Lewis on Reduction of Mind

What is Reductionism for Lewis?

As Lewis notes, many people strive to reject the label 'reductionist'. He says 'thousands' aim to be rid of the label, and while that's probably an exaggeration, it is true that the word isn't popular. He prefers to keep the term, and the old-fashioned term 'materialist', and I'll follow him in both respects. Since 'reduction' and its variants are technical terms, we can't hope to do any conceptual analysis on them. And it is foolish to aim to criticise or complain about someone's usage, except perhaps to complain that it is an unhelpful usage. But we should try and figure out what Lewis's reductionism amounts to.

Lewis says that he holds an *a priori* reductionism. The modifier 'a priori' is a little hard to figure out here. Lewis doesn't think that it is a priori just what mind reduces to. But he does think that the following principles are a priori.

(A) The facts at a world supervene on the distribution of fundamental properties and relations.

(M) Mental properties and relations are not fundamental.

(A) is just our old friend the Armstrong constraint. The argument for (M) isn't quite as clear, but I think it goes something like this.

1. Necessarily, fundamental properties and relations are not multiply realised.
2. Possibly, mental properties and relations are multiply realised.

C. So mental properties and relations are not fundamental.

This looks like a fairly compelling, and a fairly Lewisian, argument to me. So I'll assume that he'd accept it. This gets us to the conclusion that the mental features of the world supervene on the fundamental features of the world, without being fundamental features of the world. If we add the extra thesis of materialism, which for present purposes can be treated as the thesis that all the fundamental properties and relations are *physical*, we get that the mental properties supervene on the physical properties, without being physical properties.

This is a slightly odd way of putting things. Lewis is a materialist, so you might think he would want to deny that there are any non-physical properties that are instantiated. That's true under one reading, false under another. Consider the property F $=_{\text{df}}$ *being an electron or a ghost*. That's not a physical property in the sense that not necessarily everything that instantiates it is physical. But it is instantiated. What's important for materialism is that (a) the property is not fundamental, and (b) everything that actually instantiates it is physical. Mental properties for Lewis are like the F. They aren't purely physical properties in the sense that necessarily their pattern of instantiation supervenes on the distribution of physical properties. But they are consistent with materialism because (a) they are not fundamental and (b) they are only actually instantiated by material objects.

The fact that some things instantiate F is no reason to give up on materialism. Nor is the fact that F is not identical to any purely physical property in the above sense a reason to give up on *reductive* materialism. Whatever reductive materialism is, it should not be the doctrine that properties like F are never instantiated.

The supervenience of the mental on the physical is widely accepted. So is the claim that the mental properties are not purely physical in the sense I've presented. But many philosophers claim that these claims are not sufficient for *reductive* materialism, so we should find out what is intended to be the extra reductive claim. The discussion on pages 296-7 suggests something like the following condition.

(D) The mental facts about the world are *a priori* implied by the material facts.

I've called this (D) for deducibility. Now this is never made particularly unambiguous, but it seems to me that it might mean any of (D1), (D2) or (D3), or perhaps something else again. (In what follows, P refers to the physical facts, including a 'that's all' clause, and M the mental facts.)

(D1) Necessarily, *If P then M* is a priori knowable.
(D2) It is a priori that necessarily anyone who knows P is in a position to know M.
(D3) Necessarily, anyone who knows P is in a position to infer M using only premises that are a priori knowable.

(D1) implies (D2) given a weak closure principle. And (D2) implies (D3) fairly obviously. But it isn't clear that the reverse implications hold. If the person who learns P thereby learns some a posteriori epistemology, they might be able to conclude that they are warranted in believing M, although it is not a priori that knowledge of P warrants belief in M. So (D3) could be true without (D2) being true. It is harder to come up with a plausible epistemology on which (D2) is true but (D1) is false, but I don't think all such epistemologies are incoherent. So I'm inclined to think that these three claims are all distinct.

It might be objected here that my claim that (D3) is weaker than (D2) turns on an incoherent 'possibility'. I've claimed that a knower might learn some epistemological fact, namely that P warrants M, a posteriori, even though that is necessarily true. Lewis certainly doesn't have a position for that. He is a *modal rationalist*, who holds that in some sense all necessary truths are a priori knowable. Now modal rationalism does have something to be said for it. But it should not be assumed in the statement of reductionism. So we shouldn't *assume* that all necessarily true propositions are a priori knowable.

Lewis's modal rationalism follows from his view of propositions as unstructured entities, in particular as sets of possible worlds. (I want to bracket for now the question of how much of the modal rationalism one could reject while holding on to propositions as unstructured entities. All I want to stress is that for Lewis the two doctrines seem closely connected.) If propositions are sets of possible worlds, then there is a false plural in my talk about necessarily true proposition*s*. There is only one necessarily true proposition. Any agent who knows that they are self-identical knows that proposition. But plausibly to be a knower, one must know that one is self-identical. So all knowers know *the* necessarily true proposition, and hence know all necessarily true proposition. So unstructured propositions seem to lead to modal rationalism.

Many contemporary philosophers reject modal rationalism because of Kripke's examples of the necessary a posteriori, such as *water is molecular*. Lewis, following Frank Jackson, claims that these examples pose no challenge to (D2), or even to (D1). The argument he gives has a small hole in it, and seems to presuppose that propositions are unstructured. So it's worth working through this in some detail.

Let P again be the conjunction of all the physical facts, and W some claim about water, say that water covers most of planet earth. Then the worry is that although *If P then W* is necessarily true, it isn't a priori. Let H be the claim that H_2O covers most of planet earth. Perhaps it is a priori that if P then H. But we need an extra step to conclude a priori that the sentence "If P, then W" is true.

Here, in broad outlines, is the extra step. We can't say that the step is If H, then W. That is true, even necessarily true, but it is a posteriori. But consider the proposition If P and H, then W. That will be a priori true. The reason is that P tells us that H_2O plays the water role, so if P is true, and H_2O covers most of the earth, then water must cover most of the earth. So we can use the following argument, where each premise is a priori knowable.

1. If P, then H.
2. If P and H, then W.

C. So, If P then W.

The problem is that it really isn't obvious that 2 is a priori. We should agree that 2 is necessarily true for the reasons above. But it is hard to say whether that argument really shows that 2 is a priori. It might be necessary a posteriori that if P is true, then H_2O plays the water role. If that's true, then 2 will be necessary a posteriori, and so if P then W will be necessary a posteriori.

A similar problem afflicts the argument Lewis offers on page 297. Consider the part where Lewis tells us to consider a proposition Q that is true where ϕ expresses the same horizontal proposition that it actually does. There are two problems with the argument that follows. First, if we don't believe that propositions are sets of possible worlds, then we might worry that Lewis hasn't picked out a unique Q. Second, although Lewis argues convincingly that given materialism, P necessitates Q, he doesn't offer much by way of argument that *If P then Q* is a priori knowable. And that's what he needs for the argument to go through.

So I'm tempted to conclude, tentatively, that Lewis hasn't given us a good reason to believe that (D1) is true. So if reductive materialism is to be identified with claims like (D1), he hasn't given us reason to be reductive materialists.

Boiling and Reduction

But Lewis's discussion also shows a way out of this problem. Lewis twice appeals to an analogy between facts about the mental and facts about boiling. We should exploit this to get a minimal condition on a sensible account of what a reductive theory is. The constraint is that however we are to use the term 'reducible', the sentence "facts about boiling are reducible to fundamental physical facts" should turn out true.

Can we say something more specific than this? It seems clearly true that the boiling facts supervene on the physical facts, but not vice versa. So our constraint does not rule out the definition of reduction as asymmetric supervenience. Perhaps there are reasons to hope for something more.

The constraint does rule out a definition of reduction as *identity* between the properties in the reduced discourse and properties in the fundamental base. *Boiling* is not identical to any physical property. We can see this by noting that even on Twin Earth, where the physics is very different to our actual physics, kettles can still boil. I'm pretty sure that anyone who denies that kettles, or at least what is in them, boil on Twin Earth is in the grip of a (bad) theory. Anyone who denies that there are kettles on Twin Earth is really is the grip of a bad theory. (I suspect that last sentence is a null quantification!) If boiling was identical to some physical property, then it would be impossible for kettles to boil on Twin Earth, or for cauldrons to boil on Magic Earth. But these things are possible, so boiling is not identical to a physical property. Since boiling is reducible to physics, this means reduction does not require identity.

Similar reasons tell against saying that reduction is incompatible with multiple realisation. Boiling is multiply realised across worlds, but it is actually reducible to fundamental physics. It might be argued that reducible quantities must not be multiply realised in the same world. Even this seems a stretch to me. Imagine it was discovered that what constituted heat in gases was different to what constituted heat in liquids or solids. (I don't believe this is quite true, though there appear to be several respects in which laws concerning heat have to be separately stated for solids and gases.) It would be absurd to conclude that we should be non-reductive materialists about heat.

We might say that reduction requires a priori entailment of the reduced theory by the basic theory, as above. This runs into several problems. If a priori entailment means something like (D3), then there is no reason to doubt that the mental is reducible in this sense. Even if a priori entailment means something like (D1), if the broadly ‘two-dimensionalist’ account of the mental is correct, then there is little reason to doubt that the mental is reducible to the physical. If a priori entailment means something like (D1), and the broadly ‘two-dimensionalist’ account of the mental is not correct, then in all probability we’ve got ourselves a theory of reduction in which boiling does not reduce to the physical, which is absurd.

We might say that if a certain subject matter reduces to a more fundamental subject, then the laws of the reduced subject must be derivable from the more fundamental subject. It isn’t obvious that psychology has any real laws, but let’s be generous here and call any counterfactual supporting generalisation a law. Now all the ‘laws’ of psychology do supervene on the physical facts, so this doesn’t look like a reason to deny that the mental supervenes on the physical. If we insist that the laws of the reduced subject matter must supervene on just the laws of the fundamental subject matter, then we have a sense in which the mental is not reducible to the physical. But now we have a sense in which statistical thermodynamics does not reduce to fundamental physics, since to get the laws of statistical thermodynamics, we don’t need to know just the laws of fundamental physics, but also something about the initial condition of the universe. And it is absurd to think that statistical thermodynamics does not reduce to fundamental physics.

We could go on like this, but the moral seems fairly clear. There is no interesting sense of reduction that is (a) significantly stronger than supervenience and (b) plausible given that we want to say that other parts of physics are reducible to fundamental physics.

Attack on Causal Functionalism

The first half of “Reduction of Mind” ends with an attack on the theory that is sometimes called ‘functionalism’. I think the view he’s attacking is the one set out here by Jerry Fodor.

> [I]t would be quite mad to say that *being an airfoil* is causally inert. Airplanes fall down when you take their wings off; and sailboats come to a stop when you take down their sails. Everybody who isn’t a philosopher agrees that these and other such facts are explained by the story about lift being generated by causal interactions between the airfoil and the medium. If that *isn’t* the right explanation, what keeps the plane up? If that is the right explanation, how could it be that *being an airfoil* is causally inert? (Fodor “Making Mind Matter More” (1989), 62)

Lewis wants to reject this. If the physical characteristics of the airfoil are P, then Lewis says it’s that the airfoil is P that causes it to stay aloft, not that it’s an airfoil. The reason he gives is that causes are not disjunctive, and it would be a disjunctive cause if being an airfoil caused the plane to stay aloft. At least that's what Lewis is aiming for. The details are actually quite messy.

Lewis of course is talking about mental properties, particularly pains, not airfoils. And here is what he says.

> Since the highly disjunctive property of being in *M* does not occupy the *M*-role, I say it cannot be the referent of *M*. Many disagree. They would like it if *M* turned out to be a rigid designator of a property common to all who are in *M*. So the property I call 'being in *M*', they call simply *M*; and the property that I call *M*, the occupant of the *M*-role, they call 'the realisation of *M*'. They have made the wrong choice, since it is absurd to deny that *M* itself is causally efficacious. Still, their mistake is superficial. They have the right properties in mind, even if they give them the wrong names. (r307)

This is all strange on several levels. For one thing, Lewis's position seems to be subject to some fairly straightforward counterexamples. For another, it seems unnecessary to ensure his causal aims.

The following problem is based on some points made by Eric Hiddleston at the recent APA Pacific. Assume that *M* is 'pain', and assume that a particular koala is in pain in virtue of being in physical state P_K, and a human is in pain in virtue of being in physical state P_H. Now what is Lewis's account of the semantics of sentences like (1)?

(1) That koala is in pain.

Presumably he thinks that 'pain' here doesn't refer to *being in pain*, i.e. the highly disjunctive property in common to all of the creatures that are in pain. Rather it refers to P_K. So the logical form of it is something like (2), where *a* denotes the koala. (I'll ignore here the complications to do with the logical form of the complex demonstrative.)

(2) $P_K a$

Now if we consider (3), we should think that its logical form is (4), where *b* denotes the human.

(3) That koala is in pain. So is that human.
(4) $P_K a$. $P_K b$.

The reason this should be the logical form is that in general when we say that *a* is *F*, and so is *b*, then we've said that *b* is *F*. But that's the wrong result here, since the human is in pain, and $P_K b$

is false. So something has gone wrong. There is a way out of this for Lewis, but it involves reducing somewhat the gap between Lewis and Fodor. Lewis could say that we should read (3) as having a similar logical form to (5).

(5) That koala loves her mother. So does that human.

In (5), 'her mother' denotes the koala's mother, but (5) doesn't (unambiguously) say that the human also loves the koala's mother. Rather, what it says is that the human loves her own mother. So perhaps we should say that (3) is best spelled out in English as (6).

(6) That koala is in pain *for her*. So is that human.

We'll then say that 'pain for her' in the first sentence denotes P_K, just as Lewis wants. But it is a logically complex expression, and it includes a bound variable that can take a different value when we say that someone else is in the 'same' state. (This can all be said formally using lambda expressions, but I'll leave out that level of formalism.) Given that Lewis says that pain-in-K is the fundamental expression kind, perhaps this is a preferable way to the above of setting out Lewis's own view. It still seems to have problems to me.

First, it doesn't look distinctively different from the Fodorian position that Lewis is criticising. Lewis's opponent says that 'pain' denotes the diagonal property of being in the state that plays the pain role for your kind. Lewis says that 'pain' denotes the property of being in the state P, where state P is the state that plays the pain role. You might think that there is a difference here, and it would show up in embeddings under subjunctives. For instance, you might think that Lewis's opponent would assert, while Lewis would deny, (7).

(7) If state Q played the pain role in humans, and that human was in Q, she would be in pain.

But actually Lewis *endorses* (7). That's because he says that 'pain' is a non-rigid designator. So I'm not entirely sure that the position I'm attributing to Lewis here is that different to his opponent's position.

Second, if there is a difference between the position I'm attributing to Lewis and his opponent's position, it is that Lewis posits an extra variable position, for kinds, in the logical form of pain attributions. But if anything this makes things better for Lewis's opponent than they are for Lewis, for two reasons. The simple reason is that there simply doesn't seem to be such a variable, so positing it is exorbitant. The complicated reason is that Lewis's way of doing the semantics suggests that (6) should be ambiguous, the way that (5) is ambiguous. One reading, perhaps not the preferred reading but a clearly available reading, of (5) is that both the koala and the human love the koala's mother. So (6) should have a reading where it says that both the koala and the human are in the state that plays the pain role in koalas. But that is not an available reading of (6). So I don't think there is much to be gained by this move.

Let's say then that 'pain' does denote the 'diagonal' property of being in the state, whatever it is, that plays the pain role in one's kind. Does that mean that we're committed either to the disjunctive diagonal property being causally efficacious, or that pains are not causally efficacious? No, because talk about which *properties* are causally efficacious is always disguised talk about which *events* are causally efficacious, and the relationship between the disguise and what's underneath might be rather complicated.

We'll start with a rather different example that Lewis uses. Consider sentence (8), in a circumstance in which it is spoken truly.

(8) The fall caused Smith's death.

Lewis wanted to say this could be true even before he had the causation as influence theory. In other words, he wanted to say that it could be true when he had the theory that causal required counterfactual dependence of the *occurrence* of the effect on (a chain of occurrences starting with) the occurrence of the cause.

Now Smith presumably is a mortal, so he would have died whether or not he fell. So if we say that *Smith's death* is an event that occurs in any world where and only where Smith dies, it won't be true that if the fall hadn't occurred, this event wouldn't have occurred. (We'll ignore the irrelevant possibility of a chain of counterfactual dependencies.) Lewis realised all this, and he had a solution to it. He said that *Smith's death* is a much more narrowly drawn event. It occurs where and only where Smith dies *this death*. The point isn't that somehow or other 'death' denotes a particular kind or way of dying. It is rather that the event that is Smith's death is an event that might not have occurred, even if Smith died. This all seems fairly plausible, and we should apply it back to cases of mental causation like (9).

(9) The koala is eating a lot because it is in pain.

In (9), 'pain' simply means *pain*, i.e. the property that is possessed by all and only creatures that are in pain for their kind. But the event of the koala being in pain is *not* the event that occurs when and only when the koala is in the state that plays the pain role for koalas. Rather, it is the event that occurs when and only when the koala is in P_K.

Now we can say what it is for a property to be causally efficacious. A property F is causally efficacious if events that occur when and only when a particular object is F are causes. It doesn't matter just how we'd *describe* that event on various occasions.

This seems like 'splitting the difference' between Fodor and Lewis. Fodor wants to say that the plane stays aloft because it's wing is an airfoil. Lewis wants to deny that the event that occurs when and only when the wing of an object is an airfoil is a cause. These two claims are not contradictory, provided we think that the event of the plane's wing being an airfoil is not the event that occurs when and only when its wing is an airfoil. This seems a slightly odd thing to think, until we see that it's exactly what we have to (and naturally did) think about deaths.

No More Platitudes

Lewis still accepts the Ramsey sentence approach to defining psychological terms. But there is one significant retraction to his earlier view. His early view was that Ramsifying over the folk psychological platitudes gave us the referents of the psychological terms. Now he thinks that we should instead Ramsify over folk psychology itself, rather than the platitudes people offer about it.

Since folk psychology is basically common knowledge, it might be wondered how much of a change this is. I think it's actually a reasonably large change. Although we all have tacit mastery of folk psychology, we may not always be able to clearly express what we know. The platitudes don't reflect what we know about each other's psychology, but rather the things we can easily express that we take to reflect that knowledge.

This distinction is fairly familiar from syntax. Every speaker of English knows, to some degree or other, its syntax. If they didn't, they would constantly be coming out with ill formed sentences, and not noticing when other people use ill formed sentences. But in fact this doesn't happen. So there is such a thing as the 'folk' syntax of English that is common knowledge.

Still, it would be a bad mistake to say that the meanings of the terms of syntactic theory are given by Ramsifying over the *platitudes* about syntax. Look at the platitudes about syntax and you'll come up with all sorts of strange things. I suppose we can, to a first approximation, take Strunk & White to embody the syntactic platitudes. But doing this *won't* give you a very good guide to the syntactic knowledge that we all have. Getting at this knowledge is quite hard – people in linguistics departments get paid very good money for their work at digging it out. The meanings of syntactic terms is given by Ramsifying over the theory that they discover. There isn't as much work (or money?) in digging out folk psychological knowledge. But if we ever do systematically set out folk psychology, not by collecting platitudes but by systematising the knowledge that people are, must be, acting on, that will be what we get the Ramsey sentence from.