stuff’ is a physical sentence—‘the watery stuff’ is admittedly a bit of technical terminology, but it is supposed to be simply ‘water’ unrigified, and so it belongs among our everyday vocabulary. Moreover, there is a conclusive reason for extruding it from the physical lexicon: any physicalist will agree that the supervenience base for everything can be stated in a language without any expression remotely cognate with ‘the watery stuff’. Hence it does not trivially follow from the fact that the demon knows everything expressed by true physical sentences that he knows that H₂O is the watery stuff. Here is the relevant part of the quoted passage, where Jackson gives the argument that the demon does know this:

But if physicalism is true, all the information needed to yield the propositions being expressed about what the actual world is like in various physical sentences can be given in physical terms, for the actual context is given in physical terms according to physicalism.

And what is presumably the same point is made in an earlier paper as follows:

Although understanding may not even in principle be enough to yield truth-conditions, it is enough to yield how truth-conditions depend on context. But of course the context is, according to the physicalist, entirely physical. Hence, the physicalist is committed to there being an a priori story to tell about how the physical way things are makes true the psychological way things are (Jackson 1994a, 40; see also 1994c, 491).

I can only see one way of interpreting these passages, as follows. Suppose physicalism is true. Then there is a true φ such that ‘φ ⊃ H₂O is the watery stuff’ expresses a necessary proposition. Assuming that every necessary proposition is priori (which might be supported by an independent argument in favor of the possible worlds conception of a proposition), the proposition expressed by ‘φ ⊃ H₂O is the watery stuff’ is a priori. The demon therefore knows it, and he also knows the proposition expressed by φ. Does it follow that he can deductively infer that H₂O is the watery stuff? Not yet. To get it we need an instance of the following schematic principle:

(C) If the demon knows that p ⊃ q, and knows that p, he can deductively infer that q.

With that in hand, we have the desired result.

But if this is Jackson’s point, again the only part of two-dimensionalism that is doing any work is the claim that every necessary proposition is a priori. There was no need to introduce argument B: Jackson could have argued directly that, if
physicalism is true, then there is some $\phi$ such that $\neg \phi \supset \text{‘Water covers most of the Earth’ expresses a true proposition}$. This is necessary, and hence (given that every necessary proposition is a priori and the relevant instance of (C)) the demon can know that ‘Water covers most of the Earth’ expresses a true proposition. (So, substituting any true sentence for $\neg \phi \supset \text{‘Water covers most of the Earth’ expresses a true proposition}$, and running the argument again, we get the conclusion that, if physicalism is true, cosmic hermeneutics is possible.)

An argument for the crucial premise that every necessary proposition is a priori can be extracted from Jackson’s “puzzle”, discussed above in section 2.41, about understanding a necessarily true sentence without knowing that it’s true. Entertaining a proposition, the extracted argument goes, involves knowing its truth-conditions, and thus entertaining a necessary proposition involves knowing that it is true in every condition, and so true; therefore every (entertainable) necessary proposition is a priori. But the reply is essentially the same as the one to Jackson’s original puzzle. Someone who entertains that water is $\text{H}_2\text{O}$ thereby knows (more exactly: can know, if he’s conceptually sophisticated) that the proposition that water is $\text{H}_2\text{O}$ is true at a world $w$ iff, in $w$, water is $\text{H}_2\text{O}$. That is the only uncontroversial sense in which entertaining a proposition involves knowing its truth conditions, and it does not imply that someone who entertains that water is $\text{H}_2\text{O}$ can thereby know that the proposition is true at every world.$^{52}$

Alternatively, one might first try to argue for the identification of propositions with sets of possible worlds, from which the crucial premise follows. We cannot investigate the merits of this identification here.$^{53}$ But we should note that the identification provides some motivation for denying the seemingly trivial (instances of) (C). This is because the two together make it hard to avoid the unpalatable conclusion that knowledge and belief are closed under necessary consequence.$^{54}$ So I think it fair to say that the strategy of establishing the crucial premise via the possible worlds conception of a proposition may reasonably be resisted.$^{55}$

However, for all we have said so far, perhaps the first component of two-dimensionalism, that the semantics of a sentence determines a matrix of the sort explained earlier, might be turned to the advantage of cosmic hermeneutics (this is at least suggested by Jackson 1992, 1994a, b). That is what we shall finally examine.

2.43 The argument from weakened two-dimensionalism (Jackson 1992, 1994a, b)

Let us reinstate the convention that ‘(v)’ placed to the left of sentence $\alpha$ names the proposition expressed by $\alpha$, and return to Jackson’s initial example:

**Argument A**

(13) $\text{H}_2\text{O}$ covers most of the Earth.
Therefore, (14) Water covers most of the Earth.
Argument $A$, as we are understanding it in this section, has propositions, not sentences, as its premise and conclusion. Suppose we accept that $A$, despite being modally valid, is not a priori valid. That is, although (the proposition expressed by) ‘$H_2O$ covers most of the Earth ⊃ water covers most of the Earth’ is necessary, it is not a priori. We are, then, now working with a more intuitive conception of a proposition than the possible worlds sense, moreover one according to which there are necessary a posteriori propositions.

It must be stressed that the question we are now focussing on is quite different from the one that occupied most of the previous section. That question was: what more information does someone who understands ‘$H_2O$ covers most of the Earth’ and ‘Water covers most of the Earth’, and who knows that $H_2O$ covers most of the Earth, require in order to infer deductively that ‘Water covers most of the Earth’ expresses a true proposition? The present question is: what more information does someone who knows that $H_2O$ covers most of the Earth require in order to infer deductively that water covers most of the Earth? Recall that with the assumptions of the previous section our present question had an easy answer, namely that no additional information is required. But with our present assumptions that is no longer true.

However, suppose we want to retain one key idea of the two-dimensionalist analysis of the necessary a posteriori, that the reference of ‘water’ is fixed by the world of utterance. An obvious way to do that is to identify the semantic content of ‘water’ with a rigidified description: something like ‘the actual potable liquid that falls as rain and flows in streams’ (cf. Davies and Humberstone 1980, 18–20; Chalmers 1996, 59; Jackson 1992, 483–4, 1994a, 39, 1994b, 187), abbreviated as ‘the actual watery stuff’. Thus, uttered in the actual world, the sentence ‘Water covers most of the Earth’ expresses a proposition that is true at a world $w$ iff, in $w$, the stuff that is watery in the actual world covers most of the Earth. (Note: we are still working with the simplifying assumption that utterances of ‘water’ within a world do not differ in reference.)

So, according to this proposal:

(13b) Water is the actual watery stuff.

is necessary and a priori, and so:

(13c) Water is the watery stuff.

is contingent and a priori. (According to two-dimensionalism as explained earlier, the proposition that water is the watery stuff is contingent but a posteriori.)

With the assumptions now in force, the addition of (13d) turns the modally valid but a priori invalid $A$ into the a priori valid $B$:

Argument $B$

(13) $H_2O$ covers most of the Earth.
(13d) $\text{H}_2\text{O}$ is the watery stuff.
Therefore, (14) Water covers most of the Earth.

However, the problem raised in the previous section now reappears. This maneuver is only going to further the cause of cosmic hermeneutics if (13d) is either a physical fact or else follows a priori from physical facts (for what is essentially the same point, see Block and Stalnaker forthcoming).56 There is no reason to think it a physical fact, in the relevant sense: ‘the watery stuff’ is an abbreviation for a description couched in folk vocabulary—vocabulary that is not part of the austerely physical supervenience lexicon. Does (13d) then follow a priori from physical facts? Of course, if physicalism is true, then there will be some $\phi$ such that $\phi \supset \text{H}_2\text{O}$ is the watery stuff is necessary. If the expansion of ‘the watery stuff’ contains further “natural kind” terms—‘liquid’, perhaps—then $\phi \supset \text{H}_2\text{O}$ is the watery stuff will be a posteriori. (See again Block and Stalnaker forthcoming.) Suppose, using the example mentioned earlier, that ‘the watery stuff’ abbreviates ‘the potable liquid that falls as rain and flows in streams’. And suppose that ‘liquid’ is the only natural kind term occurring in this description, and is to be analysed along the lines of ‘water’ by a rigidified description we can abbreviate as ‘the actual liquidish stuff’. (Thus the proposition that any liquidish stuff is a liquid will be contingent and a priori.) Then we will be able to expand $B$ as follows:

**Argument C**

(13) $\text{H}_2\text{O}$ covers most of the Earth.
(13e) $\text{H}_2\text{O}$ is a liquidish stuff.
(13f) $\text{H}_2\text{O}$ is the potable stuff that falls as rain and flows in streams.
Therefore, (14) Water covers most of the Earth.

Ex hypothesi, this argument is a priori valid, and the sentences expressing (13e) and (13f) do not contain any terms that can be analysed as rigidified descriptions. However, these sentences do contain non-physical vocabulary, and so it is a substantive question whether (13e) and (13f) can be deductively inferred from physical facts. Nothing in Jackson's presentation, or any argument discussed earlier, provides reason to think so.

Compare the previous argument—the mislabelled “argument from two-dimensionalism”—with the one under discussion. The previous argument certainly has the desired conclusion, that cosmic hermeneutics is possible if physicalism is true. But one of its premises (that every necessary proposition is a priori) is contentious and without visible support. The present argument does not have this premise, but neither does it have the desired conclusion. Taking a premise of the argument to be that *every* natural kind term can be analysed as a rigidified description (not itself containing any natural kind terms), the conclusion is this: Kripkean examples of the necessary a posteriori involving natural kind terms, like ‘Water is $\text{H}_2\text{O}$’, are not counterexamples to the claim that
cosmic hermeneutics is possible (if physicalism is true). If this argument is sound, an objection is successfully rebutted. But that is all, for we are still left wondering whether, for example, there are a priori truths of the form "ϕ ⊆ H₂O is a liquidish stuff".57 (And anyway, this is an extremely large "if".58)

Lewis, Jackson, Kripke and others pointed out (in effect) that the proper formulation of physicalism commits it to the existence of certain necessarily true conditionals.

Kripke pointed out that the notions of necessity and a priority are distinct: the former is from metaphysics, the latter from epistemology. (He then went on to give examples where the two notions came apart, but all we need is the initial observation.)

The first insight leads us to wonder whether, if physicalism is true, cosmic hermeneutics is possible. The second insight suggests that it may well not be. We have found no reason to revise this conclusion.

Notes

*I am very grateful to Ned Block, Jim Pryor, Robert Stalnaker, Daniel Stoljar, and Ralph Wedgewood for much discussion about early drafts, and to Michael Glanzberg, Noa Latham, Sarah McGrath, and Scott Soames for comments on a later version. I am also indebted to David Chalmers for correspondence, and to an ANU (RSSS) reading group, especially Frank Jackson and (again) Daniel Stoljar.

1. ‘Fact’ is used here and throughout as a synonym of ‘true proposition’. Propositions I take to be the objects of belief and the referents of ‘that’-clauses; what sorts of entities can fill this role will concern us later.

2. ‘Fact’-terminology occurs in some passages from Chalmers 1996, discussed in 2.2 and 2.3 below. We may fairly take his usage to agree with mine.

3. Henceforth, for brevity, I shall often speak of knowing a sentence α, of inferring α, of α’s being a priori, of α’s being true at world w, and so forth. By such locutions I always mean: knowing the proposition expressed by α, inferring the proposition expressed by α, etc. Unless explicitly noted otherwise, context-dependence will be harmlessly ignored.

4. See Horgan 1983. I have also borrowed Horgan’s Laplacian demon. Scholars should note that I use ‘cosmic hermeneutics’ in a somewhat different way from Horgan. He defines it as “the radical interpretive task of ascertaining all the truths at [a] given P-world on the basis of the totality of that world’s microphysical truths” (21). (A “P-world” is a possible world that is physically like ours in various respects (for which see 19).) The most significant difference is that cosmic hermeneutics in Horgan’s sense does not restrict by definition the demon’s a posteriori knowledge to physical truths: for example, he discusses a proposal that would allow the demon knowledge of “laws that link the vocabulary of microphysics to the rest of our vocabulary” (22).
5. As discussed in the following section, cosmic hermeneutics is not possible unless the Laplacian demon is given slightly more resources than his allotment so far. With the appropriate addition made, the philosophers mentioned hold the following.

According to Chalmers (1996), cosmic hermeneutics is possible with respect to almost every fact (the exceptions being, roughly speaking, facts about consciousness). Lewis (1994; see also Horgan 1983, n18) holds, simply, that cosmic hermeneutics is possible. Jackson (1998a; see also 1994a, b, c) is primarily concerned to argue only that the physicalist is committed to the possibility of cosmic hermeneutics. In his 1982, Jackson argued against physicalism, and held a position very similar to the one Chalmers now holds, but his most recent view (1998a) is that physicalism is true, which brings him into agreement with Lewis. Latham (forthcoming) is another defender of the possibility of cosmic hermeneutics; Levine (1993) appears to be sympathetic to a position similar to Chalmers’.

Horgan’s own opinion is that, in order to be able to infer everything, the demon needs extra a posteriori knowledge, but only knowledge of “meaning constraints”: “principles which are dictated by the very meaning of our higher-level vocabulary and our microphysical vocabulary” (25). (See also Horgan and Timmons 1992.) So Horgan thinks that cosmic hermeneutics (in my sense, not his—see preceding footnote) is not possible. Chalmers’ n35 (367) is thus misleading.

One might have expected Robert Stalnaker to be on the list. For he has defended a “pragmatic picture” of belief, with propositions taken to be sets of possible worlds, which is, as he admits, congenial to the idea that knowledge is closed under necessary consequence (1984, 76). And if knowledge is closed under necessary consequence, it trivially follows that, if physicalism is true, cosmic hermeneutics is possible (see the following section). However, Stalnaker takes pains to point out various ways the closure principle might be resisted, at the end of the day leaving the matter somewhat unresolved (see also Stalnaker 1991, forthcoming). For further discussion, see footnote 54 below.

6. Remember I am talking about the propositions expressed by the sentences, rather than the sentences themselves (see footnote 3 above). If the little argument just given is spelled out more explicitly, it will be seen that it appeals to instances of the following two, apparently innocuous, schematic principles (for the left-to-right and right-to-left parts, respectively):

If the proposition that p is a priori and the proposition that p ⊃ q is a priori, then the proposition that q is a priori.

And:

(C) If the demon knows that p ⊃ q, and knows that p, he can deductively infer that q.

(C) re-enters the discussion in 2.42 below.

7. Examples can be generated using the rigidifying operator ‘actually’. If φ is contingent, ‘(Actually φ) ⊃ φ’ is contingent and arguably a priori. But we will ignore this minor complication.

8. Since one sentence true at w* is ‘the world that is actual is w*’ (so we may stipulate), duplication simpliciter for worlds is identity. (Of course the argument for the equiv-
alence in the text relies on the existence of suitably strong languages—see footnote 2 above. As far as the main point of this paper goes, this linguistically profligate assumption is only for convenience.)

10. This is certainly a stipulation. As Block and Stalnaker (forthcoming) point out in a related connection, it is not a truth of physics that physicalism is true.
11. He has since come to have doubts (Jackson 1998a, 44, n21).
12. As is often pointed out, intuitively Mary when released acquires not just knowledge, but also a concept (in my terms, she can understand ‘red-feeling’ when she sees the tomato, but not when in her black-and-white cell). In the present context this is a bit of a distraction, which we can avoid by considering a variant of the case where Mary has been shown red objects and ostensibly taught the meaning of ‘red-feeling’: intuitively there will still be many red-feeling facts she cannot deductively infer from her physical knowledge. (On these two aspects of the Mary example see Loar 1997.)
13. This kind of objection (although put in slightly different terms) first appeared in Horgan 1984.
14. If this is to stand a chance of being true, ‘δ v μ’ and ‘(δ v μ) & ~δ’ (for example) had better not count as moral/descriptive sentences, respectively, since ‘δ ⊃ (δ v μ)’ and ‘(δ v μ) & ~δ ⊃ μ’ are a priori (cf. Prior 1960, 90–1). But in any case, all that is needed to produce conflict with cosmic hermeneutics is the weaker claim that some oughts are not derivable from an is.
15. It is defended, at least as a possibility worth taking seriously, in Brink 1989, ch. 6, and (I take it) endorsed in Boyd 1988.
16. Soames 1997; Horwich 1995; see also Byrne 1993, ch. 4. (Horwich’s discussion, rather misleadingly in my opinion, revolves around deflationary vs. inflationary conceptions of truth.)
17. There is a sense in which Jackson and Chalmers are sceptical of metaphysical necessity: they see it as “logical [necessity] with an a posteriori semantic twist” (Chalmers 1996, 38). But the scepticism I have in mind is some sort of Quinean view that rejects even the anodyne Jackson/Chalmers notion as unintelligible.
18. Other possible counterexamples concern indexicals, demonstratives, and various tensed constructions. It certainly does seem plausible that the entire supervenience base for the world could be given in a language without indexicals, demonstratives, and tenses. Assume that is so. Then ‘AB is a philosopher ⊃ I am a philosopher’ (as uttered by me); ‘AB is a philosopher ⊃ that man is a philosopher’ (as uttered by someone demonstrating me); and ‘The faculty meeting is starting at t ⊃ the faculty meeting is starting now’ (as uttered at t) all express necessarily true propositions (with a caveat about worlds where AB does not exist). Yet it might well appear that none of these propositions is a priori. However, if Kaplan’s (1989) theory of such terms is correct, as I think it more-or-less is, then appearances are deceptive. This is not the place to discuss these issues: if you prefer, imagine the demon situated in the actual world, and add indexicals to the physical language.

A further problem is generated by the apparent “realist” possibility of physical facts unknowable even by our hyper-idealized Laplacian demon. But this only affects the letter, not the spirit, of cosmic hermeneutics.

For more discussion of indexicals and other difficulties that I am not going to raise, see Chalmers 1996, 81–6.
19. A qualification is needed (see footnote 5 above, and footnote 34 below).
20. The ε-δ analysis, familiar from elementary calculus, of smoothness and continuity, and Turing’s analysis of effective computability, are good examples (due to David Lewis) against the sometimes-heard view that conceptual analysis is entirely fruitless.

21. For partial realization, see Lewis 1970, 82–3. For Lewis’s latest thoughts on multiple realization, see Lewis 1997, 334.

22. This “definition” of course has problems. The problem with multiple realizability is worth mentioning: if n and b are a certain nut and its accompanying bolt, then \( \langle \lambda x \ x=n, \lambda x \ x=b \rangle \) will satisfy the realization formula of the nuts-and-bolts theory; and if \( \text{’Rx’} \) expresses an uninstantiated property, and if \( \langle \lambda x \ (Px \lor Rx), \lambda x \ (Qx \lor Rx) \rangle \). In the mental case, Lewis’s solution to these sorts of problems relies on (i) the fact that folk psychology makes many causal claims, and (ii) a “sparse” theory of causally efficacious properties. Note that property designators defined using Lewis’s method will generally be non-rigid. For a way of extracting explicit definitions that makes property designators rigid, see Block 1980.

23. If we allow that T-terms may denote even though the realization formula of T is only partially realized, “is true” should be replaced by “is largely true”.


25. For an application to color vocabulary, see Lewis 1997. That paper is largely concerned to solve what Smith (1996, 48–54) calls the “permutation problem”: the apparent fact that if an n-tuple satisfies the realization formula of “folk chromatics”, any permutation of that n-tuple will also satisfy it. Smith, by the way, goes on to argue that a similar problem also afflicts moral functionalism.

26. A broadly Kripkean alternative is: the reference of \( \tau_i \) was originally fixed by something like Lewis’s “definitions”; the reference of that term as used subsequently by speakers is determined by a causal chain linking earlier uses to later ones; the original reference-fixing description is no part of the meaning of \( \tau_i \), and may well be unknown by many competent speakers. All this raises extremely difficult issues, but it is not necessary to pursue them here.

27. There is no doubt that some idealization is required (not that Lewis supposes otherwise). First, not every (in fact, hardly any) scientific term is a name. Second, scientific terms are not typically introduced by means of postulates—there is usually some gesture at explicit definition. For example, the OED reports the first (1891) occurrence of ‘electron’ in the following context: “A charge of this amount is associated in the chemical atom with each bond... These charges, which it will be convenient to call electrons, cannot be removed from the atom; but they become disguised when atoms chemically unite”. Third, even if ‘electron’ was introduced by means of a postulate, it is certainly not our “only clue”—historians excepted, we have little idea what the postulate is, for the relevant theory has changed substantially over the last century or so (for some discussion of this last point, see the end of Lewis 1970).

28. Indeed, in the case of moral functionalism, the reductive definitions must be complex, to avoid falling foul of (one interpretation of) Moore’s “open question” argument (Jackson 1998a, 150–3).

29. It is worth pointing out that Sellars’s term-introducing myth is not, in fact, as austere as Lewis seems to imply. In “Empiricism and the Philosophy of Mind” Sellars invites us to “[i]magine a Rylean language, a language of which the fundamental descriptive vocabulary speaks of public properties of public objects located in Space and Enduring through Time” (1997, 91). The speakers of this language are our Rylean ancestors.
But Sellars does not say that they invent folk psychology. It is our “Neo-Rylean” ancestors who do that. And they speak the Rylean language enriched “with the fundamental resources of semantical discourse—that is to say, the resources necessary for making such characteristically semantical statements as “‘Rot’ means red,” and “‘Der Mond ist rund’ is true if and only if the moon is round”” (1997, 92). So, in Sellars’s Neo-Rylean myth, the O-vocabulary has semantic terms. Therefore, assuming this myth to be a good one, our mentalistic vocabulary is implicitly defined by a (partly) semantic vocabulary. Hence we have simply exchanged one sort of non-physical vocabulary for another. Pending a way to directly analyse linguistic meaning in physicalistically acceptable terms, this gets us no further forward.

This can be turned into an objection against Lewis. On the one hand, there is Lewis’s myth of our Rylean ancestors, who speak only of physicalistically acceptable matters. And on the other hand, we have Sellars’s myth of our Neo-Rylean ancestors, who speak partly of semantics. Clearly Lewis needs, and of course supposes, that the first myth is a good myth. And if the first myth is a good myth, the second must be as well: enriching the original Rylean language is not going to impede any ancestral theorist. Therefore the second myth is at least as good as the first, and in fact surely more so—the more resources our mythical ancestors have, the more likely it is that they can implicitly define our mentalistic vocabulary. But the second seems to explain the “odor of analyticity” and the plausibility of behaviourism just as well as the first. So why isn’t the cautious and proper conclusion (at best) that only the second myth is a good one?

30. McGinn (1980) objects to Lewis along the following lines. Here is another myth—the myth of our Russellian ancestors, who spoke only of sense-data. Then some genius invented the theory of material objects, with its newly introduced T-terms, to explain the regularities among our sensory data. The myth explains the odor of analyticity about the platiitudes of commonsense talk of material objects and perception and the plausibility (it does have some) of idealism. But surely this is not a persuasive argument for analytic phenomenalistm. So the form of Lewis’s argument proves too much.

Now Lewis could well say, as McGinn notes, that this is a persuasive argument for analytic phenomenalism. So the physical can be analyzed in terms of the mental, and vice versa: victory all round!

Whether or not Lewis would make this reply, I myself do not think—contra commonsense functionalism—that the platitudes connecting material objects and perception are at all analytic.

31. There are also some technical difficulties in making the proposal work for verbs like ‘believes’. For these and other problems, see Byrne 1993, ch. 3.

32. As far as commonsense functionalism goes, the other main argument for it in the literature is simply a reply to the objection that there are not enough folk-psychological platitudes to define mental terms. (For the objection see, e.g., Schiffer 1987, 29–31; for the reply, relying on the idea that a lot of the platitudes might be implicitly known, see Jackson and Pettit 1990, 34–6, and Braddon-Mitchell and Jackson 1996, 55–8.) But replying to an objection doesn’t amount to much of a positive argument.

Turning to moral functionalism, it was first suggested, although not under that name, in Jackson 1992. Some considerations in favour of this theory are supplied in a later paper (Jackson and Pettit 1995), although it seems to me that they are unpersuasive. And Jackson and Pettit might not violently disagree with this assessment: at the end of the paper, they write that “we have done very little by way of meeting possible objections; and that we have done nothing to dislodge any of the alternative doctrines
that currently do battle in the meta-ethical field. Our aim has been to float the functionalist idea, not to establish definitively that it is sea-worthy” (39). Jackson’s 1998a, however, gives the impression that moral functionalism has passed its sea-trials with flying colors.

33. At any rate this is right if we ignore—as we are doing until 2.4—complications induced by Kripkean a posteriori necessity. Given this simplification, Chalmers holds that the following are equivalent: α’s “truth is ensured by the meanings of the concepts involved” (1996, 52); α is logically necessary (52); ‘~α’ is inconceivable (66–8); α is a priori (68–9). Obviously Chalmers does not take ‘logically necessary’ to mean: valid (in some formal system) (see 35, 52). The complications are discussed in Chalmers’ ch. 2; he deals with them using the two-dimensional framework explained in 2.41 below.

34. The exceptions to “most high-level facts” are those with a “dependence on conscious experience”; “[p]erhaps the best way to phrase [Chalmers’ view] is to say that all facts supervene logically on the combination of physical facts and phenomenal facts, or that all facts supervene logically on the physical facts modulo conscious experience” (Chalmers 1996, 71–2).

35. I should confess that I do not properly understand what “conceivability” is supposed to be: the claimed equivalence between α’s being conceivable and ‘~α’’s being not a priori (see footnote 33 above) would be a perfectly good explanation if it were a definition, but it isn’t. Fortunately my dim grasp of the notion will suffice here.

36. There are other problems. (2) is presumably only inconceivable if conjoined with a posteriori information about the actual world, in particular that the wombats around these parts are creatures of entirely physical composition. But then (3) does not follow from the inconceivability of this conjunction. I stress that I am not attributing this bad argument to Chalmers; the argument he plainly has in mind is discussed immediately below.

37. Following Jackson, I shall not fuss over whether ‘Water is H₂O’ is true at worlds where there is no water (similarly, mutatis mutandis, for other Kripkean examples). If you prefer, replace this sentence with ‘If water exists, then it is H₂O’.

38. This is not to say that (7) is unobjectionable. For arguments against any instance of it, see Soames 1989.

39. See also Segerberg 1973; Stalnaker 1978; Lewis 1980; Davies and Humberstone 1980; Tichy; 1983. It should be emphasized that although Stalnaker’s work is significantly responsible for inspiring two-dimensionalism as explained here, he does not endorse it himself (see Block and Stalnaker forthcoming).

40. For the second alleged aspect restored, see footnote 43 below. Of course, it would be sheer confusion (of which, I hasten to add, Jackson is entirely innocent) to take the claim that the Twin Earthlings’ word ‘water’ refers to XYZ to imply that the English word ‘water’ is indexical or otherwise context-dependent.

41. The role Jackson intends this assumption to play in his argument is not completely clear to me. See footnote 55 below.

42. The proposition that α expresses a true proposition should not be confused with the diagonal proposition associated with α (pace Salmon 1986, ch. 6, n3). Letting α = “Water covers most of the Earth”, the former proposition, but not the latter, is false at a world where the watery stuff covers most of the Earth and where ‘Water covers most of the Earth’ is used to express the proposition that 2 + 2 = 5. Again, as we’ll see in a few paragraphs, the diagonal proposition associated with ‘Water is the watery stuff’ is necessary; but plainly the proposition that ‘Water is the watery stuff’ expresses a true proposition is contingent.
43. Jackson calls the diagonal proposition ‘the A-proposition’ (1998a, 76), Chalmers calls it ‘the primary proposition’ (1996, 63–4). A similar account is in Tichý 1983, although Tichý does not mention the two-dimensional apparatus (and clearly does not think of propositions as sets of worlds). The proposition that, according to Tichý, determines the diagonal proposition (but not conversely) is called ‘the proposition associated with the sentence’ (231). (For more on Tichý’s account, see footnote 45 below.)

In the text we have made the simplifying assumption that intra-world utterances of ‘water’ do not differ in reference. Consider utterances of ‘Water covers most of the Earth’ on Earth and Twin Earth (in the same world w), and drop the assumption. There is now no single proposition expressed by that sentence, considering w as actual. To account for this added complexity within the two-dimensional framework, the matrix row-headings need to name centered worlds (a pair of a world and a context); thus one row-heading will be ‘(w, Earthly context)’, another will be ‘(w, Twin-Earthly context)’. The corresponding rows will give the propositions expressed, respectively, by the Earthlings and Twin Earthlings. But now there is no such thing as the diagonal proposition associated with ‘Water covers most of the Earth’: instead of a set of worlds, we have a set of world-context pairs. Because the criticism of Jackson below can be stated without giving up the simplifying assumption, we can leave the matter here.

44. Jackson says that “it is the [diagonal proposition] we know in virtue of understanding a sentence” (1998a, 76), and “understanding the sentence only requires knowing the [diagonal proposition]” (77), but obviously he is not using ‘knowing’ here in the factive sense; rather, he means something like ‘grasping’. From what he says elsewhere he clearly holds that understanding ‘Water covers most of the Earth’ requires knowing (11).

45. As mentioned in footnote 44 above, Tichý does not adopt the possible worlds conception of a proposition. So why does he think that there is in such thing as a necessary a posteriori proposition? He considers two of Kripke’s examples: ‘Phosphorus is Hesperus’ and ‘Heat is molecular motion’. Regarding the first, he says that (on Kripke’s view) both ‘Phosphorus’ and ‘Hesperus’ “are connotationless proper names of Venus... The utterer [of ‘Phosphorus is Hesperus’]...imputes self-identity to Venus...what has been asserted is a necessary truth. Now it seems equally obvious that this same truth is also knowable a priori” (232). This is a good argument (cf. Salmon 1986, 135–8). Regarding the second, he says that “Kripke takes the view that the semantics of the term ‘heat’ is rather like that of ‘Phosphorus’...’heat’ names molecular motion... Thus on Kripke’s theory, all that [‘Heat is molecular motion’] says is that molecular motion is molecular motion...[which is] not only necessary but knowable a priori” (234). But this argument is mistaken. Suppose ‘N’ is a name of M, and suppose that ‘M’ is rigid. Then ‘N’ is ‘M’ is necessary. But it does not follow that ‘N is M’ expresses the same proposition as ‘M is M’. For it might be that ‘M’ is not a name of M, but instead a complex rigid referring expression (like ‘molecular motion’ or ‘the actual inventor of the zip’). Thus, granted that the proposition expressed by ‘M is M’ is a priori, it does not follow that the proposition expressed by ‘N is M’ is also a priori. (There are some complications here, because it is not clear just what Kripke means by saying that general terms like ‘heat’ are rigid: see Soames 1998b.)

46. Stalnaker (if we trample over any worries he might have about the a priori) is an example of someone who rejects the first and retains the second.

47. Changing Jackson’s numbering throughout.
48. Jackson’s (13a) is in fact ‘H₂O is the watery stuff of our acquaintance’. In keeping with our simplifying policy of ignoring the presumed intra-world variation in the reference of ‘water’, I have omitted ‘of our acquaintance’.

49. Omitting ‘of our acquaintance’. See the previous footnote.

50. There is a complication here. A two-dimensionalist might well say that in normal contexts, an utterance of ‘NN knows that water covers most of the Earth’ reports that NN stands in the belief relation, not to the proposition expressed by ‘Water covers most of the Earth’, but to the diagonal proposition associated with it (cf. Stalnaker 1986, 73–4; 1978; 1981; Jackson 1998a, 76). What’s even worse, a two-dimensionalist might further hold that an utterance of the proposition expressed by ‘Water covers most of the Earth’ in normal contexts refers to the diagonal proposition associated with ‘Water covers most of the Earth’! But I trust my meaning is clear.

51. All this suggests that Jackson is concerned primarily to argue that if physicalism is true cosmic hermeneutics in a revised sense is possible, with the revision of the demon’s resources and mission being something like:

The demon understands the language of physics and English (perhaps: any language). He knows everything expressed by true physical sentences and which physical sentences are true. His task is to infer deductively which English sentences (perhaps: which sentences of any language) are true.

No problem if so, for we shall see that Jackson’s argument fails to establish that cosmic hermeneutics in either this sense or my official sense is possible if physicalism is true.

52. See also Yablo forthcoming for more discussion, and a similar complaint.

53. For familiar reasons, it is a hard row to hoe. It is hoed about as well as it could be in Stalnaker 1984. For discussion of some of the familiar reasons, see Schiffer 1986; Field 1986; and Stalnaker 1986. (This is not to say, of course, that the identification cannot be an illuminating simplification for many theoretical purposes; quite the contrary, as the work of, in particular, Lewis and Stalnaker has convincingly shown.)

54. Since (C) is equally plausible for ordinary rational beings like ourselves, at any rate on a broad reading of ‘can deductively infer’, we can replace ‘the demon’ by ‘someone’. Then the instances of the revised (C) and the possible worlds conception of a proposition yield:

(C*) For all propositions P, Q, if someone knows P, and Q is a necessary consequence of P, then he can deductively infer (and thereby know) Q.

The problem is to stop (C*) leading to the stronger conclusion:

(C+) For all propositions P, Q, if someone knows P, and Q is a necessary consequence of P, then he knows Q.

The way to block (C+)—that knowledge is closed under necessary consequence—is to defend the view that deductive inference is a process: by reasoning deductively, one can come to know something one did not know before, namely a necessary consequence of what one knows. If that’s right, then of course one may hold (C*) without
holding \((C^+)\). For the latter says that one *simultaneously knows* the necessary consequences of what one knows. Now the reason why deductive inference *seems* to be a process is that proving something from a set of sentences certainly *is* a process. However, on the possible worlds conception of a proposition, proving something from a set of sentences is not all it appears to be: in the case of mathematics, in particular, it cannot be described as *coming to know* the proposition expressed by the sentence at the last line of the proof, for that proposition is the necessary proposition, which the person producing the proof already knows (cf. Stalnaker 1984, 24–5). Thus, the possible worlds conception of a proposition removes the obvious reason for thinking that deductive inference is a process, and so, once \((C^+)\) is in place, \((C^+)\) threatens. That is why (instances of) \((C)\) and the possible worlds conception of a proposition together support the closure of knowledge (and belief) under necessary consequence.

55. It may well not be Jackson’s strategy, because there is textual evidence to suggest that he sees the possible worlds conception of a proposition not as an important assumption, but merely a useful simplification. In his 1992 and 1994b the argument is stated without mentioning propositions at all. In his 1998a he seems to imply that his practice of calling sets of possible worlds ‘propositions’ is just a terminological stipulation, not a substantive thesis about, e.g., the objects of belief (see the quotation at the beginning of 2.41 above; 76; and 76, n32). In his 1994a there is the remark: “First, the issue is an issue about sentences...and not about propositions, or at least not propositions thought of as sets of possible worlds” (37, my italics; see also 1994c, 489), the implication apparently being that there might well be other equally good ways of thinking of propositions.

To complicate these exegetical matters, Jackson’s own view is that propositions (the objects of belief and the referents of that-clauses) are sets of possible worlds (Braddon-Mitchell and Jackson, 190–5).

56. And in fact, the appearance of chemical vocabulary in B is somewhat misleading. It follows from (13) and (13d) that the watery stuff covers most of the Earth, and it follows from *that* that water covers most of the Earth. If our Laplacian demon knows (13), we can give him knowledge of (14) by telling him (13d). But equally, we could have told him straight off that the watery stuff covers most of the Earth.

57. Chalmers also uses the two-dimensional apparatus to argue that physicalism implies that cosmic hermeneutics is possible (1996, esp. 65–70, 131–8; see also forthcoming). But his argument is more dialectically complex, so for reasons of space I have chosen only to discuss Jackson’s argument from two-dimensionalism here (Yablo forthcoming has a nice comparison of Jackson and Chalmers). I examine Chalmers’ version in Byrne 1998.

58. There are three main objections to the rigidified-description theory (take ‘water’ as the example, and forget about the problem of purging the description of natural kind terms).

First, an objection (or more properly a class of objections) familiar from *Naming and Necessity* and various papers by Putnam (1962, 1970, 1975). For example, surely a speaker may use ‘water’ with its customary meaning, even if the descriptive content he associates with the word is scant or is in fact false of water. (Jackson at one point offers ‘the stuff which actually falls from the sky, fills the oceans, is odourless and colourless, is essential for life, is called ‘water’ by experts,..., or which satisfies enough of the foregoing’ (1994a, 39). But the “experts” may well call water ‘H2O’, not ‘water’, and in any case, putting the word itself into the descriptive content has the undesirable consequence that the semantics for the English word ‘water’ differs from
that of the French ‘eau’.) (For a defence against these sorts of objections, see Jackson 1998b.)

Second, an objection due to Soames (1998a, 14–6). ‘Actual’, in its rigidifying use, is an indexical like ‘I’. I believe I am a philosopher, and you might believe that very proposition (you wouldn’t express it using ‘I’, of course). Could someone on a remote planet also believe that proposition? Surely not—such a person has had no contact with me at all. Now, some of us believe that water is wet. And some inhabitants of remote possible worlds also believe that very proposition. On the rigidified-description theory, for someone in a world w to believe that proposition it is necessary that he believes something about our world. But, especially if we imagine w to be remote from our world, how could he have such a belief? (Cf. the analogous example of ‘I’.)

We have been making the simplification (by the rigidified-description theorist’s lights) that intra-world tokens of ‘water’ do not differ in reference. The third objection is that it is hard to see how this can be unproblematically removed. Jackson’s suggested description in his 1998a (see footnote 49 above) is ‘the watery stuff of our acquaintance’, with the prepositional phrase removing the simplification. But that won’t work, for it falsely implies that an Earthly traveller visiting Twin Earth on a day trip will speak truly when he utters ‘Water covers most of Twin Earth’s surface’ (cf. Burge 1982, 103–7). ‘The watery stuff we were originally acquainted with’ is no better either, for it falsely implies that Earthly settlers on Twin Earth will never speak truly when they utter ‘Water covers most of Twin Earth’s surface’. What about ‘the stuff that was the (right kind of) causal origin of our use of the word ‘water’’? That is another of Jackson’s suggestions (1994a, n25); it suffers from a defect noted in the second paragraph above. And there are other problems. For example, ‘the right kind of causal origin’ can either be understood as a schema, ‘right kind’ to be filled in by your favorite causal theory of reference, or else the description can be taken to be something like ‘the stuff that caused (in the reference-determining way, whatever that is) our use of ‘water’”. But the former can be dismissed: no one has any idea how ‘right kind’ can be filled in correctly, so the correct filling can hardly be part of the semantics of a word we all understand. And the latter is circular (cf. Kripke 1980, 68–70).

A final point. If the desired conclusion is that cosmic hermeneutics in the full-blown sense is possible if physicalism is true (where we do not restrict the facts the demon has to know to facts expressible in a natural language like English), then it has to be argued, not merely that English contains no Kripkean natural kind terms, but that they are not part of any possible language (at any rate any possible language that could be understood).

References


Byrne, A. 1998. “Chalmers’ two-dimensional argument against physicalism.” MS.


Latham, N. Forthcoming. “Chalmers on the addition of consciousness to the physical world.” Philosophical Studies.


Soames, S. 1998b. “What is it for a general term to be a rigid designator?” MS.