IX*—MENTAL CAUSATION AND MENTAL REALITY

by Tim Crane

I

The Problems of Mental Causation. Functionalism in the philosophy of mind identifies mental states with their dispositional connections with other mental states, perceptions and actions. Many theories of the mind have sailed under the Functionalist flag. But what I take to be essential to Functionalism is that mental states are individuated *causally*: the reality of mental states depends essentially on their causal efficacy.

But the very idea of mental causation has long been thought problematic, and has recently been challenged by a number of new arguments. In their general outline, these arguments go like this. A cause has its effects only in virtue of some of its properties. When I burn my hand on a red hot poker, it is not the colour of the poker that causes the burn, but its heat. So when a mental cause has some effect, we can ask: does it have this effect in virtue of its mental properties, or in virtue of its non-mental properties? Is the *mentality* of the mental cause 'efficacious' in producing its effect? The opponents of mental causation argue that it is not. But if the mental properties of these causes are inefficacious, then there is strictly speaking no genuinely mental causation. Our thoughts and sensations never literally make things happen.

If there is no mental causation, then Functionalism's causal conception of mental reality is mistaken. So it is crucial for Functionalists to understand why mental causes are supposed not to have their effects in virtue of their mental properties.

*Meeting of the Aristotelian Society held in the Senior Common Room, Birkbeck College, London, on Monday, 24th February, 1992 at 8.15 p.m.

In the recent literature, a number of distinct assumptions have been offered in support of this claim.[1] They are:

> (1) *The Anomalism of the Mental*: the mental is not subject to laws, neither to purely mental laws nor to mental–physical laws.

It is generally accepted that the existence of a causal relation between A and B entails that A and B fall under some law. So on the face of it, if there are no mental laws, then there is no mental causation.

> (2) *The 'Causal Closure' of the Physical*: all physical effects have sufficient physical causes.

I move my hand away from the fire because I am in pain. But if physical causes alone suffice, then surely the *pain* is unnecessary for this effect?

> (3) *The Inefficacy of the Functional*: functional properties themselves do not have effects, only the 'structural' properties that 'realise' them are efficacious.

Being a mousetrap is a functional property which may be 'realised' by e.g. an arrangement of wood, springs and cheese. It is not the *mousetrapness* of the mousetrap which is responsible for its effects, but the 'realising' properties of the wood, the spring etc. Likewise, it is claimed, with mental properties—if they are functional properties.

> (4) *The Inefficacy of the Semantic*: the 'semantic' properties of mental states cannot be responsible for the effects of those states.

When I think about Chicago, the 'semantic' properties of my thought are those in virtue of which it represents Chicago. But how can the fact that something represents something else have any effects? Nothing short of magic, it seems, could get a relation in which I stand to Chicago to be immediately responsible for what I do here in London.

1 For (1), see Davidson 1970, Macdonald and Macdonald 1986 and Le Pore and Loewer 1987. For (2), see Papineau 1990, and Kim 1991. For (3), see Jackson and Pettit 1990, and Block 1990. For (4), see Dretske 1988, Fodor 1990, Block 1990, and Segal and Sober, forthcoming.

Some of these assumptions are related, though not in a very straightforward way. Assuming that causes need laws and that the mental-physical distinction is exhaustive, then assumption (1), the *Anomalism of the Mental*, obviously entails (2), the *Causal Closure of the Physical*. But (1) does not entail (3), the *Inefficacy of the Functional*: one could hold that functional properties are efficacious, as long as none of them are mental. (1) does, however, entail (4), the *Inefficacy of the Semantic*, on the assumption that all semantic properties are, or derive from, mental properties.

Assumption (2), *the Causal Closure of the Physical*, does not entail (1): one can believe in *non-causal* mental laws. Nor does (2) entail (3): one could believe that there are efficacious functional physical properties. But (2) does entail (4), again on the assumption that all semantic properties are mental.

Assumption (3), *the Inefficacy of the Functional*, does not entail (1) for the same reason that (2) doesn't—even if all mental properties are functional. And (3) does not entail (2), on the usual meaning of 'physical': one could hold that some physical effects have sufficient non-functional and non-*physical* causes. Neither does (3) entail (4), since the semantic may not be functional.

Finally, assumption (4), *the Inefficacy of the Semantic*, does not entail (1), if only because not all mental states are semantic. (4) does not entail (2), since non-semantic mental states (e.g. sensations) may be efficacious. And (4) does not entail (3), since many functional properties are not semantic.

So (1) entails (2) and (4) on reasonable assumptions, and (2) entails (4), but apart from that, the four assumptions are logically independent. What this means is that there is not one problem of mental causation—there are a number, some of which are unrelated to the others.

In the rest of this paper, I will critically examine these assumptions. Since I have elsewhere argued against (1) and (2), my aim here is not so much to refute them but to show exactly what is involved in accepting them. This will take up section II. I will then offer arguments against (3) in section III, in the course of providing a better understanding of Functionalism.

The upshot is that the friend of mental causation need not be troubled by (1)–(3). But (4) does give rise to a genuine problem of mental causation, and in section IV I say what it is. In sketching the

shape a solution to this problem should take, I briefly address in section V another live debate: that between 'Language of Thought' and 'Connectionist' conceptions of mental reality. I conclude that the conception of mental causation outlined in section III gives us a clearer view of what is at stake in this debate. The take-home message—mental causation vindicates mental reality—is not new. What may be new is the conception of mental reality defended.

II

Mental Laws And Mental Properties. As is well known, Anomalous Monism identifies token mental events with token physical events while denying that there are any 'strict' mental laws.[2] Yet it claims to allow mental causation. A number of critics have therefore asked: when a mental token has an effect, does it do so in virtue of its mental properties or in virtue of its physical properties? If it is in virtue of its mental properties, then we have psychophysical causal (and thus law-like) connections, which is inconsistent with the theory. If it is in virtue of its physical properties, then the mental is causally redundant. So Anomalous Monism is either inconsistent or epiphenomenalist (see Honderich 1982).

How good is this objection? If Anomalous Monism is committed to the existence of mental *properties*, then I think the objection is decisive. For if one believes both that events are mental in virtue of possessing mental properties, and that causation entails laws involving those properties, it is hard to see how one can keep the theory consistent yet avoid counting those *properties* as causally idle.

There have recently been a number of attempts to defend such a version of Anomalous Monism from this objection (see, e.g. Macdonald and Macdonald 1986 and Le Pore and Loewer 1987). But for reasons I have given elsewhere (Crane 199 1b), I think these

2 See Davidson 1970. 'Strict' in this discussion either means 'deterministic', or 'free from *ceteris paribus* clauses', and the paradigms of strict laws are the laws of basic physics. Since I believe that not all laws of physics are deterministic, and their statements *do* entail *ceteris paribus* clauses (see Wilson 1985 p. 240), I do not think there are any strict laws in this sense. So (*pace* McLaughlin 1989) the distinction is irrelevant.

responses do not get round this basic objection. If there are efficacious mental properties, there are mental laws: Anomalism must go. To my mind, this is no great loss, since there are plenty of excellent reasons for believing in psychological and psychophysical laws (see Crane and Mellor 1990, §4).

So an Anomalous Monism that is committed to mental properties is vulnerable to the charge of epiphenomenalism-or-inconsistency. But it is worth pointing out that *Davidson's* version of Anomalous Monism is not vulnerable to this charge. For to formulate the charge properly, you have to be able to ask: in virtue of *which property* of the cause does it have its effect? But this is a question that Davidson's nominalist ontology does not let you ask.

Davidson believes that singular causal relations hold between particular events, but they do not hold 'in virtue of' some property of these events.[3] Events, according to Davidson, are mental (or physical) 'only as described' (1970 p. 215): whether an event is a mental event depends on its being described in a mental vocabulary. And the mere description of an event is not responsible for its effects: it is just something we use in *explaining* the effects. For these descriptions do not, on Davidson's view, pick out *properties* of these events, since on his nominalistic view there are no properties. His Anomalous Monism therefore says that all events having mental descriptions have physical descriptions too. It does not mention properties.[4]

The only way to maintain Anomalous Monism, it seems, is to hold the following two contentious doctrines: (a) there are no properties,

3 *Cf.* 'How could Smith's actual fall, with Smith weighing, as he did, twelve stone, be any more efficacious in killing him than Smith's actual fall?' (Davidson 1967, p. 150). As Armstrong and Heathcote (1991, p. 68) put it, 'Davidson's Quinean nominalism does not enable him to say that it is properties of the cause and effect that are the main players in causation'. This point tells against Jackson and Pettit (1990 p.197n), and defenders of Davidson, such as Le Pore and Loewer (1987 p. 633) and McLaughlin (1989, p. 122) who mention it but ignore it in the remainder of their defence of Davidson. My impression is that these Davidsonians are (rightly) shy of accepting the radical consequences of Davidson's position. A notable exception is Melchert 1986.

4 'The theory under discussion is silent about processes, states and attributes if these differ from individual events' (Davidson 1970 p. 210). I use the term 'nominalism' for any doctrine that denies the existence of universals. A nominalist in this sense need not deny the existence of all abstract objects; and Davidson and Quine of course do not.

and (b) causation is a relation solely between particulars. It is the nominalist doctrine, (a), that makes Davidsonians say that mental events are not mental because they have some mental property—they are mental because they are described as such. And it is (b) that makes them say that the same particular cause can be correctly described in different ways, as mental or as physical.

Although I would dispute both these doctrines, I shall not do so here. Instead I shall focus on the crucial denial of properties. If there are no properties, then causation cannot depend on them. We cannot literally say, as I did in section I, that a cause has its effects in virtue of certain of its properties. But few facts about causation seem more obvious than this.[5] So if we do insist that causation depends on properties, then it turns out that the chief problem with Davidson's Anomalous Monism is not, as most critics claim, that it admits *inefficacious* mental properties, but that it does not admit mental properties at all.

Moreover, if causation does depend on properties, then we should also reject Davidson's doctrine (b) that it is a relation *solely* between particulars. The obvious alternative is to see causation as a relation between complex entities, particulars having properties. We can call these entities *facts*, and the properties the *constituents* of the facts.[6] It will not matter here which account of facts we adopt, since for the purposes of this discussion all that matters is that insofar as they are causes, facts have properties as constituents.

But what should we say about properties? Since most causal relations between facts hold independently of our discovery of them, a property cannot simply be explained in terms of the semantic value of a predicate (see Lewis 1983, Mellor 1991). For this would mean that there are at once too many and too few properties. Too many because there is no more reason to think that

5 See the useful discussion by Jackson and Pettit: 1990 pp. 196–197. Nominalists must account for this fact in whatever way they see fit: see Lewis 1983.

6 For the argument that the relata of singular causation are facts rather than particulars, see Mellor 1987. Bennett 1988 contains a useful discussion of events, facts and their place in causation. Some people, following Kim and others, call causes construed as property-instantiations 'events'. The point is terminological: I reserve the term 'events' for the particulars Davidson calls 'events', and call events in Kim's sense 'facts': see Bennett 1988, pp. 76–78.

all the predicates in a language pick out real properties than there is to think all singular terms pick out real particulars. Too few because there will be many properties that are not yet picked out by any predicate.

I cannot address the question of how we know when a real property is picked out by a predicate—except to say that it is not generally an *a priori* matter, any more than it is with singular terms and particulars. We learn of the existence of most properties empirically: through the effects on us of the facts of which they are constituents.

On this view, then, a mental cause is a fact with a mental property as constituent. Thinking of causation in this non-Davidsonian way certainly leaves no obstacles to mental causation. But if assumption (2) is right, then physical causes always suffice. So barring causal overdetermination, the mental reduces to the physical, or 'supervenes' upon it, or is wholly inefficacious (see Papineau 1990).

But it is not at all obvious why we should believe in (2). Essentially the problem is that the definition of 'physical' in (2) gives rise to a dilemma. Either the thesis that all physical effects have sufficient physical causes just is a definition of 'physical', in which case the the physical is the causal in a quite trivial sense. Or the 'physical' means the properties incorporated in the laws of physics proper, in which case (2) rules out perfectly adequate causes—for instance, biological causes—simply because they are not studied by physics.

But I have pursued this claim elsewhere, so I shall not repeat its details here.[7] My aim in this section has been to spell out just what the defenders of (1) and (2) are committed to, and to suggest how implausible these commitments are. Defenders of (1) must adopt the doctrines (a) and (b) above; and defenders of (2) will find themselves impaled on one of the horns of the above dilemma.

7 See Crane and Mellor 1990 and Crane 1991a. David Papineau (1990) argues for the definitional version, but claims that it is then not trivial what will count as physical. Jackson and Pettit (1990 pp. 210–213) assert, but do not argue for, the non-trivial version.

III

Functionalism And Mental Causation. That's all I shall say here about assumptions (1) and (2). To assess assumption (3) we must return to Functionalism, and its causal conception of mental reality.

As I said in section I, what is essential to Functionalism is the claim that mental states are identified by their dispositional relations to perceptions, actions and other mental states: these relations are often called their *causal roles*, or their *functional roles*. Consider, for example, Robert Stalnaker's correlative definitions of belief and desire:

> To desire that *P* is to be disposed to act in ways that would tend to bring it about that *P* in a world in which one's beliefs, whatever they are, were true. To believe that *P* is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which *P* (together with one's other beliefs) were true. (Stalnaker 1984 p. 15)

This could be one fragment of a Functionalist account of belief and desire. The full account would detail all the ways in which these states combine to produce their effects. This full account, with the mental terms removed and replaced by bound variables, would be the *Ramsey Sentence* for the theory of belief and desire (see Lewis 1970). The Ramsey Sentence for belief and desire, Functionalists claim, makes essential reference to what beliefs and desires cause. It thus specifies what beliefs and desires *are*—by specifying what being in those states disposes us to *do*.

Though they do, in a sense, define beliefs and desires, the clauses in the Ramsey Sentence need not all be known *a priori*. For we can find out more about these mental states by empirical investigation—facts about which introspective commonsense does not always deliver clear and unequivocal opinions. For instance, the complete Ramsey Sentence might tell us that beliefs need not be conscious, or that beliefs and desires come by degrees. These features of beliefs and desires need not be accessible to *a priori* conceptual analysis alone.

It might seem, therefore, that Functionalism is invulnerable to the problems of mental causation, since the functional roles specified in the complete Ramsey Sentence will include the typical effects of mental states, and therefore presuppose that they have

causal powers. So if the Ramsey Sentence is true, then mental causation is vindicated.

But assumption (3) says that functional properties are inefficacious (see Jackson and Pettit 1990, p. 201). The idea is that a functional property has to be 'realised' by some non-functional property in order for it to have any impact on the world. As Block puts it,

> Functional properties are properties that consist in the having of some properties or other (say non-functional properties) that have certain causal relations to one another and to inputs and outputs. In the production of these outputs, it is the non-functional properties that are standardly the causally relevant ones, not the functional properties. (Block 1990, p. 155)

This is why although a *mousetrap* is functionally characterised—as a device that takes unsuspecting live mice as input and yields dead mice as output—it is the properties that 'realise the mousetrap', or 'occupy the mousetrap-role', that are efficacious in these murderous transactions. If this is right, and mental properties are functional properties, then they have no effects in their own right—they need causal ratification by some other non-functional property.

This claim seems obviously true of mousetraps. But given the fact that properties are discovered *a posteriori*, the fact that the mousetrap's 'mousetrapness' has no effects might just mean that there is no such property as *being a mousetrap*. The objection has to show what general feature of *real* functional properties prohibits them from being efficacious. So can the argument be shown to work against *all* functional properties, and therefore against mental properties as conceived by Functionalism?

The argument needs the distinction between functional and non-functional properties (or between 'role' and 'occupant'/'realiser' properties) to be a distinction in nature. If there is a problem about the efficacy of functional properties, it should not arise just because of the way we happen to describe the property in question.

But it turns out that often whether a property is 'functional' depends precisely on the theoretical vocabulary chosen to define it. To borrow a nice example of Mark Wilson's, if Newtonian physics is formulated in the vocabulary of Newton's *Principia* ('F=ma' etc.) then *potential energy* is functionally defined. But in Hamiltonian or Lagrangian

formulations of physics, potential energy is a primitive notion, and the primitives of *Principia* (e.g. *gravitational force*) come out as functionally defined. So the question arises: is a system's potential energy a functional property of that system or not?[8]

To explain away compelling examples like these, the defender of (3) needs to provide a principle that identifies *essentially* functional properties: those properties that are essentially characterised in terms of the typical behaviour of the things that have them. We would then know which properties stand in need of causal ratification by non-functional properties. Since, as Jackson and Pettit observe (Jackson and Pettit 1990, p. 203, properties essentially characterised by the behaviour their possession produces are *dispositional* properties, what we need is a principle that distinguishes the essentially dispositional from the essentially non-dispositional.

But there is no such principle.[9] All the characteristic features of dispositions are possessed by the properties we identify as their 'categorical bases'. The theoretical descriptions that we give of these properties will include or entail conditionals that specify the difference that having these properties makes to the things that have them. Even a paradigmatically 'structural' or 'categorical' property such as *mass* is characterised, *inter alia*, in terms of the difference that its instantiation makes to a body's acceleration under a given force. But this is just to say that *having a certain mass* is characterised in terms of a disposition: in terms of what an object with that mass would do, if certain things were done to it.

This does not mean that dispositions do not have 'bases'. It's just that these bases only need be other dispositions. Rather than thinking of the base of a disposition, D, as a property, P, of an entirely other kind that 'realises' the disposition, we should think of P as just *another* property nomologically secured by the

8 Wilson 1985 p. 235. The idea that there is no distinction in nature between 'functional' and 'structural' properties is also well defended in Chapter 4 of Lycan 1987. Lycan is right to point out that the term 'structural' (rather like the term 'cognitive architecture') is crucially ambiguous. But I disagree with him that 'functional' and 'teleological' need go together here.

9 See Mellor 1974, to which I am indebted.

possession of *D*. And since this *P* will be another disposition, we should not then suppose that it is *P rather than D* that is efficacious, simply on account of *D*'s dispositionality.

The 'functionality' of functional properties amounts to their dispositionality. So if this claim about dispositions is right, then we either conclude that (3) is false, or that no properties are efficacious. And it is obvious what we should say: the fact that a property is functional does not *ipso facto* prevent it from being efficacious. So the efficacy of the mental is untouched by their functionality. If dispositionality goes 'all the way down' (Blackburn 1990 p. 63), then realism can go all the way up.

This claim can be illustrated by comparing mental properties with a property whose causal powers ought to be unproblematic: the property *temperature*. The traditional type-identity theory claimed that the identity-statement 'temperature = mean molecular kinetic energy' is an analogy for the identity of mental properties and brain properties. The standard objection is that mental properties, unlike temperatures, can be 'variably realised' from creature to creature—so a given mental property cannot be identical with one brain property.

But it has not been sufficiently recognised that the identity theory's claim about temperature is false. As Patricia Churchland points out,

> the temperature of gases is one sort of affair (mean kinetic energy of constituent molecules), but the temperature of a *plasma* cannot be the same, since a plasma is constituted not by molecules but by sub-atomic particles; the temperature of solids is different yet again, and so also is the temperature of *empty space* as embodied in its background black-body radiation. (Churchland 1982 pp. 101–102)[10]

So temperatures, like mental states, are variably realised across substances. Someone might conclude from this that temperatures are, contrary to appearances, inefficacious. But this is a hopeless manoeuvre. For the molecules whose mean kinetic energy 'realises' the temperature of a gas are in their turn 'realised' by atoms, and the

10 For other important correctives to the usual interpretation of the temperature analogy, see Enç 1983, p. 289; Wilson 1985 pp. 228–229.

atoms 'realised' by sub-atomic particles. And the mean kinetic energy itself will be 'variably realised' in different kinds of gases. On what principle, then, do we decide that one 'realising' level is the level of efficacious properties? Or are there, once again, no efficacious properties?

The answer is of course to block the regress by denying that the efficacy of temperature is undermined by these other microscopic facts about it. We should therefore draw two important morals from the temperature analogy. First, the property *temperature* cannot be identical with any of the realising properties. Second, and more relevant here, the fact that a mental property is 'variably realised' cannot render that property inefficacious, since parity of reasoning would falsely show that temperatures and other variably realised physical properties are inefficacious.

We would not, perhaps, get ourselves into the position of denying these obvious facts about temperature if we did not think in terms of an absolute 'role'–'realiser' distinction. I have claimed (following Wilson 1985 and Lycan 1987) that this is not a distinction in nature, but an artifact of theory. Of course, having a given temperature nomologically requires having certain other properties too. But to avoid the regress we should not think of these properties as—objectively—the 'realisers of the temperature-roles', but simply as other properties nomologically secured by a things's possession of temperature. And as with thermodynamic reality, so—I would claim—with mental reality.

IV

The Problem Of Intentionality: The Efficacy Of Content. So far I have claimed that three alleged problems of mental causation can be dissolved by adopting conceptions of properties, causation and dispositions that are in any case independently desirable. These general charges of epiphenomenalism, then, are no threat to Functionalism.

But what about our assumption (4), the claim that the 'semantic' properties of mental states are causally idle? Here we do encounter a genuine problem about mental causation: the problem of intentionality.

This problem arises because it is hard to see how those mental states that represent the world—the intentional states or propositional attitudes—can have their effects *in virtue of the fact* that they represent the world. The reasoning behind this line of thought is as follows. Intentional states seem to be essentially *relational*—they are relations to propositions or 'contents' (see Fodor 1985). It is in virtue of being relations to these propositions that intentional states represent the world. But, whether they are Fregean Thoughts, Russellian Propositions or sets of possible worlds, propositions are abstract objects, with no spatio-temporal location. And how can a relation to an abstract object have any effects? Our beliefs and desires do certainly cause our actions—but how, if they are relations to abstract propositions?

In current jargon, the 'semantic' features of intentional states are those features in virtue of which they represent the world, and their 'syntactic' features are some of those that don't. So if intentional states represent the world in virtue of being related to a proposition (or 'content') then assumption (4) may be expressed by saying that content is not relevant to the effects of intentional states: intentional states do not have their effects *in virtue of their contents*.

I think that this problem of the efficacy of content is the most troubling fact about intentionality. Intentionality is often described as mysterious—but mystery alone does not create a philosophical problem. The reason intentionality is so mysterious is that it seems to make the causal powers of beliefs and desires impossible. So the problem of intentionality is predominantly a causal problem: to solve the problem is to show how intentional states have the causal powers we know they have.

The first step in a solution, it seems to me, must show that each intentional state needs some *intrinsic* feature or features that serve as the mechanism for its interactions with other states, and ultimately with actions. These intrinsic features will be the *local causal surrogates* for the contents of the states. The Ramsey Sentence for intentional states will therefore need supplementing with specifications of these surrogate properties. I shall not say much here about what these surrogates are like, except that they need not be physical properties—though they may be. They just need to be intrinsic.

The problem of intentionality then reduces to the problem of showing how the local causal surrogates are surrogates for the contents of intentional states. Since causal surrogacy means that intentional states have semantic *and* non-semantic properties, these states may be thought of as incorporating *symbols* or *representations*. Thus the problem of intentionality then takes on its more usual form: how does any mental symbol or representation have its content?

Suppose this question can be satisfactorily answered. The worry may still be raised that on this story, it would be the causal surrogates that are 'doing all the causal work', and not the content: content will still not be efficacious.

But the content *itself* was never supposed to be efficacious. On any theory, contents are abstract objects—Fregean Thoughts, or sets of possible worlds—and as such are causally inert. What is supposed to be efficacious is the intentional *state*—the belief or the desire. Functionalism says that these states have certain characteristic causes and effects. I have claimed that for this to be true, the Ramsey Sentence for intentional states also has to specify certain intrinsic properties of their subjects which act as causal surrogates. This would be an account of what it is for an intentional state to have its effects *in virtue of its content*. For it is only as surrogates *for* abstract contents that the causal surrogates earn their keep. It is no part of this proposal that the causal surrogates could have their effects even if they were not representations (see Crane 1990). So the proposal would not eliminate *content*, but would (we realists hope) vindicate the causal efficacy of *states with content*.

V

Connectionism And The Language Of Thought. Of course, it is one thing dogmatically to say this, and quite another to show it. But if the proposal is on the right lines, it provides an obvious route from a Functionalist theory of the mind to a computational theory. For causal surrogates or mental symbols are postulated to account for the causal transactions between intentional states and actions. But these transactions have a certain systematic character, which the computational theory says is best explained by postulating

computations—causal sequences of symbols manipulated according to formal rules.

This may appear to be an *a priori* argument for Fodor's Language of Thought (LOT) hypothesis. This hypothesis says that mental states are computational relations to sentences of a mental language—items with semantic and 'syntactic' structure. Fodor himself has argued for the LOT by saying that there needs to be a mechanism for the relation between an organism and a proposition (see Fodor 1981 p. 202).

If the LOT hypothesis is understood as the claim that mental processes involving intentional states just need *some* intrinsic mechanism or other, then the argument of the last section is indeed an *a priori* argument for the hypothesis. (And indeed, sometimes the defenders of the hypothesis appear to see it this way—see Fodor 1987 p. 80). On this understanding of the hypothesis, 'mental syntax' is just another name for causal surrogacy.

Although there is nothing wrong with using the term 'syntax' in this way, it will mislead if it is not distinguished from syntax in a more specific sense. In this sense, an intentional state has syntactic structure iff it has non-semantic properties whose causal behaviour 'mirrors' the semantic structure of the state's content. So if I desire coffee and I think that this is coffee in front of me, the causal process involving these two states involves two representations which are tokens of the same semantic 'coffee'-type, *and* of the same syntactic type too.

It is this sense of 'syntactic' which the LOT hypothesis is more usually taken as intending. And obviously, the claim that there is mental syntax in this sense is a substantial empirical hypothesis, which the argument of section IV does not entail.[11] For it is by no means incoherent to suppose that the local causal surrogates of states with the same content have very little in common apart from the fact that they are surrogates for the same contents.

In fact this is exactly what some Connectionist accounts of mental processing say (though not in so many words: see

11 *Pace* Martin Davies, (forthcoming), who has argued that *a priori* considerations can establish the LOT hypothesis in this stronger sense.

Smolensky 1988). A Connectionist network has a large number of very simple processors which may be individually activated or inhibited in a simple way. Combinations of these simple units can be activated in patterns that function as representations in the network. But token representations of the same semantic type need not have *non*-semantic features in common in the way the LOT hypothesis claims. One pattern of activation may represent *coffee in a cup*, and another represent *coffee beans*, and the two patterns need have nothing interestingly non-semantic in common (Smolensky 1988 pp. 16–17). Representations in a Connectionist network need not, therefore, have a mental syntax in the stronger sense.

Would the causal efficacy of beliefs and desires be undermined by the truth of this sort of Connectionist theory? Does the truth of folk psychology require mental syntax? There are many complex issues surrounding these questions, which I cannot address in this brief discussion. But I do think that on the Functionalist picture outlined—'realism all the way up'—the short answer is: no. For the efficacy of beliefs and desires, on this picture, depends on the truth of their Ramsey Sentence. In section IV, I argued that the Ramsey Sentence for beliefs and desires will involve reference to local causal surrogates for their contents. But that is simply an extra claim about the causal structure of the beliefs and desires *themselves*. The existence of the surrogates does not eliminate the beliefs and desires from the Ramsey Sentence, and it does not undermine their efficacy. It is just an extra fact about them.

We do not yet know whether the structure of the surrogates—the mechanism of thought—is syntactic or Connectionist, or a bit of both. This is one of the most important questions addressed by contemporary cognitive science. But the present point is that nothing in the idea of these surrogates rules out their having a Connectionist structure. So nothing in what we now know of the Ramsey Sentence for beliefs and desires entails that Connectionism is false.[12]

Department of Philosophy
University College London
Gower Street
London WC1E 6BT

REFERENCES

D. M. Armstrong and Adrian Heathcote, (1991) 'Causes and Laws', *Nous* 25, 63–73.

Jonathan Bennett, (1988) *Events and their Names*, (Cambridge: Cambridge University Press).

Simon Blackburn, (1990) 'Filling in Space', *Analysis* 50, 62–65.

Ned Block, (1990) 'Can the Mind Change the World?', in *Meaning and Method: Essays in Honour of Hilary Putnam*, edited by George Boolos (Cambridge: Cambridge University Press) 137–170.

Patricia Smith Churchland, (1982) Review of: Austen Clark, *Psychological Models and Neural Mechanisms*, *Journal of Philosophy* 79, 98–111.

Tim Crane, (1990) 'The Language of Thought: no Syntax without Semantics', *Mind and Language* 5, 187–212.

Tim Crane, (1991a) 'Why Indeed? Papineau on Supervenience', *Analysis* 51, 32–37.

Tim Crane, (1991b) 'All God Has To Do', *Analysis* 51, 235–244.

Tim Crane and D.H. Mellor, (1990) 'There is no Question of Physicalism', *Mind* 99, 185–206.

Donald Davidson, (1967) 'Causal Relations', reprinted in *Essays on Actions and Events* (Oxford: Oxford University Press 1980) 149–162.

Donald Davidson, (1970) 'Mental Events', reprinted in *Essays on Actions and Events* (Oxford: Oxford University Press 1980) 207–224.

Martin Davies, (forthcoming) 'Concepts, Connectionism and the Language of Thought', in *Philosophy and Connectionist Theory*, edited by D. Rumelhart, W. Ramsey and S. Stich (Hillsdale, NJ: Lawrence Erlbaum Associates).

Fred Dretske, (1988) *Explaining Behavior: Reasons in a World of Causes*, (Cambridge, MA: MIT Press).

Berent Enç, (1983) 'In Defense of the Identity Theory', *Journal of Philosophy* 80, 279–298.

Jerry A. Fodor, (1981) *Representations*, (Hassocks: Harvester Press).

Jerry A. Fodor, (1985) 'Fodor's Guide to Mental Representation', *Mind* 94, 76–100.

Jerry A. Fodor, (1987) 'A Situated Grandmother?', *Mind and Language* 2, 64–81.

Jerry A. Fodor, (1990) 'Making Mind Matter More', in *A Theory of Content and Other Essays*, (Cambridge, MA: MIT Press).

Ted Honderich, (1982) 'The Argument for Anomalous Monism', *Analysis* 42, 59–64.

Frank Jackson and Philip Pettit, (1990) 'Causation in the Philosophy of Mind', *Philosophy and Phenomenological Research* 50, 195–214.

Jaegwon Kim, (1991) 'Dretske on How Reasons Explain Behaviour', in *Dretske and his Critics*, edited by Brian McLaughlin, (Oxford: Blackwell) 52–72.

Ernest Le Pore and Barry Loewer, (1987) 'Mind Matters', *Journal of Philosophy* 84, 630–642.

David Lewis, (1970) 'How to Define Theoretical Terms', *Journal of Philosophy* 67, 427–446.

12 I am very grateful to Jonathan Wolff for useful comments on an earlier draft of this paper.

David Lewis, (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61, 343–377.
William G. Lycan, (1987) *Consciousness*, (Cambridge MA: MIT Press).
Cynthia Macdonald and Graham Macdonald, (1986) 'Mental Causes and the Explanation of Action', *The Philosophical Quarterly* 36, 145–158.
Colin McGinn, (1989) *Mental Content*, (Oxford: Blackwell).
Brian P. McLaughlin, (1989) 'Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical', in *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, edited by James E. Tomberlin, (Atascadero, CA: Ridgeview) 109–135.
Norman Melchert, (1986) 'What's Wrong with Anomalous Monism?', *Journal of Philosophy* 83, 265–274.
D. H. Mellor, (1974) 'In Defense of Dispositions', *Philosophical Review* 83, 157–181.
D. H. Mellor, (1987) 'The Singularly Affecting Facts of Causation', in *Metaphysics and Morality*, edited by Philip Pettit et al., (Oxford: Blackwell) 111–136.
D. H. Mellor, (1991) 'Properties & Predicates', in *Matters of Metaphysics*, (Cambridge: Cambridge University Press) 170–182.
David Papineau, (1990) 'Why Supervenience?', *Analysis* 50, 66–71.
Gabriel Segal and Elliott Sober, (forthcoming) 'The Causal Efficacy of Content', *Philosophical Studies*
Paul Smolensky, (1988) 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences* 11, 1–74.
Robert C. Stalnaker, (1984) *Inquiry*, (Cambridge, MA: MIT Press).
Mark Wilson, (1985) 'What is this Thing Called "Pain"? —The Philosophy of Science Behind the Contemporary Debate', *Pacific Philosophical Quarterly* 66, 227–267.