LEWIS' STRAWMAN

By Eric Margolis and Stephen Laurence

In a survey of his views in the philosophy of mind, David Lewis criticizes much recent work in the field by attacking an imaginary opponent, 'Strawman'. His case against Strawman focuses on four central theses which Lewis takes to be widely accepted among contemporary philosophers of mind. The theses concern (1) the language of thought hypothesis and its relation to folk-psychology, (2) narrow content, (3) de se content, and (4) rationality. We respond to Lewis, arguing (amongst other things) that he underestimates Strawman's theoretical resources in a variety of important ways.

I. INTRODUCTION

In a recent survey of his own work in the philosophy of mind, David Lewis critically discusses the views of an imaginary philosopher he calls 'Strawman'.[1] Strawman is primarily intended to serve as a foil for presenting Lewis' own views, but as Lewis notes, Strawman resembles a number of real philosophers. Though no particular philosopher may hold exactly the constellation of views which Strawman does, 'very many are to be found in his near vicinity' (p. 308). As a result, Lewis' critique of Strawman can be seen as a general critique of much of contemporary philosophy of mind.

Here is what Strawman believes (pp. 308–9). First, there is a language of thought, that is, an internal system of representation which has language-like syntactic and semantic structure. Secondly, whether a sentence in the language of thought is a belief or a desire is a matter of its having an appropriate functional role (the belief role or the desire role). Thirdly, words in the language of thought have their content in virtue of causal links to their referents in accordance with a theory of content in the Kripke–Putnam tradition.[2] Fourthly, all of these commitments are supposed to be part of 'folk-psychology', the set of common-sense principles that guide everyday thinking about the mind and behaviour.

[1] D. Lewis, 'Reduction in Mind', repr. in his *Papers in Metaphysics and Epistemology* (Cambridge UP, 1999), pp. 291–324.
[2] See S. Kripke, *Naming and Necessity* (Harvard UP, 1972); H. Putnam, 'The Meaning of "Meaning"', in his *Philosophical Papers*, Vol. II (Cambridge UP, 1975), pp. 215–71.

Lewis' case against Strawman is built around four objections. They are:

1. Strawman erroneously takes folk-psychology to be committed to the language of thought
2. Strawman takes all content to be wide, but some is narrow (how exactly this distinction is to be drawn is a matter of controversy, but for Lewis the difference is that the wide content of a belief depends upon the external things the thinker is acquainted with, while narrow content is independent of any such acquaintance)
3. Strawman has no account of *de se* content, i.e., 'content that concerns not the world but oneself' (p. 317)
4. Strawman ignores rationality, and cannot account for the fact that rationality is constitutive of content.

Having set out these objections, Lewis sums up his own positive account as follows (p. 324):

> With Strawman as 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, in so far as it is defined at all, by constitutive rationality on the basis of its typical causal role. The 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.

Though we think Lewis' positive account is problematic in a number of important respects, our concern here is with Lewis' critique of the family of views associated with Strawman. We argue that none of Lewis' objections raises serious problems for Strawman; despite Lewis' criticisms, the general theoretical orientation that Strawman represents remains perfectly cogent.

II. FOLK-PSYCHOLOGY, INTUITION AND THE LANGUAGE OF THOUGHT HYPOTHESIS

We shall begin with Lewis' first objection to Strawman, namely, that folk-psychology is not committed to a language of thought. This is certainly correct. Folk-psychology may traffic in intentional states, but it does not hold that these states are, or correspond to, sentences in the head. Lewis draws this point out by noting that there is an alternative to the claim that beliefs have sentence-like structure. His suggestion is that they may instead be *map-like*, following Frank Ramsey's remark that 'beliefs are ... "a map

... by which we steer"'.[3] Lewis does not have a lot to say about what being map-like consists in, or what the resulting theory of mind looks like, but he does note that maps represent things in such a way that individual parts of a map do not express individual truth-evaluable contents. For instance, you cannot isolate a part of a map of California that represents just the content that San Francisco is north of Los Angeles. Any subsection of the map that represents this one relationship will invariably express other information about these cities and the surrounding area. In contrast, the sentence 'San Francisco is north of Los Angeles' does not say anything about their size, their distance from other landmarks, etc. It expresses just the content that San Francisco is north of Los Angeles.

Though we do not really want to press the point here, it is worth noting that Lewis' map view faces a number of serious problems. For one thing, it has difficulty accommodating the very observation Lewis makes regarding sentences. Suppose that the sentence 'San Francisco is north of Los Angeles' does express just the content that San Francisco is north of Los Angeles. Given also the reasonable assumptions that language expresses thought and that understanding a sentence generally involves having a thought with the same content as that sentence, there must be a *thought* which has just this content as well. So it is not just sentences that contrast with maps. Thoughts do too: that is, thoughts have individual truth-evaluable contents. Map-like representation is also problematic in that it is limited in the sorts of contents it can accommodate. For example, what property of a map corresponds to the belief that it is not true that either determinism is false or people do not have free will? Or to the belief that Sam believes that Cathy said that Dave hoped that she would win? Or to the belief that if you were to put some sugar in this coffee it would dissolve?

These problems suggest that map-like representation is not as strong an alternative as Lewis thinks. None the less our main concern is not with its prospects *per se* but with the significance which Lewis attaches to it. He argues that if his map view *is* a live option, then the language of thought hypothesis cannot be part of folk-psychology. Rather it is an *empirical* question whether thoughts have language-like structure. Of course Strawman could revise his view by maintaining that ultimately he is not interested in *a priori* analysis, and that he is perfectly happy to endorse the language of thought hypothesis as an empirical thesis. (Among real philosophers, this would not really be a revision, since the language of thought hypothesis is almost universally understood to be an empirical thesis.) However, as Lewis sees it, this creates another set of problems (p. 324):

[3] Lewis, 'Reduction in Mind', p. 310, quoting Ramsey, 'General Propositions and Causality', in his *Philosophical Papers*, ed. D.H. Mellor (Cambridge UP, 1990), p. 146.

> 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'. 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.

In other words, if philosophical theories of belief are to be tied to empirical research in cognitive science, philosophers cannot rely on their intuitions, as they often do, in the construction and evaluation of a theory of belief.

Unfortunately, this first objection to Strawman is deeply flawed, and for a number of reasons. First, suppose that Strawman continues to embrace some form of conceptual analysis in appealing to folk-psychology in his enquiry into the nature of belief. There is still no reason why Strawman should think that folk-psychology is both exhaustive and entirely correct in what it says about belief. As Lewis himself is happy to admit, folk-psychology may be wrong about certain details. If there is not something that exactly fulfils the role which folk-psychology assigns to belief, we can still say that beliefs are the things which 'nearly' fulfil that role (see, e.g., p. 301). What this means, however, is that even if folk-psychology is agnostic about some matter, e.g., the dispute between sentence-like and map-like representation, this may just be a case of simple omission. So given Lewis' own way of construing conceptual analysis, there is plenty of room for believing in a language of thought without having to abandon conceptual analysis at all.

Secondly, we are rather puzzled by Lewis' implication that appealing to folk-psychological principles is incompatible with empirical enquiry. Folk-theories of the sort that philosophers like Lewis are concerned with are, generally speaking, subject to the influence of empirical or *quasi*-empirical theorizing. Folk-psychology is no exception. This is partly what accounts for the many modifications that have occurred in folk-psychology over the years (witness the introduction of the four humours, Freudian mechanisms of repression, parental style as a determinant of personality, and so on). Lewis tells Strawman that he has to choose between folk-psychology and cognitive science. But surely something has gone wrong. Why should Strawman not be allowed to do what the folk themselves do under the impact of culturally propagated theories? If they can incorporate empirical findings, so can Strawman.

Thirdly, the conflict between folk-psychology and empirical enquiry is spurious in another way. It is agreed on all sides that empirical investigation is relevant to determining which entities realize the roles that folk-psychology specifies for such states as belief and desire. Lewis generally assumes that these entities will be neurological states (perhaps relativized to species), but empirical enquiry could equally reveal that the entities in question are mental sentences. In that case the empirical hypothesis that there is a language of thought would be on a different level from Lewis' folk-psychology. It would provide part of the account of how folk-psychological states are implemented.

In short, there is no reason why Strawman has to give up conceptual analysis. At the same time, Strawman need not be worried about abandoning conceptual analysis and forming a more intimate alliance with cognitive science. Lewis simply begs the question against his straw opponent by saddling him with the view that '"belief" is a folk-psychological name for a kind of state posited by folk-psychology'. After all, Strawman adopts a causal theory of reference. For him, beliefs are whatever entities stand in the appropriate causal relations with our concept or term 'belief'.[4] For this reason, Strawman should maintain that it is no mystery that we can discover new facts about beliefs. Allowing for such discoveries is one of the characteristic selling points of causal theories.

Suppose, then, that Strawman decides to treat the language of thought hypothesis as an empirical thesis. Having made this decision, he may or may not be inclined to 'set up thought-experiments and ask us what we want to say about them'. Some philosophers in his vicinity are sceptical about thought-experiments and any reliance on intuition. But Strawman would not be *forced* to renounce the use of intuitions. Even within a broadly empirical framework, and even assuming a causal theory of reference, intuitions can provide evidence concerning the kinds that our terms and concepts pick out. One way in which they do this is by correlating with the factors that determine the application of our terms and concepts.

For example, on a nomic–causal theory of content, categorization dispositions are not constitutive of content *per se*; content is constituted by the existence of a nomic dependence between a mental representation and the property it expresses. All the same, this relation is generally mediated by cognitive mechanisms that incorporate our categorization dispositions. As a result, categorization dispositions correlate with content to some

---

[4] In this he agrees with William Lycan (among others): 'As in Putnam's examples of "water", "tiger", and so on, I think the ordinary word "belief" (*qua* theoretical term of folk-psychology) points dimly towards a natural kind that we have not fully grasped and that only mature psychology will reveal': Lycan, *Judgement and Justification* (Cambridge UP, 1988), p. 32.

significant extent. Since it is also generally agreed that intuitions reflect our categorization dispositions, it follows that intuitions offer *prima facie* evidence regarding the referents of our concepts. In this way, intuition may very well tell us something about belief.[5] Importantly, however, intuitions are not entirely trustworthy, since their correlation with content is less than perfect. One problem is that intuitions do not always track categorization. What we actually do say about what we would say (i.e., in some hypothetical situation) is not always right. Another problem is that intuitions are subject to variation from a great variety of cognitive and emotional influences, including many different sorts of reasoning biases.[6] Finally, and most significantly, categorization does not always track content, since categorization can be guided by incomplete or erroneous information, information which is subject to correction by empirical enquiry. The implication, however, is not that intuition is to be disregarded as a source of evidence. Rather it is that the evidence intuition provides is limited and empirically defeasible.

In sum, though folk-psychology is not committed to a language of thought, that need not be a problem for Strawman. He can endorse the language of thought hypothesis as an empirical claim *and* continue to rely on intuitions whether he goes in for conceptual analysis or not.

III. WIDE AND NARROW CONTENT

Lewis' second objection is that Strawman's account is deficient because it delivers only wide content. Lewis claims that we need to embrace narrow content as well. His discussion is based on a comparison of you and your twin earth *Doppelgänger*, you and your Swampman counterpart, and you and your brain in a vat. In each case there will be a difference in wide content, because you and your counterparts are acquainted with different things. Yet, as Lewis sees it, you and your counterparts share contentful states too. For example, you and your twin share the belief that 'the stuff he has heard under the name "water" falls from clouds' (p. 316). Such beliefs are narrow, since a difference in acquaintance does not render a difference in content. The stuff he is acquainted with is XYZ, while the stuff you are acquainted with is H_2O, yet this difference has no bearing on whether you share the belief in question. For Lewis, this means that an adequate theory of the mind cannot do without a notion of narrow content.

[5] For further discussion, see Laurence and Margolis, 'Concepts and Conceptual Analysis', *Philosophy and Phenomenological Research*, forthcoming.

[6] See, e.g., D. Kahneman *et al.* (eds), *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge UP, 1982); R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs: Prentice Hall, 1980).

The first thing to note with regard to this criticism is that it applies only to a subset of the positions adopted by the real philosophers who have an affinity with Strawman. In particular, it does not apply to two-factor theories, which maintain that mental content has both a narrow and a wide component, despite the fact that such theories clearly dominate the general theoretical neighbourhood which Strawman represents.[7]

Another problem is that even among theorists who shun narrow content, there are a number of responses which can be made to deal with the sorts of examples that generate Lewis' objection. For example, Jerry Fodor's theory of wide content, based on nomic causal relations, has no problem with Swampman, since Swampman's internal states enter into the very same nomic dependencies as his counterpart's.[8]

Yet a more general problem for Lewis is this: Strawman could respond that the commonalities on which Lewis depends may not be a matter of twins (etc.) sharing contentful states, but rather their sharing a dispositional property. For example, twin-you, swamp-you and vat-you all share the property that if they were embedded in your actual current and historical environment, they would have the same contentful states as you have now. This property could in turn explain why twins (etc.) engage in corresponding behaviour and state-transitions, and in general why they seem to share so much of their psychology.

So whether Strawman supplements his causal account of wide content with narrow content or not, Lewis' demands for narrow content fail to undermine Strawman's position. One might also ask whether Strawman, or anyone, should be all that much concerned with the sorts of considerations which Lewis cites on behalf of narrow content. After all, twins, brains in vats and swamp creatures are all bizarre philosophical inventions. Do we really want our theorizing about the mind to be dominated by them? Perhaps the right thing to say is that our intuitions in these cases are not of much interest and that they may simply have to be disregarded.[9] What is more, if Strawman opts for a causal theory of reference, then he has all the more reason to be sceptical about these intuitions. For him, the nature of belief is ultimately an empirical question, one which is no more subject to the

---

[7] See, e.g., C. McGinn, 'The Structure of Content', in A. Woodfield (ed.), *Thought and Object* (Oxford UP, 1982), pp. 207–58; B. Loar, 'Conceptual Role and Truth Conditions', *Notre Dame Journal of Formal Logic*, 23 (1982), pp. 272–83; and especially N. Block, 'Advertisement for a Semantics for Psychology', in P. French *et al.* (eds), *Midwest Studies in Philosophy*, Vol. x: *Studies in the Philosophy of Mind* (Univ. of Minnesota Press, 1986), pp. 615–78.

[8] Fodor, *The Elm and the Expert* (MIT Press, 1994).

[9] See, e.g., D. Dennett, 'Cow-sharks, Magnets, and Swampman', *Mind and Language*, 11 (1996), pp. 76–7; and D. Papineau, 'Doubtful Intuitions', *Mind and Language*, 11 (1996), pp. 130–2.

pressure of intuition than any other empirical enquiry. Since science in general is not guided by such fantastical examples, why should it be any different with the science of the mind?

IV. *DE SE* CONTENT

Lewis' third objection is that Strawman cannot account for *de se* content. No doubt there is a substantial difference in thoughts about oneself that occur in a first-person and a third-person way. In particular, one has to allow that there is a big difference between, say, Joe's believing *Joe is Joe* and his believing *I am Joe.* He might very well believe the first, but not the second, if he does not know who he is. As Lewis notes, 'However much I may know about the things that make up the world, their properties and their arrangements, it is something extra to know which one of all these things is me' (p. 317).

Be that as it may, there is no reason why Strawman has to say that the I-thought and the Joe-thought are indistinguishable. The difference between such thoughts seems to be largely a matter of differing functional roles. However, we have already pointed out that among the real theorists in Strawman's vicinity, many are happy to supplement a causal theory of wide content with an account of narrow content. And narrow content is typically construed in terms of functional role. This suggests that Strawman might say that the representation corresponding to 'I' has a distinctive functional role and that this role accounts for the psychological difference between having a first-person and a third-person thought about oneself. A further advantage of this explanation is that it accounts for the simil- arity between your thinking *I am tired* and someone else's thinking *I am tired.* Though these thoughts have different wide contents, they may be said to have the same narrow content or functional role.[10] Besides, even theorists who shun narrow content and favour a (relatively) pure causal theory appeal to functional role to explain the contents of some terms. For instance, Fodor offers a functional-role account of the content of logical connectives.[11] There is no reason why Strawman could not account for *de se* content in much the same way, that is, by adopting a restricted amount of functional role.

[10] See Block, 'Advertisement for a Semantics for Psychology', for a development of a two- factor theory that is partly supported by exactly these sorts of cases.

[11] Fodor, 'A Theory of Content, 2', in his *A Theory of Content and Other Essays* (MIT Press, 1990), pp. 89–136.

V. RATIONALITY

Lewis' fourth and final objection is that Strawman ignores the parts of folk-psychology that concern human rationality. In particular, Strawman is supposed to err in his failure to recognize that rationality is constitutive of mental content. Here is how Lewis puts the argument (p. 321):

> 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.

This argument is flawed in a number of different ways. One thing to note is that the most it could establish is that it is constitutive of the *state*, e.g., being a belief that *p*, that this must satisfy certain constraints of rationality, not that those constraints are constitutive of *the content* of this state. The latter is an additional inference for which Lewis provides no justification.

A far more serious problem is that Lewis' argument simply begs the question against Strawman, since it clearly *presupposes* that content is simply a matter of functional role. But this is to ignore the substantial role that Strawman assigns to the causal–external determinants of content (which ironically Lewis himself emphasizes in his previous objections). For this reason alone, Strawman should not be moved by Lewis' argument.

Moreover, even if Lewis were not begging the question, his argument would not work, since Strawman is free to deny that mental states are necessarily constrained by any particular aspects of the folk-psychological roles which conceptual analysis might assign to them. As we argued earlier, there is no need for Strawman to buy into Lewis' project of conceptual analysis, and certainly few of the philosophers who identify with Strawman do. (An anonymous referee has questioned whether this criticism is consistent with our earlier claim that intuitions may still be of use in investigating the kinds to which our concepts refer. But the conflict is spurious, as the role that we identified for intuitions is far from the one that Lewis employs, and in no way vindicates Lewis' commitment to conceptual analysis. On our model, intuitions offer only highly defeasible evidence concerning the kinds our concepts pick out.)

Lewis' charge that Strawman ignores those parts of folk-psychology concerning rationality carries an additional irony. In point of fact, language of thought theorists have an explicit general model of how rationality could be mechanized and thereby of how it might fit into the natural world. Some

would even go so far as to argue that computationalism taken in conjunction with the language of thought hypothesis is the *only* plausible general model of rationality that anyone has ever proposed.[12] The idea is that thought processes are to be treated as rule-governed operations over syntactically structured representations. Because valid inferences can be mimicked by syntactic relations, a physical device that is sensitive only to syntactic properties can draw valid inferences all on its own, respecting the semantic properties of the states involved. The advantage of this model of thinking is that it shows how the state-transitions in a mind can be truth-preserving and yet occur without the need for a mysterious inner agent or homunculus who appreciates their contents.

Finally, it should also be noted that rationality does not always play an important role in accounting for the transitions among contentful psychological states. Humans are subject to various sorts of irrationality – the gambler's fallacy, the fallacy of ignoring base rates, etc. So humans are, at best, imperfectly rational. Lewis agrees, claiming that this imperfect rationality is still constitutive of content (p. 321). But what he fails to see is that many psychological processes are neither rational nor irrational; they are arational. For example, it is hardly appropriate to evaluate the processes of visual object recognition, or face perception, or syntactic parsing, in terms of how rational they are. To be sure, they involve transitions among contentful states – transitions that depend on the particular contents that these states have – but these are hardly 'rational' inferences. Instead it seems that the best one can say is that they just work in the way they do.

In sum, contrary to what Lewis claims, Strawman has a balanced, well supported and sophisticated treatment of the place of rationality in the study of the mind.

VI. CONCLUSION

Lewis ends his discussion of Strawman with the following remark (p. 324):

> 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 cannot 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.

It is of course true that with enough *ad hoc* qualifications and amendments, just about any philosophical view can be made to withstand criticism. But

---

[12] See, e.g., Fodor, *The Mind Doesn't Work That Way: the Scope and Limits of Computational Psychology* (MIT Press, 2000).

Strawman is not in this position at all. In fact there is nothing in Lewis' critique that poses a serious challenge to the general theoretical orientation which Strawman represents. The fact that folk-psychology is agnostic about the language of thought hypothesis is not a worry, whether Strawman goes in for conceptual analysis or for some more naturalistic alternative. Nor should Strawman be worried about Lewis' insistence on narrow content; there are a variety of perfectly plausible options for Strawman to take in response to these worries. *De se* content can also be accommodated by supplementing a causal account of content with a bit of functional role, or by embracing a more thoroughgoing two-factor theory; but this is exactly what real philosophers with Strawman's outlook do anyway. Finally, in spite of Lewis' claims, Strawman can hardly be seen as ignoring rationality. Indeed, Strawman arguably has the most detailed and plausible account available of how rationality is compatible with a broadly materialist framework. Far from defending a losing position, Strawman has a powerful and compelling picture of the mind. Perhaps that is why he has so many allies.[13]

Rice University; University of Sheffield