Blackwell Companions to Philosophy

A Companion to the Philosophy of Mind

Edited by

SAMUEL GUTTENPLAN

BLACKWELL
Reference

First published 1994

Blackwell Publishers, the publishing imprint of
Basil Blackwell Ltd
108 Cowley Road, Oxford OX4 1JF, UK

Basil Blackwell Inc.
238 Main Street
Cambridge, Massachusetts 02142, USA

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A Companion to the philosophy of mind / edited by Samuel Guttenplan.
p. cm. – (Blackwell companions to philosophy)
Includes bibliographical references and index.
ISBN 0-631-17953-4 (hard : alk. paper)
1. Philosophy of mind. I. Guttenplan, Samuel D. II. Series.
BD418.3.C62 1993
128'.2-dc20 93-39595
CIP

Typeset in 10½ on 12½pt Photina by Acorn Bookwork, Salisbury, Wilts
Printed in Great Britain by Hartnolls, Bodmin

This book is printed on acid-free paper

Lewis, David: 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'. (*See* FUNCTIONALISM; IDENTITY THEORIES; PHYSICALISM; RYLE.)

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 (*see* Lewis, 1983a and 1986a, pp. 59–69). 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* (Lewis, 1992, p. 218). (*See* SUPERVENIENCE.)

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; *see* Lewis, 1983a, pp. 365–70). We have no a priori guarantee of it, but we may reasonably think that present-day physics 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 (Feinberg, 1966). 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 (Lewis, 1983a, pp. 361–5).

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 (1982). 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 happened. (*See* CONSCIOUSNESS; QUALIA.)

Materialists have said many things about what does happen in such a case. I myself, following Nemirow (1990), call it a case of know-how: Mary gains new imaginative abilities (Lewis, 1990). 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 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.

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 (Lewis, 1986a, pp. 212–13), 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 (1992), is founded upon the 'two-dimensional' analysis of necessity a posteriori put forward by Stalnaker (1978), Davies and Humberstone (1980), and Tichý (1983). Two-dimensionalism says that there is no such thing as a necessary *a posteriori* proposition. However, one single sentence θ 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 θ 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 θ at the various worlds: and let H be the proposition horizontally expressed by Θ at the actual world. The proposition *diagonally expressed* by θ is the proposition D that holds at any world W iff the proposition horizontally expressed by θ at W is true at W. So if we know D, we know that θ horizontally expresses some truth or other, but we may not know which truth. Sentence θ is necessary *a posteriori* iff H is necessary but D is knowable only a posteriori. Likewise, a proposition P *necessarily* implies that θ iff P implies H; but P a priori implies that θ iff P implies D. Our worry was that when θ was about the mind, and P was a premise made true by fundamental physics, P might imply that θ 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 θ 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 θ. 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 miscellaneously 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. (*Pace* Lewis, 1972, p. 256, eliciting the general principles of folk psychology is no mere matter of gathering platitudes.) 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 the 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 (1931a, pp. 212–236) and Carnap (1963, pp. 958–66); see also Lewis (1970, 1972). 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 team-mate 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 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 team-mate some distance away, *then* the impact of your foot upon the ball would have propelled the ball onto a trajectory bringing it within the team-mate'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 team-mate 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-inducing 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 the 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 (Lewis, 1970, 1972) that in this case too the name M had no referent. 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 (1973): and it is consistent with, though not required by, the treatment of Carnap (1963). 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 (Lewis, 1980). 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' (Armstrong, 1968, p. 68).

When we describe mental state M as the occupant of the M-role, that is what Smart (1959) calls a topic-neutral description. 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:

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.

(*See* Lewis, 1966, 1972; *and see* Armstrong, 1968, for an independent and simultaneous presentation of the same position, with a much fuller discussion of what the definitive causal roles might be.)

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 roles (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 (1972) vigorously intuits that some names for mental states, in particular 'pain', are rigid designators: that is, it's not contingent what their referents are. 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 controverter 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.

(The controverter just imagined would agree with the discussion in Kripke, 1972, pp. 344–42. But I don't mean to suggest that Kripke would agree with him. At any rate, the words I have put into his mouth are not Kripke's.)

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 (Lewis, 1986a, pp. 86–92). 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-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 (Lewis, 1980). 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-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 (1982), that this recasting of the analyses would not be easy. 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-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. (In Lewis, 1970, I called it the 'diagonalized sense' of M.) 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 (Lewis, 1986c). 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 realization 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 fuctionalist.

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*.

CONTENT

A mind is an organ of REPRESENTATION. Many things are true according to it; that is, they are believed. Or better, they are more or less probable according to it; that is, they are believed or disbelieved to varying degrees. Likewise, many things are desired to varying positive or negative degrees. What is believed, or what is desired, we call the CONTENT of BELIEF or DESIRE.

(I think it an open question to what extent other states with content – doubting, wondering, fearing, pretending, . . . – require separate treatment, and to what extent they can be reduced to patterns in belief and desire and contentless feeling. Be that as it may, I shall ignore them here (*see* EMOTION).)

What determines the content of belief and desire? – The occupation of folk-psychological roles by physical states, presumably neural states; and ultimately the pattern of coinstantiation of fundamental physical properties and relations. But to say just that is to say not much. Those who agree with it can, and do, approach the problem of content in very different ways.

I can best present the approach I favour by opposing it to an alternative. A crude sketch will suffice, so in fairness I name my opponent *Strawman*. I doubt there is anyone real who takes exactly the position that Strawman does – but very many are to be found in his near vicinity.

Strawman says that folk psychology says – and truly – that there is a LANGUAGE OF THOUGHT. It has words, and it has syntactic constructions whereby those words can be combined into sentences. Some of these sentences have a special status. Strawman says they are 'written in the belief box' or 'in the desire box', but even Strawman doesn't take that altogether literally. There are folk-psychological causal roles for the words, for the syntactic constructions, and for the belief and desire boxes. It is by occupying these roles that the occupants deserve their folk-psychological names.

The question what determines content then becomes the question: what determines the semantics of the language of thought? Strawman says that folk psychology specifies the semantic operations that correspond to syntactic constructions such as predication. As for the words, Strawman says that folk psychology includes, in its usual tacit and unsystematic way, a causal theory of reference (more or less as in Kripke, 1972). There are many relations of acquaintance that connect the mind to things, including properties and relations, in the external world. Some are relations of perceptual acquaintance. Others are less direct: you are acquainted with the thing by being acquainted with its traces. Often, you are acquainted with the thing by way of its *linguistic* traces that is, you have heard of it by name. Somehow, in virtue of the different causal roles of different words of the language of thought, different words are associated with different relations of acquaintance, which connect them to different external things. Whatever a word is thus connected to is the referent of that word.

Once the words of the language of thought have their referents, the sentences have their meanings. These are structures built up from the referents of the words in a way that mirrors the syntactic construction of the sentences from the words. Take predication – Strawman's favourite example. A word F of your language of thought is connected by one relation of acquaintance to the property of being French. Another word A is connected by another relation of acquaintance to the man André. (The first

relation might be linguistic, for instance, and the second perceptual.) The syntactic construction of predication builds a sentence F(A). Its meaning is the ordered pair of the property of being French and André. Such a pair is a 'singular proposition', true just in case its second element instantiates its first. (Other singular propositions are triples, quadruples, . . ., with relations in the first place.) If you have F(A) written in your belief box, you thereby believe that André is French.

Strawman's account of content is sketchy, as I said it would be. Even with help from all his allies, I doubt he will find it easy to fill the gaps. I especially wonder what he can say about how the words get hooked up to the right relations of acquaintance. A causal theory of reference for public language might usefully mention mutual expectations among language-users, intentions to instill beliefs, semantic intentions, or other such instances of mental content. But even if we had corresponding expectations or intentions about our own language of thought, Strawman could not without circularity use them in a general account of mental content.

Suppose, all the same, that Strawman's account could be completed successfully by its own lights. I would still have four objections.

First, I don't believe that folk psychology says there is a language of thought. Rather, I think it is agnostic about how MENTAL REPRESENTATION works – and wisely so.

What is the issue? Of course everybody should agree that the medium of mental representation is somehow analogous to language. A raven is like a writing-desk. Anything can be analogized to anything. And of course nobody thinks the head is full of tiny writing.

A serious issue, and one on which I take folk psychology to be agnostic, concerns the relation between the whole and the parts of a representation. Suppose I have a piece of paper according to which, *inter alia*, Collingwood is east of Fitzroy. Can I tear the paper up so that I get one snippet that has exactly the content that Collingwood is east of Fitzroy, nothing more and nothing less? If the paper is covered with writing, maybe I can; for maybe 'Collingwood is east of Fitzroy' is one of the sentences written there. But if the paper is a map, any snippet according to which Collingwood is east of Fitzroy will be a snippet according to which more is true besides. For instance, I see no way to lose the information that they are adjacent, and that a street runs along the border. And I see no way to lose all information about their size and shape.

(A hologram, or famously a connectionist network (*see* CONNECTIONISM), differs even more from a paper covered with writing. If we make a hologram of the map and break it into snippets, detail will be lost in blur. But the arrangement of *all* the suburbs, provided it was shown with sufficient prominence on the original map, will remain to the last.)

Mental representation is language-like to the extent that parts of the content are the content of parts of the representation. If our beliefs are 'a map . . . by which we steer', as Ramsey said (1931b, p. 238), then they are to that extent not language-like. And to that extent, also, it is misleading to speak in the plural of beliefs. What is one belief? No snippet of a map is big enough that, determinately, something is true according to it, and also small enough that, determinately, nothing is true according to any smaller part of it. If mental representation is map-like (let alone if it is hologram-like) then 'beliefs' is a bogus plural. You have beliefs the way you have the blues, or the mumps, or the shivers.

But if mental representation is language-like, one belief is one sentence written in the belief box, so 'beliefs' is a genuine plural. Whether the plural is bogus or genuine is not settled by rules of grammar. Rather, it is an empirical question, and a question that folk psychology leaves open. 'The shivers' might be a parallel case. Is there such a thing as one shiver? – Maybe and maybe not. I don't think one cycle of vibration should be called 'one shiver', but there might be a better candidate. What if one

firing of a control neuron would set you shivering for four seconds, and prolonged shivering is caused by this neuron firing every two seconds? If so, I think the shivering set off by one firing could well be called 'one shiver', and then it is right to say that shivering consists of a sequence of overlapping shivers. Under this hypothesis, the plural is genuine. Under other hypotheses, the plural is bogus.

Of course you might say, under the hypothesis that mental representation is map-like, that any proposition true according to the mental map is one belief. Or you might say that the one belief that Collingwood is east of Fitzroy is the highly disjunctive state of having some mental map or other according to which that *inter alia* is true. Say so if you like. But I only insist that if you say either thing, then you may not also assume that 'one belief' is the sort of thing that can occupy a causal role. You may still say '. . . because he believes that Collingwood is east of Fitzroy', but only if you mean by it '. . . because he has beliefs' – bogus plural! – 'according to which *inter alia* Collingwood is east of Fitzroy'.

If Strawman heeds the advice of some of his allies, he will respond by changing his position. He will give away conceptual analysis and folk psychology, and market his wares as 'cognitive science' (see COGNITIVE PSYCHOLOGY). No problem, then, if the folk are agnostic about the language of thought. Let it be a new hypothesis, advanced because it best explains . . . What? Well-known facts about belief? – But 'belief' is a folk-psychological name for a kind of state posited by folk psychology. If Strawman leaves all that behind him, where shall he find his evidence? He can never again set up thought experiments and ask us what we want to say about them. That would only elicit our folk-psychological preconceptions. He can make a fresh start if he really wants to – I assume he will not want to – but he cannot have his cake and eat it too. (*See* Jackson, 1992.)

If Strawman stands his ground, on the other hand, he will insist that folk psychology is far from agnostic about the language of thought. It has plenty to say, after all, about our 'concepts' (or 'ideas') of things. Our concept of a concept, says Strawman, is just our concept of a word of the language of thought. – I doubt it. I haven't much of any concept of Elsternwick. I have little idea what the place looks like, what sort of people live there, . . . All I know is that there is a place of that name, and roughly where it is. But I *do* have the word. (At least, I have the word 'Elsternwick' of our public language. *If* I have a language of thought, presumably this word has been borrowed into it.) My lack of a concept isn't lack of a word; rather, I lack any very rich cluster of associated descriptions.

Strawman can reply that even if I haven't *much* concept of Elsternwick, still I have enough of one that I can think about Elsternwick (for instance, when I think how little I know about it). It is this minimal concept of a concept, he says, that is our folk-psychological concept of a word of the language of thought. – Yes, I have a concept of Elsternwick in the minimal sense that I have *whatever it takes* to be able to think about it. But must the basis of such an ability, in general or even in this case, be the possession of a word? On that question, the folk and I remain agnostic.

My second objection to Strawman's account is that it delivers only wide content. Which singular propositions you believe depends upon which external things are suitably connected by relations of acquaintance to the words of your language of thought.

Strawman holds that all content is wide because he has learned the lesson of TWIN EARTH. Recall the example (Putnam, 1975, pp. 139–42). Oscar the Earthling believes that water often falls from clouds. Twoscar on Twin Earth is in no way acquainted with water, that is, with H_2O. Rather, Twoscar is acquainted with XYZ, a superficially similar liquid that is abundant on Twin Earth, in exactly the way that Oscar is acquainted with H_2O. There is no other relevant difference between Twoscar and Oscar. We are invited to agree that Twoscar does not believe that water falls from clouds, and

believes instead that XYZ falls from clouds. Strawman does agree.

And so do I, but with many reservations. For one thing, I think agreement is not compulsory. Like any up-to-date philosopher of 1955, I think that 'water' is a cluster concept. Among the conditions in the cluster are: it is liquid, it is colourless, it is odourless, it supports life. But, *pace* the philosopher of 1955, there is more to the cluster than that. Another condition in the cluster is: it is a natural kind. Another condition is indexical: it is abundant hereabouts. Another is metalinguistic: many call it 'water'. Another is both metalinguistic and indexical: *I* have heard of it under the name 'water'. When we hear that XYZ off on Twin Earth fits many of the conditions in the cluster but not all, we are in a state of semantic indecision about whether it deserves the name 'water'. (*See* Unger, 1984, pp. 79–104. But while I agree with Unger about what happens in various cases, I don't endorse all the morals he draws.) When in a state of semantic indecision, we are often glad to go either way, and accommodate our own usage temporarily to the whims of our conversational partners (Lewis, 1979b). So if some philosopher, call him Schmutnam, invites us to join him in saying that the water on Twin Earth differs in chemical composition from the water here, we will happily follow his lead. And if another philosopher, Putnam (1975), invites us to say that the stuff on Twin Earth is not water – and hence that Twoscar does not believe that water falls from clouds – we will just as happily follow his lead. We should have followed Putnam's lead only for the duration of that conversation, then lapsed back into our accommodating state of indecision. But, sad to say, we thought that instead of playing along with a whim, we were settling a question once and for all. And so we came away lastingly misled.

The example half suceeds. It is not compulsory, but certainly it is permissible, to say that Oscar does believe that water falls from clouds and differently acquainted Twoscar does not. Therefore wide content does serve a purpose. It enters into the analysis of some sentences that are about belief, or at least partly about belief; or at least it does so under some permissible disambiguations of these sentences.

Other examples are similar. Twoscar is acquainted with molybdenum as Oscar is with aluminium; with a disease of bone as Oscar is with a disease of joints; with spy robots as Oscar is with cats; and so on. It seems to matter little whether Twoscar is our neighbour, or whether he lives on a remote planet, or whether he lives in a different possible world. In each case we find that the difference in what Twoscar and Oscar are acquainted with makes a difference to the truth value, under some disambiguation, of some sentences that are at least partly about belief. But that is all we find. There is nothing here to support Strawman's thesis that wide content is the only kind of content; or that it is in any way pre-eminent or basic.

We should not jump to the conclusion that just any belief sentence is susceptible to Twin Earth examples. Oscar thinks that square pegs don't fit round holes; I don't think you can tell an even halfway convincing story of how Twoscar, just by being differently acquainted, fails to think so too. Oscar believes there's a famous seaside place called 'Blackpool'; so does differently acquainted Twoscar, though of course it may not be Blackpool – not *our* Blackpool – that he has in mind. Oscar believes that the stuff he has heard of under the name 'water' falls from clouds. So does Twoscar – and so does Twoscar even if you alter not only his acquaintance with water but his relations of acquaintance to other things as well. You know the recipe for Twin Earth examples. You can follow it in these cases too. But what you get falls flat even as an example of how content is sometimes wide, let alone as evidence that content is always wide.

The famous brain in a bottle is your exact duplicate with respect to brain states and their typical causal roles; but is acquainted only with aspects of the computer that fabricates its virtual reality. You and the brain share no objects of acquaintance. So,

according to Strawman, you and the brain share no common beliefs whatever.

Newborn Swampman, just this moment formed by an unlikely chance assembly of atoms, also is your exact duplicate with respect *inter alia* to brain states and their typical causal roles (Davidson, 1987). But so far, he hasn't had time to become acquainted with much of anything. Therefore, according to Strawman, he believes not much of anything.

Strawman and his allies may think that we have here two remarkable philosophical discoveries. I think, rather, that Strawman's thesis that all content is wide has here met with a twofold *reductio ad absurdum*. Granted, the brain in a bottle shares no wide content of belief with you. Granted, Swampman has no wide content of belief at all. Yet there must be some good sense in which both the brain and Swampman are your mental twins; some good sense in which they believe just what you do. (And in our less extreme cases, there must be some good sense in which Twoscar believes just what Oscar does.) Strawman's position is unacceptable. Not because it posits wide content; but because it omits narrow content, content independent of what one is acquainted with. It omits the sort of content that you and the brain and Swampman, and likewise Oscar and Twoscar, have in common.

(Narrow content is independent of what you are acquainted with, but that does not mean that it is altogether intrinsic to you. For it still depends on the causal roles of your brain states; causation depends on the laws of nature; and if some sort of regularity theory of lawhood is true, living under such-and-such laws is not intrinsic to you. Further, it is the typical causal roles of your brain states that matter. But you may be an atypical member of your kind; hence what is typical of your kind is not intrinsic to you. So I can say only this: if X and Y are intrinsic duplicates, and if they live under the same laws of nature, and if they are the same in kind, then they must be exactly alike in narrow content.)

In insisting on the existence of narrow content, I am not guided by any preconception about what sort of properties may figure in causal explanation, or in truly scientific explanation. I dare say the fundamental laws of physics must concern perfectly natural, intrinsic properties. But that's irrelevant, since causal and scientific explanation seldom consists in subsumption under these fundamental laws. Rather, it is a matter of giving information about how things are caused (Lewis, 1986b). Such information can come in many forms, both within science and without, and there is no reason to proscribe extrinsic classifications. (Lynne Baker told me a nice example: the science of economics is all about extrinsic properties like poverty and debt. Yet there is nothing wrong, and nothing unscientific, in saying that Fred stays poor because of his burden of debt.)

I am guided, rather, by my tacit mastery of the principles of folk psychology. I said: Oscar believes that the stuff he has heard of under the name 'water' falls from clouds; and so does Twoscar. (And so do you, and so does the brain in a bottle, and so does Swampman.) These are ordinary folk-psychological belief sentences; but narrow ones, as witness the fact that they are not susceptible to Twin Earth examples.

This narrow content is content, rightly so-called: something is true according to the belief-system in question. The content is true on condition that the stuff the believer has heard of under the name 'water' does indeed fall from clouds; otherwise false. It is not 'purely syntactic content' – something I take to be a contradiction in terms. Nor is it a mere function that delivers genuine content as output when given circumstances of acquaintance as input. Nor is it merely phenomenalistic content, restricted in subject matter to the believer's experience.

However, it is not content that can be given by a singular proposition, and that leads to my third objection against Strawman's account.

Strawman's singular propositions suffice to specify which things have which properties. If all else supervenes upon the pattern

of coinstantiation of fundamental properties, that in turn will suffice to specify the way the world is. But much of the content of our knowledge and belief is *de se*: it concerns not the world but oneself. (*See* Perry, 1977; Lewis, 1979a; Chisholm, 1979.) However much I may know about the things that make up the world, their properties and their arrangement, it is something extra to know which one of all these things is *me*. This is *de se* knowledge, whereby I locate myself in the world and self-ascribe the properties I think myself to possess, but is not knowledge of how the world is. Its content cannot be captured by singular propositions. What singular proposition is expressed when I say, or I think, 'I am DL'? – Just the proposition that DL = DL. And when I self-ascribe the property F? – Just the proposition that DL is F. But I can know these propositions without knowing who I am, or whether I am F. (And you can know them too.) Strawman's only recourse is to say that *de se* knowledge is characterized not by its *de se* content but some other way – and if he says that, he confesses that his account of content is inadequate. Belief that falls short of knowledge can likewise have *de se* content. If you take *your*self to be DL, your false belief and my true belief have their *de se* content in common. Desire also has *de se* content. If you desire to be F and I believe myself to be F, again the two attitudes have their *de se* content in common.

There is also tensed content. The world is spread out over many times; but we can have knowledge, or belief or desire, about which of these times is now. Again, this is not knowledge of how the whole spread-out world is. It is something extra. Some would speak of content *de se et nunc*, but I would subsume *de nunc* under *de se*. For I think we persist through time by consisting of many time-slices, or momentary selves; and in the last analysis, it is these momentary selves that do our thinking. So when I think 'It's now time for lunch', that's one of my momentary selves self-ascribing *de se* the property of being located at lunchtime.

The 'propositions', if we may call them that, which make up *de se* content are true or false not absolutely, as singular propositions are, but relative to a subject. (Or to a subject at a time, if you don't believe in momentary selves.) The content of my knowledge *de se* that I am DL is something that is true for me but not for you. Its linguistic expression requires a first-person pronoun, or some equivalent device. We could call it an 'egocentric proposition' (or 'egocentric and tensed'). Or we can simply identify it with the property that I self-ascribe: the property of being DL. Likewise the *de se* content of my belief that I have F is just the property of F itself; the *de se* content of my belief that it's lunchtime is the property (possessed not by the whole of me but by some of my momentary selves) of being located at lunchtime; and so on. A *de se* self-ascription of a property is true just on condition that the self-ascriber possesses the self-ascribed property.

(May I say, then, that *de se* belief has 'truth conditions'? Not if Strawman has his way. He goes in for terminological piracy. He transforms one term after another into a mere synonym for 'singular proposition'. He has taken 'object of thought'. He has taken 'content'. He has taken 'proposition'. He is well on the way to taking 'truth condition'. When he has taken all the terms for his own, dissident thoughts will be unsayable.)

Since Strawman has no place for *de se* content, it makes sense that he overlooks narrow content as well. For narrow content is very often *de se*. To revisit our previous example: Oscar self-ascribes having heard, under the name of 'water', of a liquid that falls from clouds. He also self-ascribes the property of being at a place (and time) in the vicinity of which the most abundant liquid is one that falls from clouds. Differently acquainted Twoscar self-ascribes these same two properties, and in this way Oscar and Twoscar share the same *de se* narrow content of belief.

On my own view, it is just such *de se* narrow content that underlies wide content. The semantics of the alleged language of thought needn't enter into it. To the extent that language enters my story at all, it is not by way of the language of thought, but

rather by way of thought about language – about the ordinary public language, whereby, for instance, Oscar heard of something under the name 'water'.

Here is one recipe (Lewis, 1979a, pp. 538–43): if R is a relation of acquaintance, and subject S self-ascribes being R-acquainted uniquely with something that has property F (the narrow part), and if S is R-acquainted uniquely with A (the wide part), S thereby widely believes the singular proposition that A has F. There are variants on the recipe. Our example of French André was a case in which property F as well as individual A enters indirectly as an object of acquaintance; we must of course let in cases where the property F gives way to a relation with two or more *relata*; maybe sometimes we should drop the qualification 'uniquely'; and maybe sometimes the relation R is not, or not entirely, a matter of acquaintance. But in every case, wide belief in a singular proposition derives from narrow *de se* self-ascription plus facts about what the subject is related to.

Often we know a lot about which singular propositions someone believes in this wide and derivative way; but we know less about *how* – in virtue of just which self-ascriptions and relations of acquaintance – he believes those singular propositions. So it's no surprise to find that our ordinary-language belief sentences often seem to be ascriptions of wide content. Often; but not always. In these last few paragraphs I've been talking about *de se* narrow content, and I've been talking about it in plain English. (Such bits of jargon as I used were first explained in plain English.)

There are still other dimensions to the semantic complexity and the multifarious ambiguity of ordinary-language belief sentences. Think of the belief sentences that show up as test cases in articles advocating one semantic analysis or another. I *always* want to say: 'in a sense that's true, in a sense false'. One complication is that we get direct-quotational effects even in what is ostensibly indirect quotation (*see* Rieber, 1992). An example: Fred knows perfectly well that the house he lives in is made of wood, but Fred also thinks that 'abode' is the English word for a house made of mud-brick. 'Fred believes that he has an abode – yes or no?' In at least some contexts (this isn't one of them) I'd be prepared to insist on 'no'. Wouldn't you? Moral: if you hope to understand the folk psychology of belief by studying the linguistic phenomenology of ordinary belief sentences, you're in for big trouble.

I've said that narrow content is very often *de se*, but by resorting to a cheap trick I can change 'often' to 'always'. Take an apparent exception: the narrow belief that square pegs won't fit in round holes. Take this to be the *de se* self-ascription of the property of inhabiting a world wherein square pegs won't fit in round holes. A peculiar property, since either all the inhabitants of the world share it or else none do; and, like many other self-ascribed properties, very far from fundamental: but in a broad enough sense of the word, a property all the same. Likewise you can self-ascribe the property of inhabiting a world where there's a famous seaside place called 'Blackpool'. And so on, until all narrow content has been included as *de se*. Hoky, but maybe worthwhile for the sake of uniform treatment.

My final objection is that Strawman ignores large parts of the folk psychology of belief and desire: the parts that characterize aspects of our RATIONALITY. Folk psychology says that a system of beliefs and desires tends to cause behaviour that serves the subject's desires according to his beliefs. Folk psychology says that beliefs change constantly under the impact of perceptual evidence: we keep picking up new beliefs, mostly true, about our perceptual surroundings; whereupon our other beliefs (and our instrumental desires) change to cohere with these new beliefs. Folk psychology sets presumptive limits to what basic desires we can have or lack: *de gustibus non disputandum*, but still a bedrock craving for a saucer of mud would be unintelligible (Anscombe, 1958, pp. 69–71). Likewise it sets limits to our sense of plausibility: which hypotheses we find credible prior to evidence, hence

which hypotheses are easily confirmed when their predictions come true. And it sets presumptive limits on what our contents of belief and desire can be. Self-ascribed properties may be 'far from fundamental', I said – but not *too* far. Especially gruesome gerrymanders are *prima facie* ineligible to be contents of belief and desire. (See Lewis, 1983a, pp. 370–7; Lewis, 1986a, pp. 38–9 and 105–8.) In short, folk psychology says that we make sense. It credits us with a modicum of rationality in our acting, believing, and desiring.

(Beware. 'Rationality' is an elastic word, and here I've stretched it to cover a lot. If you'd rather use it more narrowly – just for the serving of desires according to beliefs, say – no harm done. So long as you don't just ignore the several other departments of rationality that I listed, it doesn't matter what you call them.)

If mental states are to be analysed as occupants of folk-psychological roles, and if the folk psychology of belief and desire has a lot to say about rationality, and if what it says is framed in terms of content, then it seems that constraints of rationality are constitutive of content. Yet Strawman's account of content makes no place for constitutive rationality. Why not?

Perhaps Strawman thought, wisely, that it would be better to say too little than too much. It wouldn't do to conclude that, as a matter of analytic necessity, anyone who can be said to have beliefs and desires at all must be an ideally rational *homo economicus*! Our rationality is very imperfect, Strawman knows it, and he knows that the folk know it too. Of course we overlook options and hypotheses, we practice inference to the third-best explanation, we engage in double think, and so on, and on, and on.

But there is no cause for alarm. Folk psychology can be taken as a theory of imperfect, near-enough rationality, yet such rationality as it does affirm can still be constitutive. And even if folk psychology did set too high a standard – even if, to take the worst case, it were a theory of ideal rationality – still an imperfect but near-enough occupant of a folk-psychological role could thereby be an imperfect but near-enough deserver of a folk-psychological name. Remember also that the *typical* occupant of a role needn't occupy it in every case. In short, constitutive rationality leaves plenty of room for human folly.

(I think that systematic theories of ideal rationality – decision theory, for instance, and the theory of learning from experience by conditionalizing a subjective probability distribution – are severely idealized versions of parts of folk psychology. They are founded upon our tacit knowledge of folk psychology, elicited in the guise of 'intuition'. But folk psychology also supplies the grains of salt to be applied to these idealizations. Sometimes it supplies complementary pairs of opposite idealizations: a quantitative theory of subjective probabilities and utilities precise to however many decimal places, and alongside it a non-quantitative theory of beliefs and desires that don't admit of degree at all.)

Constitutive rationality is part of the legacy of behaviourism, and that is a second reason why Strawman mistrusts it. A behaviourist analysis might say, roughly, that a subject's beliefs and desires are those beliefs and desires, attribution of which would best make sense of how the subject is disposed to behave, and of how his changing behavioural dispositions depend on the changing perceptible features of his surroundings. But Strawman is a robust realist about beliefs and desires. He takes them to be genuine inner states, and causes of behaviour. He won't like an analysis that dispenses with efficacious inner states in favour of mere patterns of dispositions. Still less would he like it if the behaviourist went on to say that attributions of belief and desire governed by constitutive rationality were instrumentally useful, or warranted by rules of assertability, but not straightforwardly true.

I applaud these misgivings. I too am a robust realist about beliefs and desires. (About whole systems of beliefs and desires, anyway, though maybe not about all the little snippets – the sentences written in the belief and desire boxes – of which these systems may or may not be composed.) But

I say the proper remedy is not to shun constitutive rationality, but to apply it differently. The behaviourist applies it directly to the subject; I say we should apply it to the subject's inner state. The behaviourist says that the subject *has* that system of beliefs and desires that best makes sense of how the subject is disposed to behave. Whereas I'd say that the inner state *is* that system of beliefs and desires that best makes sense of the behaviour which that state is apt for causing in subjects. Thus I'd use constitutive rationality not to dispense with causally efficacious inner states, but rather to define their content.

A third reason why Strawman shuns constitutive rationality is that sometimes it needs to be applied not to the singular propositions that are the wide content of belief and desire, but instead to the underlying *de se* narrow content. The furniture of the *Lebenswelt* which presents us with our problems of decision and learning consists, in the first instance, of objects given *qua* objects of acquaintance, and individuated by acquaintance. (*See* Hintikka, 1972; Lewis, 1983b.) That is a matter of narrow content. If you are lucky, and you're never wrong or uncertain about whether you're really R-acquainted with something, and you're never wrong or uncertain about whether the thing you're R_1-acquainted with is or isn't the same as the thing you're R_2-acquainted with, then we can talk about your beliefs and desires entirely in terms of wide content. We can safely let things *simpliciter* stand in for things-*qua*-objects-of-acquaintance. But if you're not so lucky, that won't work. Take unlucky Pierre (Kripke, 1979). He self-ascribes being R_1-acquainted with a pretty city and being R_2-acquainted with an ugly city. But in fact he is R_1-acquainted and R_2-acquainted with the same city, London. Thereby he believes both that London is pretty and that London is ugly. (Kripke derives this conclusion from certain premises, but I find the conclusion at least as obvious as the premises.) I take this to be a conflict in wide content: Pierre widely believes two singular propositions that predicate conflicting properties of the same thing. Folk psychology says that by careful attention we can detect and eliminate conflicts in our beliefs – especially if we're good at logic, as Pierre is. But plainly that was never meant to apply to Pierre's conflict of singular propositions. Mere thought can't save him. What he needs is the information *de se* that he is R_1-acquainted and R_2-acquainted with the very same thing. (*See* Lewis, 1981.)

And suppose Pierre believes that by boarding the bus before him, he can be taken to London for a week of sight-seeing. Would boarding that bus serve his desires according to his beliefs? It helps not at all to know that he widely believes both that London is pretty and that London is ugly. What does help is the information already given about his narrow self-ascriptions, plus one further thing: he also self-ascribes having a bus before him that would take him to the place he is R_1-acquainted with. (*See* Lewis, 1986a, p. 58.)

(On constitutive rationality, *see* Stalnaker, 1984, pp. 1–42; Lewis, 1974, 1986a, pp. 27–40. But see Lewis, 1974, with caution: it began as a conversation with Donald Davidson, and I went rather too far in granting undisputed common ground. (1) I gave an important place to the subject Karl's beliefs as expressed in Karl's own language; that certainly suggests language-of-thoughtism, though I hope I committed myself to nothing more than the safe thesis that Karl's medium of mental representation is *somehow* analogous to language. (2) I was too individualistic: I ignored the possibility that deviant Karl might believe something in virtue of the causal role of his inner state not in Karl himself but in others who are more typical members of Karl's kind. (3) I had not yet come to appreciate the role of *de se* content. *Also see* Lewis, 1986a, pp. 27–40, with caution: besides endorsing constitutive rationality, I also stated it within a controversial framework of realism about unactualized *possibilia*. I still think that's a good way to state it; but I never said it was the only way. Constitutive rationality and realism about *possibilia* needn't be a package deal!)

This completes my list of objections against Strawman's program for explaining content. Doubtless you can think of ever so many ways of amending Strawman's theses to get around my objections. Some lists of amendments would take us to the positions really held by real people. Of course I can't show that no version of Strawman-amended can work. But for myself, I pin my hopes on a more radical reversal of Strawman's position.

With Strawman for a foil, my own approach can be summed up quickly. The contentful unit is the entire system of beliefs and desires. (Maybe it divides up into contentful snippets, maybe not.) That system is an inner state that typically causes behaviour, and changes under the impact of perception (and also spontaneously). Its content is defined, insofar as it is defined at all, by constitutive rationality on the basis of its typical causal role. This content is in the first instance narrow and *de se* (or *de se et nunc* if you'd rather steer clear of momentary selves). Wide content is derivative, a product of narrow content and relationships of acquaintance with external things.

See also DAVIDSON; DENNETT; FODOR; PHILOSOPHY AND PSYCHOLOGY; PROPOSITIONAL ATTITUDES; REASONS AND CAUSES; THOUGHTS; THOUGHT AND LANGUAGE.

BIBLIOGRAPHY

Anscombe, G.E.M. 1958. *Intention*. Oxford: Blackwell.

Armstrong, D.M. 1968. *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.

Carnap, R. 1963. Replies and expositions. In *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp. Cambridge University Press, 859–1013.

Chisholm, R.M. 1979. The indirect reflexive. In *Intention and Intentionality: Essays in Honour of G. E. M. Anscombe*, ed. C. Diamond and J. Teichman. Brighton: Harvester.

Davidson, D. 1987. Knowing one's own mind. *Proceedings and Address of the American Philosophical Association*, 60, 441–58.

Davies, M.K., and Humberstone, I.L. 1980. Two notions of necessity. *Philosophical Studies*, 38, 1–30.

Feinberg, G. 1966. Physics and the Thales problem. *Journal of Philosophy*, 66, 5–13.

Field, H. 1973. Theory change and the indeterminacy of reference. *Journal of Philosophy*, 70, 462–481.

Hintikka, J. 1972. Knowledge by acquaintance – individuation by acquaintance. In *Bertrand Russell: A Collection of Critical Essays*, ed. D. Pears. Garden City, N.J.: Doubleday.

Jackson, F.C. 1982. Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–36.

Jackson, F.C. 1992. Armchair metaphysics. Presented at a conference on the place of philosophy in the study of the mind, University of New South Wales.

Jackson, F.C., Pargetter, R., and Prior, E.W. 1982. Functionalism and type-type identity theories. *Philosophical Studies*, 42, 209–25.

Kripke, S. 1972. Naming and necessity. In *Semantics of Natural Language*, ed. D. Davidson and G. Harman. Dordrecht: Reidel.

Kripke, S., 1979. A puzzle about belief. In *Meaning and Use*, ed. Avishai Margalit.

Lewis, D. 1966. An argument for the identity theory. *Journal of Philosophy*, 63, 17–25. Reprinted with additions in Lewis, 1983c.

—— 1970. How to define theoretical terms. *Journal of Philosophy*, 67, 427–46. Reprinted in Lewis, 1983c.

—— 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249–58.

—— 1974. Radical interpretation. *Synthèse*, 23, 331–44. Reprinted with postscripts in Lewis, 1983c.

—— 1979a. Attitudes *de dicto* and *de se*. *Philosophical Review*, 88, 513–43. Reprinted with postscripts in Lewis, 1983c.

—— 1979b. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 339–59. Reprinted in Lewis, 1983c.

—— 1980. Mad pain and Martian pain. In *Readings in Philosophy of Psychology. Vol. 1*, ed. N. Block. Cambridge, MA: Harvard University Press. Reprinted with postscript in Lewis, 1983c.

—— 1981. What puzzling Pierre does not believe. *Australasian Journal of Philosophy*, 59, 283–89.

—— 1983a. New work for a theory of uni-

versals. *Australasian Journal of Philosophy*, 61, 343–77.
—— 1983b. Individuation by acquaintance and by stipulation. *Philosophical Review*, 92, 3–12.
—— 1983c. *Philosophical Papers, Vol. 1*. Oxford University Press.
—— 1986a. *On the Plurality of Worlds*. Oxford: Basil Blackwell.
—— 1986b. Causal explanation. In Lewis, *Philosophical Papers, Vol. 2*. Oxford University Press.
—— 1986c. Events. In Lewis, *Philosophical Papers, Vol. 2*. Oxford University Press.
—— 1990. What experience teaches. In *Mind and Cognition: A Reader*, ed. W. G. Lycan. Oxford: Basil Blackwell.
—— 1992. Critical notice of D. M. Armstrong, *A Combinatorial Theory of Possibility*. *Australasian Journal of Philosophy*, 70, 211–24.
Nemirow, L. 1990. Physicalism and the cognitive role of acquaintance. *In Mind and Cognition: A Reader*, ed. W.G. Lycan. Oxford: Basil Blackwell.
Perry, J. 1977. Frege on demonstratives. *Philosophical Review*, 86, 474–97.
Putnam, H. 1975. The meaning of 'meaning'. In *Language, Mind and Knowledge*, ed. K. Gunderson. Minneapolis, MN: University of Minnesota Press.
Ramsey, F.P. 1931a. Theories. In Ramsey, *The Foundations of Mathematics*. London: Routledge & Kegan Paul.
Ramsey, F.P. 1931b. General propositions and causality. In Ramsey, *The Foundations of Mathematics*. London: Routledge & Kegan Paul.
Rieber, S. 1992. A test for quotation. *Philosophical Studies*, 68, 83–94.
Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review*, 68, 141–156.
Stalnaker, R. 1978. Assertion. *Syntax and Semantics*, 9, 315–32.
Stalnaker, R. 1984. *Inquiry*. Cambridge, MA.: M.I.T. Press.
Tichý, P. 1983. Kripke on necessity a posteriori. *Philosophical Studies*, 43, 225–41.
Unger, P. 1984. *Philosophical Relativity*. Minneapolis, MN: University of Minnesota Press.

Thanks to the Boyce Gibson Memorial Library and the philosophy department of Birkbeck College; and to the editors, Ned Block, Alex Byrne, Mark Crimmins, Allen Hazen, Ned Hall, Elijah Millgram, Thomas Nagel, and especially Frank Jackson.

DAVID LEWIS