

## 4

### D. Lewis, "Reduction of Mind"

I am a realist and a reductive materialist about mind. I hold that mental states are contingently identical to physical – in particular, neural states. My position is very like the 'Australian materialism' of Place, Smart, and especially Armstrong. Like Smart and Armstrong, I am an ex-Rylean, and I retain some part of the Rylean legacy. In view of how the term is contested, I do not know whether I am a 'functionalist'.

#### Supervenience and analysis

My reductionism about mind begins as part of an *a priori* reductionism about everything. This world, or any possible world, consists of things which instantiate fundamental properties and which, in pairs or triples or . . . , instantiate fundamental relations. Few properties are fundamental: the property of being a club or a tub or a pub, for instance, is an unnatural gerrymander, a condition satisfied by miscellaneous things in miscellaneous ways. A fundamental, or 'perfectly natural', property is the extreme opposite. Its instances share exactly some aspect of their intrinsic nature. Likewise for relations.<sup>1</sup> I hold, as an *a priori* principle, that every contingent truth must be made true, somehow, by the pattern of coinstantiation of fundamental properties and relations. The whole truth about the world, including the mental part of the world, supervenes on this pattern. If two possible worlds were exactly isomorphic in their patterns of coinstantiation of fundamental properties and relations, they would thereby be exactly alike *simpliciter*.<sup>2</sup>

It is a task of physics to provide an inventory of all the fundamental properties and relations that occur in the world. (That's because it is also a task of physics to discover the fundamental laws of nature, and only the fundamental properties and relations may appear in the fundamental laws.<sup>3</sup>) We have no *a priori* guarantee of it, but we may reasonably think that present-day physics already goes a long way toward a complete and correct inventory. Remember that the physical nature of ordinary matter under mild conditions is very well understood.<sup>4</sup> And we may reasonably hope that future physics can finish the job in the same distinctive style. We may think, for instance, that mass and

charge are among the fundamental properties; and that whatever fundamental properties remain as yet undiscovered are likewise instantiated by very small things that come in very large classes of exact duplicates. We may further think that the very same fundamental properties and relations, governed by the very same laws, occur in the living and the dead parts of the world, and in the sentient and the insentient parts, and in the clever and the stupid parts. In short: if we optimistically extrapolate the triumph of physics hitherto, we may provisionally accept that all fundamental properties and relations that actually occur are physical. This is the thesis of materialism.

(It was so named when the best physics of the day was the physics of matter alone. Now our best physics acknowledges other bearers of fundamental properties: parts of pervasive fields, parts of causally active spacetime. But it would be pedantry to change the name on that account, and disown our intellectual ancestors. Or worse, it would be a tacky marketing ploy, akin to British Rail's decree that second class passengers shall now be called 'standard class customers'.)

If materialism is true, as I believe it is, then the *a priori* supervenience of everything upon the pattern of coinstantiation of *fundamental* properties and relations yields an *a posteriori* supervenience of everything upon the pattern of coinstantiation of fundamental *physical* properties and relations. Materialist supervenience should be a contingent matter. To make it so, we supply a restriction that makes reference to actuality. Thus: if two worlds were physically isomorphic, and if no fundamental properties or relations alien to actuality occurred in either world, then these worlds would be exactly alike *simpliciter*. Disregarding alien worlds, the whole truth supervenes upon the physical truth. In particular, the whole mental truth supervenes. So here we have the common core of all materialist theories of the mind.<sup>5</sup>

A materialist who stops here has already said enough to come under formidable attack. An especially well-focused version of the attack comes from Frank Jackson.<sup>6</sup> Mary, confined in a room where all she can see is black or white, studies the physics of colour and colour vision and colour experience (and any other physics you might think relevant) until she knows it all. Then she herself sees colour for the first time, and at last she knows what it's like to see colour. What is this knowledge that Mary has gained? It may seem that she has eliminated some possibilities left open by all her previous knowledge; she has distinguished the actual world from other possible worlds that are exactly like it in all relevant physical respects. But if materialist supervenience is true, this cannot be what has happened.

Materialists have said many things about what does happen in such a case. I myself, following Nemirow, call it a case of know-how: Mary gains new imaginative abilities.<sup>7</sup> Others have said that Mary gains new relations of acquaintance, or new means of mental representation; or that the change in her is just that she has now seen colour. These suggestions need not be taken as rival alternatives. And much ink has been spent on the question whether

## “REDUCTION OF MIND”

these various happenings could in any sense be called the gaining of ‘new knowledge’, ‘new belief, or ‘new information’. But for a materialist, the heart of the matter is not what *does* happen but what *doesn't*: Mary does not distinguish the actual world from other worlds that are its physical duplicates but not its duplicates *simpliciter*.\*

Imagine a grid of a million tiny spots – pixels – each of which can be made light or dark. When some are light and some are dark, they form a picture, replete with interesting intrinsic gestalt properties. The case evokes reductionist comments. Yes, the picture really does exist. Yes, it really does have those gestalt properties. However, the picture and the properties reduce to the arrangement of light and dark pixels. They are nothing over and above the pixels. They make nothing true that is not made true already by the pixels. They could go unmentioned in an inventory of what there is without thereby rendering that inventory incomplete. And so on.

*properties of the picture = neural processes*

Such comments seem to me obviously right. The picture reduces to the pixels. And that is because the picture supervenes on the pixels: there could be no difference in the picture and its properties without some difference in the arrangement of light and dark pixels. Further, the supervenience is asymmetric: not just any difference in the pixels would matter to the gestalt properties of the picture. And it is supervenience of the large upon the small and many. In such a case, say I, supervenience is reduction. And the materialist supervenience of mind and all else upon the arrangement of atoms in the void – or whatever replaces atoms in the void in true physics – is another such case.

Yet thousands say that what's good about stating materialism in terms of supervenience is that this avoids reductionism! There's no hope of settling this disagreement by appeal to some uncontested definition of the term ‘reductionism’. Because the term *is* contested, and the aim of some contestants is to see to it that whatever position they may hold, ‘reductionism’ shall be the name for something else.

At any rate, materialist supervenience means that for anything mental, there are physical conditions that would be sufficient for its presence, and physical conditions that would be sufficient for its absence. (These conditions will include conditions saying that certain inventories are complete: an electron has only so-and-so quantum numbers, for instance, and it responds only to such-and-such forces. But it's fair to call such a condition ‘physical’, since it answers a kind of question that physics does indeed address.) And no matter how the world may be, provided it is free of fundamental properties or relations alien to actuality, a condition of the one sort or the other will obtain. For all we know so far, the conditions associated with a given mental item might be complicated and miscellaneous – even infinitely complicated and miscellaneous. But so long as we limit ourselves just to the question of how this mental item can find a place in the world of fundamental physics, it is irrelevant how complicated and miscellaneous the conditions might be.

It may seem unsatisfactory that physical conditions should always settle whether the mental item is present or absent. For mightn't that sometimes be a vague question with no determinate answer? A short reply to this objection from vagueness is that if it did show that the mental was irreducible to fundamental physics despite supervenience, it would likewise show that boiling was irreducible to fundamental physics – which is absurd. For it is a vague matter just where simmering leaves off and boiling begins.

A longer reply has three parts. (1) If the physical settles the mental insofar as anything does, we still have materialist supervenience. Part of what it means for two physically isomorphic worlds to be just alike mentally is that any mental indeterminacy in one is exactly matched by mental indeterminacy in the other. (2) Whenever it is a vague question whether some simplistic mental classification applies, it will be determinate that some more subtle classification applies. What's determinate may be not that you do love him or that you don't, but rather that you're in a certain equivocal state of mind that defies easy description. (3) If all indeterminacy is a matter of semantic indecision,<sup>8</sup> then there is no indeterminacy in the things themselves. How could we conjure up some irreducible mental item just by failing to decide exactly which reducible item we're referring to?

It may seem that when supervenience guarantees that there are physical conditions sufficient for the presence or absence of a given mental item, the sufficiency is of the wrong sort. The implication is necessary but not *a priori*. You might want to say, for instance, that black-and-white Mary really did gain new knowledge when she first saw colour; although what she learned followed necessarily from all the physics she knew beforehand, she had remained ignorant because it didn't follow *a priori*.

A short reply to this objection from necessity *a posteriori* is that if it did show that the mental was irreducible to fundamental physics, it would likewise show that boiling was irreducible to fundamental physics – which is absurd. For the identity between boiling and a certain process described in fundamental physical terms is necessary *a posteriori* if anything is.

(A longer reply, following Jackson, is founded upon the 'two-dimensional' analysis of necessity *a posteriori*.<sup>9</sup> Two-dimensionalism says that there is no such thing as a necessary *a posteriori* proposition. However, one single sentence  $\phi$  may be associated in two different ways with two different propositions, one of them necessary and the other one contingent; and the contingent one can be known only *a posteriori*. Suppose we choose to adopt a conception of meaning under which our conventions of language sometimes fix meanings only as a function of matters of contingent fact – for example, a conception on which the meaning of 'boils' is left dependent on which physical phenomenon turns out to occupy the boiling-role. Then if we interpret a sentence  $\phi$  using the meanings of its words as fixed in world  $W_1$ , we get proposition  $H_1$ ; using the meanings as fixed in  $W_2$ , we get  $H_2$ ; and so on. Call these the propositions *horizontally expressed* by  $\phi$  at the various worlds; and let  $H$  be the

proposition horizontally expressed by  $\phi$  at the actual world. The proposition *diagonally expressed* by  $\phi$  is the proposition  $D$  that holds at any world  $W$  iff the proposition horizontally expressed by  $\phi$  at  $W$  is true at  $W$ . So if we know  $D$ , we know that  $\phi$  horizontally expresses some truth or other, but we may not know which truth. Sentence  $\phi$  is necessary *a posteriori* iff  $H$  is necessary but  $D$  is knowable only *a posteriori*. Likewise, a proposition  $P$  necessarily implies that  $\phi$  iff  $P$  implies  $H$ ; but  $P$  *a priori* implies that  $\phi$  iff  $P$  implies  $D$ . Our worry was that when  $\phi$  was about the mind, and  $P$  was a premise made true by fundamental physics,  $P$  might imply that  $\phi$  necessarily but not *a priori*. But if so, and if you think it matters, just take another proposition  $Q$ : let  $Q$  be true at exactly those worlds where  $\phi$  horizontally expresses the same proposition  $H$  that it actually does.  $Q$  is true. Given the materialist supervenience of everything,  $Q$  as well as  $P$  is made true by fundamental physics.  $P$  and  $Q$  together imply *a priori* that  $\phi$ . So the gap between physical premises and mental conclusion is closed. Anyone who wants to reopen it – for instance, in order to square materialist supervenience with Mary’s supposed ignorance – must somehow show that the two-dimensional analysis of necessity *a posteriori* is inadequate.)

If we limit ourselves to the question how mind finds a place in the world of physics, our work is done. Materialist supervenience offers a full answer. But if we expand our interests a little, we’ll see that among the supervenient features of the world, mind must be very exceptional. There are countless such features. In our little toy example of the picture and the pixels, the supervenient properties number 2 to the power: 2 to the millionth power. In the case of materialist supervenience, the number will be far greater. The infinite cardinal beth-3 is a conservative estimate. The vast majority of supervenient features of the world are given only by miscellaneous infinite disjunctions of infinitely complex physical conditions. Therefore they are beyond our power to detect, to name, or to think about one at a time. Mental features of the world, however, are not at all beyond our ken. Finite assemblies of particles – us – can track them. Therefore there must be some sort of simplicity to them. Maybe it will be a subtle sort of simplicity, visible only if you look in just the right way. (Think of the Mandelbrot set: its overwhelming complexity, its short and simple recipe.) But somehow it must be there. Revealing this simplicity is a job for conceptual analysis.

Arbiters of fashion proclaim that analysis is out of date. Yet without it, I see no possible way to establish that any feature of the world does or does not deserve a name drawn from our traditional mental vocabulary. We should repudiate not analysis itself, but only some simplistic goals for it. We should allow for semantic indecision: any interesting analysandum is likely to turn out vague and ambiguous. Often the best that any one analysis can do is to fall safely within the range of indecision. And we should allow for semantic satisficing: analysis may reveal what it would take to deserve a name perfectly,

but imperfect deservers of the name may yet deserve it well enough. (And sometimes the perfect case may be impossible.) If so, there is bound to be semantic indecision about how well is well enough.

I offer not analyses, but a recipe for analyses. We have a very extensive shared understanding of how we work mentally. Think of it as a theory: folk psychology. It is common knowledge among us; but it is tacit, as our grammatical knowledge is. We can tell which particular predictions and explanations conform to its principles, but we cannot expound those principles systematically.<sup>10</sup> Folk psychology is a powerful instrument of prediction. We are capable of all sorts of behaviour that would seem bizarre and unintelligible, and this is exactly the behaviour that folk psychology predicts, rightly, will seldom occur. (But we take a special interest in questions that lie beyond the predictive power of folk psychology; wherefore ingrates may fairly complain of a lack of *interesting* predictions!) Folk psychology has evolved through thousands of years of close observation of one another. It is not the last word in psychology, but we should be confident that so far as it goes – and it does go far – it is largely right.

Folk psychology concerns the causal relations of mental states, perceptual stimuli, and behavioural responses. It says how mental states, singly or in combination, are apt for causing behaviour; and it says how mental states are apt to change under the impact of perceptual stimuli and other mental states. Thus it associates with each mental state a typical causal role. Now we have our recipe for analyses. Suppose we've managed to elicit all the tacitly known general principles of folk psychology. Whenever *M* is a folk-psychological name for a mental state, folk psychology will say that state *M* typically occupies a certain causal role: call this the *M*-role. Then we analyse *M* as meaning 'the state that typically occupies the *M*-role'. Folk psychology implicitly defines the term *M*, and we have only to make that definition explicit.

Since the causal roles of mental states involve other mental states, we might fear circularity. 'The remedy is due in its essentials to Ramsey.'<sup>11</sup> Suppose, for instance, that folk psychology had only three names for mental states: *L*, *M*, *N*. We associate with this triplet of names a complex causal role for a triplet of states, including causal relations within the triplet: call this the *LMN*-role. Folk psychology says that the states *L*, *M*, *N* jointly occupy the *LMN*-role. That implies that *M* occupies the derivative role: coming second in a triplet of states that jointly occupy the *LMN*-role. Taking this as our *M*-role, we proceed as before. Say that the names *L*, *M*, *N* are *interdefined*. The defining of all three via the *LMN*-role is a package deal.

We might fear circularity for another reason. The causal roles of mental states involve responses to perceptual stimuli. But the relevant feature of the stimulus will often be some secondary quality – for instance, a colour. We cannot replace the secondary quality with a specification of the stimulus in purely physical terms, on pain of going beyond what is known to folk psychology. But if we analyse the secondary quality in terms of the distinctive mental states its presence is apt to evoke, we close a definitional circle. So we

should take interdefinition further. Let folk psychology include folk psychophysics. This will say, for instance, that the pair of a certain colour and the corresponding sensation jointly occupy a complex causal role that consists in part, but only in part, of the former being apt to cause the latter. Now we have a derivative role associated with the name of the colour, and another associated with the name of the sensation: the role of coming first or coming second, respectively, in a pair that jointly occupies this complex role.

We might worry also about the behaviour that mental states are apt for causing. Often we describe behaviour in a mentally loaded way: as action. To say that you kicked the ball to your teammate is to describe your behaviour. But such a description presupposes a great deal about how your behaviour was meant to serve your desires according to your beliefs; and also about the presence of the ball and the playing surface and the other player, and about the social facts that unite players into teams. More threat of circularity? More need for interdefinition? I don't know how such further interdefinition would work; and anyway, it would be well to call a halt before folk psychology expands into a folk theory of the entire *Lebenswelt!*

Describing the behaviour in purely physical terms – the angle of the knee, the velocity of the foot – would get rid of those presuppositions. But, just as in the case of the stimuli, it would go beyond what is known to folk psychology. Further, these descriptions would never fit the behaviour of space aliens not of humanoid shape; and yet we should not dismiss out of hand the speculation that folk psychology might apply to aliens as well as to ourselves.

Fortunately there is a third way to describe behaviour. When you kicked the ball, your body moved in such a way that *if* you had been on a flat surface in Earth-normal gravity with a suitably placed ball in front of you and a suitably placed teammate some distance away, *then* the impact of your foot upon the ball would have propelled the ball onto a trajectory bringing it within the teammate's reach. That description is available to the folk. They wouldn't give it spontaneously, but they can recognize it as correct. It presupposes nothing about your mental states, not even that you have any; nothing about whether the ball and the playing field and the gravity and the teammate are really there; nothing about your humanoid shape, except that you have some sort of foot. It could just as well describe the behaviour of a mindless mechanical contraption, in the shape of a space alien (with a foot), thrashing about in free fall.

(I don't say that we should really use these 'if – then' descriptions of behaviour. Rather, my point is that their availability shows how to unload the presuppositions from our ordinary descriptions.)

If *M* means 'the state that typically occupies the *M*-role' and if that role is only imperfectly occupied, what are we to do? – Satisfice: let the name *M* go to a state that deserves it imperfectly. And if nothing comes anywhere near occupying the *M*-role? – Then the name *M* has no referent. The boundary between the cases is vague. To take an example from a different term-introducing theory, I suppose it to be indeterminate whether 'dephlogisticated air' refers to oxygen or to nothing. But folk psychology is in far better shape than

phlogiston theory, despite scare stories to the contrary. We can happily grant that there are no perfect deservers of folk-psychological names, but we shouldn't doubt that there are states that deserve those names well enough.

What to do if the *M*-role, or the *LMN*-role, turns out to be doubly occupied? I used to think that in this case too the name *M* had no referent.<sup>12</sup> But now I think it might be better, sometimes or always, to say that the name turns out to be ambiguous in reference. That follows the lead of Field; and it is consistent with, though not required by, the treatment of Carnap.<sup>13</sup> Note that we face the same choice with phrases like 'the moon of Mars'; and in that case too I'd now lean toward ambiguity of reference rather than lack of it.

My recipe for analyses, like Rylean analytic behaviourism, posits analytic truths that constrain the causal relations of mental states to behaviour. (We have no necessary connections between distinct existences, of course; the necessity is verbal. The state itself could have failed to occupy its causal role, but would thereby have failed to deserve its mental name.) But the constraints are weak enough to be credible. Because the state that typically occupies a role need not occupy it invariably, and also because a state may deserve a name well enough in virtue of a role that it occupies imperfectly, we are safe from the behaviourist's bugbears. We have a place for the resolute deceiver, disposed come what may to behave as if his mental states were other than they really are. We have a place for the total and incurable paralytic with a rich mental life and no behavioural dispositions whatever. We even have a place for a madman whose mental states are causally related to behaviour and stimuli and one another in a totally haywire fashion.<sup>14</sup> And yet not anything goes. At some point – and just where that point comes is a matter of semantic indecision – weird tales of mental states that habitually offend against the principles of folk psychology stop making sense; because at some point the offending states lose all claim to their folk-psychological names. To that extent, analytic behaviourism was right. To quote my closest ally in these matters, '... outward physical behaviour and tendencies to behave do in some way enter into our ordinary concept of mind. Whatever theory of mind is true, it has a debt to pay, and a peace to be made, with behaviourism.'<sup>15</sup>

When we describe mental state *M* as the occupant of the *M*-role, that is what Smart calls a topic-neutral description.<sup>16</sup> It says nothing about what sort of state it is that occupies the role. It might be a non-physical or a physical state, and if it is physical it might be a state of neural activity in the brain, or a pattern of currents and charges on a silicon chip, or the jangling of an enormous assemblage of beer cans. What state occupies the *M*-role and thereby deserves the name *M* is an *a posteriori* matter. But if materialist supervenience is true, and every feature of the world supervenes upon fundamental physics, then the occupant of the role is some physical state or other – because there's nothing else for it to be. We know enough to rule out the chip and the cans, and to support the hypothesis that what occupies the role is some pattern of neural activity. When we know more, we shall know what pattern



of neural activity it is. Then we shall have the premises of an argument for psychophysical identification:<sup>17</sup>

mental state  $M$  = the occupant of the  $M$ -role (by analysis),  
physical state  $P$  = the occupant of the  $M$ -role (by science),  
therefore  $M = P$ .

That's how conceptual analysis can reveal the simple formula – or anyway, the much less than infinitely complicated formula – whereby, when we know enough, we can pick out a mental feature of the world from all the countless other features of the world that likewise supervene on fundamental physics.

The causal-role analyses would still hold even if materialist supervenience failed. They might even still yield psychophysical identifications. Even if we lived in a spook-infested world, it might be physical states that occupied the causal rules (in us, if not in the spooks) and thereby deserved the folk-psychological names. Or it might be non-physical states that occupied the roles. Then, if we knew enough parapsychology, we would have the premises of an argument for psycho-*non*physical identification.

When our argument delivers an identification  $M = P$ , the identity is contingent. How so? – All identity is self-identity, and nothing could possibly have failed to be self-identical. But that is not required. It's contingent, and it can only be known *a posteriori*, which physical (or other) states occupy which causal roles. So if  $M$  means 'the occupant of the  $M$ -role' it's contingent which state is the referent of  $M$ ; it's contingent whether some one state is the common referent of  $M$  and  $P$ ; so it's contingent whether  $M = P$  is true.

Kripke vigorously intuits that some names for mental states, in particular 'pain', are rigid designators: that is, it's not contingent what their referents are.<sup>18</sup> I myself intuit no such thing, so the non-rigidity imputed by causal-role analyses troubles me not at all.

Here is an argument that 'pain' is not a rigid designator. Think of some occasion when you were in severe pain, unmistakable and unignorable. All will agree, except for some philosophers and faith healers, that there is a state that actually occupies the pain role (or near enough); that it is called 'pain'; and that you were in it on that occasion. For now, I assume nothing about the nature of this state, or about how it deserves its name. Now consider an unactualized situation in which it is some different state that occupies the pain role in place of the actual occupant; and in which you were in that different state; and which is otherwise as much like the actual situation as possible. Can you distinguish the actual situation from this unactualized alternative? I say not, or not without laborious investigation. But if 'pain' is a rigid designator, then the alternative situation is one in which you were not in pain, so you could distinguish the two very easily. So 'pain' is not a rigid designator.

Philosophical arguments are never incontrovertible – well hardly ever. Their purpose is to help expound a position, not to coerce agreement. In this case,

the controvertor might say that if the actual occupant of the pain role is not a physical state, but rather is a special sort of non-physical state, then indeed you can distinguish the two situations. He might join me in saying that this would not be so if the actual occupant of the role were a physical state – else neurophysiology would be easier than it is – and take this together with intuitions of rigidity to yield a *reductio* against materialism. Myself, I don't see how the physical or non-physical nature of the actual occupant of the role has anything to do with whether the two situations can be distinguished. Talk of 'phenomenal character' and the like doesn't help. Either it is loaded with question-begging philosophical doctrine, or else it just reiterates the undisputed fact that pain is a kind of experience.<sup>19</sup>

If there is variation across worlds with respect to which states occupy the folk-psychological roles and deserve the folk-psychological names (and if this variation doesn't always require differences in the laws of nature, as presumably it doesn't) then also there can be variations within a single world. For possibility obeys a principle of recombination: roughly, any possible kind of thing can coexist with any other.<sup>20</sup> For all we know, there may be variation even within this world. Maybe there are space aliens, and maybe there will soon be artificial intelligences, in whom the folk-psychological roles are occupied (or near enough) by states very different from any states of a human nervous system. Presumably, at least some folk-psychological roles are occupied in at least some animals, and maybe there is variation across species. There might even be variation within humanity. It depends on the extent to which we are hard-wired, and on the extent of genetic variation in our wiring.

We should beware, however, of finding spurious variation by overlooking common descriptions. Imagine two mechanical calculators that are just alike in design. When they add columns of numbers, the amount carried goes into a register, and the register used for this purpose is selected by throwing a switch. Don't say that the carry-seventeen role is occupied in one machine by a state of register *A* and in the other by a state of register *B*. Say instead that in both machines alike the role is occupied by a state of the register selected by the switch. (Equivalently, by a state of a part of the calculator large enough to include the switch and both registers.) If there is a kind of thinking that some of us do in the left side of the brain and others do in the right side, that might be a parallel case.

If *M* means 'the occupant of the *M*-role' and there is variation in what occupies the *M*-role, then our psychophysical identities need to be restricted: not plain  $M = P$ , but  $M\text{-in-}K = P$  where *K* is a kind within which *P* occupies the *M*-role. Human pain might be one thing, Martian pain might be something else.<sup>21</sup> As with contingency, which is variation across worlds, so likewise with variation in a single world: the variability in no way infects the identity relation, but rather concerns the reference of the mental name.

The threat of variation has led many to retreat from 'type-type' to 'token-token' identity. They will not say that  $M = P$ , where *M* and *P* are names

for a state that can be common to different things at different times – that is, for a property had by things at times. But they will say that  $m = p$ , where  $m$  and  $p$  are mental and physical names for a particular, unrepeatable event. Token–token identities are all very well, in their derivative way, but the flight from type–type identities was quite unnecessary. For our restricted identities, of the form  $M\text{-in-}K = P$ , are still type–type.

But don't we at least have a choice? Couldn't our causal role analyses be recast in terms of the causal roles of tokens, and if they were, would they not then yield token–token identities? After all, the only way for a type to occupy a causal role is through the causes and effects of its tokens. The effects of pain are the effects of pain-events – I think, following Jackson, Pargetter, and Prior, that this recasting of the analyses would not be easy.<sup>22</sup> There are more causal relations than one. Besides causing, there is preventing. It too may figure in folk-psychological causal roles; for instance, pain tends to prevent undivided attention to anything else. Prevention cannot straightforwardly be treated as a causal relation of tokens, because the prevented tokens do not exist – not in this world, anyway. It is better taken as a relation of types.

If a retreat had been needed, a better retreat would have been to 'subtype–subtype' identity. Let  $MK$  name the conjunctive property of being in state  $M$  and being of kind  $K$ ; and likewise for  $PK$ . Do we really want psychophysical identities of the form  $MK = PK$ ? – Close, but I think not quite right. For one thing,  $M\text{-in-}K$  is not the same thing as  $MK$ . The former but not the latter can occur also in something that isn't of kind  $K$ . For another thing, it is  $P$  itself, not  $PK$ , that occupies the  $M$ -role in things of kind  $K$ .

Non-rigidity means that  $M$  is different states in different possible cases; variation would mean that  $M$  was different states in different actual cases. But don't we think that there is *one* property of being in the state  $M$  – one property that is common to all, actual or possible, of whatever kind, who can truly be said to be in state  $M$ ? – There is. It is the property such that, for any possible  $X$ ,  $X$  has it just in case  $X$  is in the state that occupies the  $M$ -role for  $X$ 's kind at  $X$ 's world.<sup>23</sup> The gerund 'being in  $M$ ' can be taken, at least on one good disambiguation, as a rigid designator of this property. However, this property is not the occupant of the  $M$ -role. It cannot occupy that or any other causal role because it is excessively disjunctive, and therefore no events are essentially havings of it.<sup>24</sup> To admit it as causally efficacious would lead to absurd double-counting of causes. It would be like saying that the meat fried in Footscray cooked because it had the property of being either fried in Footscray or boiled in Bundoora – only worse, because the disjunction would be much longer and more miscellaneous.

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.

It is unfortunate that this superficial question has sometimes been taken to mark the boundary of 'functionalism'. Sometimes so and sometimes not – and that's why I have no idea whether I am a functionalist.

Those who take 'pain' to be a rigid designator of the highly disjunctive property will need to controvert my argument that 'pain' is not rigid, and they will not wish to claim that one can distinguish situations in which the pain-role is differently occupied. Instead, they should controvert the first step, and deny that the actual occupant of the pain-role is called 'pain'. I call that denial a *reductio*.

### Notes

- 1 See David Lewis, 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61 (1983), pp. 343–377; and David Lewis, *On the Plurality of Worlds* (Blackwell, 1986), pp. 59–69.
- 2 See David Lewis, 'Critical Notice of D. M. Armstrong, *A Combinational Theory of Possibility*', *Australasian Journal of Philosophy* 70 (1992), pp. 211–224.
- 3 Lewis, 'New Work for a Theory of Universals', pp. 365–370.
- 4 Gerald Feinberg, 'Physics and the Thales Problem', *Journal of Philosophy* 66 (1966), pp. 5–13.
- 5 Lewis, 'New Work for a Theory of Universals', pp. 361–365.
- 6 Frank Jackson, 'Epiphenomenal Qualia', *Philosophical Quarterly* 32 (1982), pp. 127–136.
- 7 Laurence Nemirow, 'Physicalism and the Cognitive Role of Acquaintance' and David Lewis, 'What Experience Teaches', both in *Mind and Cognition: A Reader*, ed. by W. G. Lycan (Blackwell, 1990).
- 8 Lewis, *On the Plurality of Worlds*, pp. 212–213.
- 9 Frank Jackson, 'Armchair Metaphysics', in *Philosophy in Mind*, ed. by J. O'Leary Hawthorne and M. Michael (Kluwer, 1994); Robert Stalnaker, 'Assertion', *Syntax and Semantics* 9 (1978), pp. 315–332; M. K. Davies and I. L. Humberstone, 'Two Notions of Necessity', *Philosophical Studies* 38 (1980), pp. 1–30; and Pavel Tichý, 'Kripke on Necessity *A Posteriori*', *Philosophical Studies* 43 (1983), pp. 225–241.
- 10 See David Lewis, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (1972), pp. 249–258, eliciting the general principles of folk psychology is no mere matter of gathering platitudes.
- 11 F. P. Ramsey, 'Theories' in Ramsey, *The Foundations of Mathematics*. (Routledge & Kegan Paul, 1931), pp. 212–236; Rudolf Carnap, 'Replies and Expositions' in *The Philosophy of Rudolf Carnap*, ed. by P. A. Schilpp (Cambridge University Press, 1963), pp. 958–966. See also David Lewis, 'How to Define Theoretical Terms', *Journal of Philosophy* 67 (1970), pp. 427–446, reprinted in Lewis, *Philosophical Papers*, Vol. 1 (Oxford University Press, 1983); and Lewis, 'Psychophysical and Theoretical Identifications'.
- 12 David Lewis, 'How to Define Theoretical Terms' and 'Psychophysical and Theoretical Identifications'.

- 13 Hartry Field, 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70 (1973), pp. 462–481; Carnap *loc. cit.*
- 14 David Lewis, 'Mad Pain and Martian Pain', in *Readings in Philosophy of Psychology*, Vol. 1, ed. by N. Block (Harvard University Press, 1980), reprinted with postscript in Lewis, *Philosophical Papers*, Vol. 1 (Oxford University Press, 1983).
- 15 D. M. Armstrong, *A Materialist Theory of Mind* (Routledge & Kegan Paul, 1968), p. 68.
- 16 J. J. C. Smart, 'Sensations and Brain Processes', *Philosophical Review*, 68 (1959), pp. 141–156.
- 17 See David Lewis, 'An Argument for the Identity Theory', *Journal of Philosophy* 63 (1966), pp. 17–25, reprinted with additions in Lewis, *Philosophical Papers*, Vol. 1 (Oxford University Press, 1983); and Lewis, 'Psychophysical and Theoretical Identifications'. See Armstrong, *A Materialist Theory of the Mind*, for an independent and simultaneous presentation of the same position, with a much fuller discussion of what the definitive causal roles might be.
- 18 Saul Kripke, *Naming and Necessity* (Blackwell, 1980), pp. 147–148.
- 19 The controvertor just imagined would agree with the discussion in Kripke, *Naming and Necessity*, pp. 144–155. But I don't mean to suggest that Kripke would agree with him. At any rate, the words I have put in his mouth are not Kripke's.
- 20 Lewis, *On the Plurality of Worlds*, pp. 86–92.
- 21 Lewis, 'Mad Pain and Martian Pain'.
- 22 Frank Jackson, Robert Pargetter, and Elizabeth Prior, 'Functionalism and Type-type Identity Theories', *Philosophical Studies* 42 (1982), pp. 209–225.
- 23 In 'How to Define Theoretical Terms' I called it the 'diagonalized sense' of *M*.
- 24 David Lewis, 'Events' in Lewis, *Philosophical Papers*, Vol. 2 (Oxford University Press, 1986).