

## CAUSAL RELEVANCE AND THOUGHT CONTENT

BY KIRK A. LUDWIG

### I

It is natural to think that our ordinary practices in giving explanations for our actions, for what we do, commit us to claiming that content properties are causally relevant to physical events such as the movements of our limbs and bodies, and events which these in turn cause. If you want to know why my body ambulates across the street, or why my arm went up before I set out, we suppose I have given you an answer when I say that I wanted to greet a friend on the other side of the street, and thought that my arm's going up would be interpreted by him as a signal to stop for a moment. This widely held view<sup>1</sup> might be disputed, but I shall not argue for it in this paper. I want to start with the view that our beliefs and desires and other propositional attitudes are causally relevant, in virtue of their modes and particular contents, to our movements, in order to investigate the consequences for analyses of thought content. For this purpose, I argue, in §II, for three necessary conditions on causal relevance: (a) a nomic sufficiency condition, (b) a logical independence condition, and (c) a screening-off condition. In §III, I apply these conditions to relational and functional theories of thought content, arguing that these theories cannot accommodate the causal relevance of content properties to our behaviour. I argue further that, on two plausible assumptions, one about the dependence of the mental on the physical, and the other about the availability in principle of causal explanations of our movements in terms of our non-relational physical properties, content properties can be causally relevant only if they are nomically type-correlated, relative to certain circumstances, with non-relational physical properties of our bodies. In §IV, I respond to a number of objections that might be made to my argument.

<sup>1</sup> See, e.g., Sosa; LePore and Loewer; Fodor 1989; Kim; Antony. In all of these discussions the worry is that the mental will turn out to be causally irrelevant to our movements and what these cause.

By a content property, I mean any property attributable in a sentence of the form 'a has F' in which 'has F' is replaced by a psychological attitude verb followed by a clause which gives its content. A simple example is the property attributed in the sentence 'Richard Burton believed that the source of the Nile lay in the Mountains of the Moon', which attributes the content property of believing that the source of the Nile lies in the Mountains of the Moon to Richard Burton, where the grammatical object 'the source of the Nile lies in the Mountains of the Moon' gives the content. Likewise, 'Albert intends to win the Tour de France' attributes a content property, the property of intending to win the Tour de France, where the indirect object 'win the Tour de France' gives the content of the intention. In contrast, the sentence 'Mary wants something' does not attribute a content property in my sense, because the grammatical object of the attitude verb does not specify a particular content, that is, it does not tell us what Mary wants. The others do tell us what Burton believed and what Albert intends.

## II

### (a) *Nomic sufficiency*

I treat the relation of causal relevance as a relation between properties, or, alternatively, between event types, relativized to background conditions. Thus causal relevance is a three-place relation between features of events, or, we could say, between event types, and certain background conditions. In the following, I talk about event types and properties indifferently. Events I treat as changes, which can be thought of as an object's or spatial region's coming to have a certain property. The type of an event is determined by what property an object or spatial region comes to have when a change in it instantiates the event type. We can thus convert talk of event types into talk of properties, and convert talk of properties into talk of event types, by talking about the event of something's coming to have that property.

Causal relevance must be distinguished from causation. Causation is a relation that holds between particulars, no matter how described, while causal relevance holds between properties or event types. If two events are causally related, then there are properties of them that stand in the causal relevance relation. But for the two events to be described as causally related, they need not be picked out using descriptions which indicate a pair of properties that stand in the causal relevance relation.

As I am thinking of it, causal relevance is closely tied to causal explanation. To give a causal explanation of the occurrence of an event of a certain type, it is not sufficient to cite another event which caused it. Rather one must cite the latter event under some description which explains why it caused the former. If you want to know why Kennedy died on 22 November 1963, it is no answer to be told that it was because of an event occurring at 12.30 p.m. on that day at the far right corner of the third floor of the Texas School Book Depository facing Dealey Plaza. Some feature of that event must be cited which is causally relevant to Kennedy's death.

Causal relevance is not the same as causal explanation. While every causal explanation must cite some factor causally relevant to the occurrence of the *explanandum* event type, not every causally relevant factor causally explains the occurrence of an *explanandum* event type. Explanations are offered in response to requests for information, and thus are interest-relative. What I already know, or what is assumed relative to an explanatory request about the causal conditions for the occurrence of an event of a certain type, cannot be cited in explanation of it. For example, it is not usually to the point to be told about the presence of oxygen upon requesting an explanation of a fire.

Causal relevance is what is left over when we strip away the interest-relativity of causal explanation. Since in causal explanations we are interested not just in a cause of an event, but in what feature of the salient cause of an event was causally responsible for an effect of the type we are interested in, we want a feature of the cause which explains what it was in virtue of which the first event caused the second. Causal relations hold in virtue of the events which stand in those relations being subsumed by causal laws. Thus we can cite a feature F of an event as a feature in virtue of which it is a cause of another event of a certain type G, provided that the events fall under a causal law connecting events of type F and G in circumstances of the type in which the one event caused the other (I use capital letters throughout for types or properties, and lower case letters for particulars). This, I suggest, gives us both a necessary and sufficient condition for one event type to be causally relevant to another:<sup>2</sup>

[1] Event type  $E_1$  is causally relevant to event type  $E_2$  relative to circumstances of type C iff it is a causal law that the occurrence of

<sup>2</sup> This view is shared by Fodor 1989.

an event of type  $E_1$  is followed by the occurrence of an event of type  $E_2$  in circumstances of type  $C$ .

The interest-relativity of causal explanations enters in our selection of a salient causal factor, and so in the selection or indication of a causal law relevant to our interests.

While I believe that [1] states a necessary and sufficient condition for two event types to stand in the relation of causal relevance, in the absence of an analysis of the notion of a causal law it is not very informative. We do not know how to apply this to analyses of thought content unless we have an account of the conditions under which a nomic connection is a causal, as opposed to non-causal, nomic connection. I do not have a solution to the difficult problem of analysing the notion of a causal law. Instead, I shall argue for three necessary conditions on causal relevance, which will still allow us to draw some important conclusions about theories of thought content, given our initial assumption that content properties are causally relevant to our movements.

The first necessary condition follows from the requirement that types causally relevant to one another be related by a causal law: if events  $e_1$  and  $e_2$  stand in the causal relation in virtue of being subsumed by a causal law relating event types  $E_1$  and  $E_2$  in circumstances  $C$ , then a necessary condition for  $E_1$  being causally relevant to  $E_2$  in  $C$  is that  $E_1$  be nomically sufficient for  $E_2$  in  $C$ . Canonically:

[2] An event type  $E_1$  is causally relevant to an event type  $E_2$  in circumstances of type  $C$  only if it is nomically necessary that an event of type  $E_2$  follow one of type  $E_1$  in circumstances of type  $C$ .

For convenience, I represent the antecedent of this as ' $CR(E_1, E_2, C)$ ', read as 'Event type  $E_1$  is causally relevant to event type  $E_2$  in circumstances of type  $C$ '. Let me represent the consequent as ' $L(E_1, E_2, C)$ ', read as 'It is a law that an event of type  $E_2$  follows one of type  $E_1$  in circumstances of type  $C$ '. We require that the predicates which characterize permissible types be purely qualitative, that is, be such that their analyses do not contain singular referring terms. That  $L(E_1, E_2, C)$  holds, of course, is not sufficient for it being the case that  $CR(E_1, E_2, C)$  because the nomic sufficiency of  $E_1$  for  $E_2$  in  $C$  may be explained by conditions in  $C$  being sufficient for both.

(b) *Logical independence*

We arrive at a second necessary condition on causal relevance by noting (the Humean insight) that causal relations are contingent. By this I mean that they do not hold in virtue of logical or conceptual connections between properties of the events that stand in the causal relation. Since two events stand in the causal relation in virtue of their falling under event types which stand in the causal relevance relation, it must be a condition on the relation of causal relevance between two properties or event types that they are not logically related. This means at least that given one property P and another Q, P cannot be causally relevant to Q if that  $x$  has P entails, given the circumstances, that there is some  $y$  such that  $y$  has Q, or *vice versa*. This rules out, e.g., the property pair {being P, being the cause of P} from being in the causal relevance relation, as well as the pair {being P, being the effect of P}. Being exposed to the sun can be causally relevant to getting burned, but not to sunburn, to which the relation is logical, not causal.<sup>3</sup>

However, we need a still stronger logical independence relation than this. For consider the following two cases. First, consider the property of being caused by an event involving a coloured object. Call this being colour-caused. Thus a photograph of someone wearing a bright red shirt is colour-caused. We would say that being coloured is not causally relevant to being colour-caused because the relevance of the first to the second is logical, not causal. That something is colour-caused entails that some event involving a coloured object occurred. But now consider the property of being red. This is no more causally relevant to being colour-caused than being coloured is, but in this case that an event that was colour-caused occurred does not entail that an event involving a red object occurred. Thus, simple logical independence is too weak a criterion to rule out being an event involving a red object from standing in the causal relevance relation to being colour-caused. Here is a second case in which we need to say the same thing. Consider the property of being a colour-cause, i.e., the cause of an event involving a coloured object. For example, a cause of a colour photograph is a colour-cause. This property will not be causally relevant to the property of being an event

<sup>3</sup> See Dardis for a fuller and slightly different discussion of this kind of condition on causal relevance. Similar considerations are advanced, though for different purposes, in Fodor 1991: see esp. pp. 10–11, 19. My thinking on causal and logical connections here was originally prompted by Davidson's 1963 discussion, though, as is made clear below, I think there is more to the logical connection objection than Davidson realized.

involving a coloured object, because its occurrence logically requires the occurrence of such an event. However, consider the property of being an event involving a red object. Being a colour-cause does not require the occurrence of an event involving a red object. Yet we do not want to say that being a colour-cause could be causally relevant to the occurrence of an event involving a red object, for the occurrence of such an event is logically sufficient for the occurrence of an event logically required by the occurrence of a colour-cause. So, again, we need to strengthen the logical independence requirement. These two cases illustrate the following principle:

[3] If event type  $E_1$  logically requires event type  $E_2$ , in circumstances of type  $C$ , then, in circumstances of type  $C$ , event type  $E_1$  is not causally relevant to any event type  $E_2+$  such that an event's falling under that event type is logically sufficient for it to be of type  $E_2$ .

Our problem is to strengthen the logical independence condition so that it has this as a consequence.

We can construct a condition which has the appropriate strength in two stages. First, we define 'weakly logically independent'.

[4] An event type  $E_1$  is weakly logically independent of an event type  $E_2$  in circumstances  $C$  iff it is possible for an event of type  $E_1$  to occur in  $C$  without an event of type  $E_2$  occurring and it is possible for an event of type  $E_2$  to occur in  $C$  without an event of type  $E_1$  occurring.

Second, we define 'strongly logically independent' as follows:

[5] An event type  $E_1$  is strongly logically independent of an event of type  $E_2$  in circumstances of type  $C$  iff event type  $E_1$  is weakly logically independent of event type  $E_2$  in  $C$  and every event type which  $E_1$  is not weakly logically independent of in  $C$  is weakly logically independent of  $E_2$  in  $C$  and every event type which  $E_2$  is not weakly logically independent of in  $C$  is weakly logically independent of  $E_1$  in  $C$ .

I abbreviate 'event type  $E_1$  is strongly logically independent of event type  $E_2$  in circumstances of type  $C$ ' as 'SLI( $E_1, E_2, C$ )'. Then our second necessary condition on causal relevance can be stated as follows:

[6]  $CR(E_1, E_2, C)$  only if  $SLI(E_1, E_2, C)$

Now we can see that [6] entails [3], as required.

(c) *Screening off*<sup>4</sup>

The third necessary condition on causal relevance will be that an event type not be, in a certain sense to be explained below, *screened off* from the effect type in the circumstances. This condition is designed to rule out of the causal relevance relation such pairs of event types as a fall in mercury level and the onset of a storm.

We often want to know, in vaguely specified conditions, whether a given event type is causally relevant to another. It may appear to be, because, relative to certain conditions, whenever an instance of the one event type occurs one of the other does as well, as in the case of a fall in the mercury level and the occurrence of a storm, and this remains so while other factors vary. But we may suspect that it is not that event type, but another event type, specified in the background conditions as occurring, which is what is really causally relevant to the effect type of interest. If this were the case, we could say that the second event type screened off the first from the effect type.

How in practice do we test for whether an event is screened off from some effect type of interest? There are, I think, two tests which we apply. To see how this works, let us take an example.<sup>5</sup> Suppose that whenever a soprano gets to a certain point in a song, if there is a glass on the table next to her, the glass breaks. Let us suppose that, keeping certain conditions fixed, we vary other things and discover that, in all those situations, the glass breaks. Someone suggests that the meaning of the word which the soprano sings at that point, a factor we had kept fixed while varying other factors, is what is causally responsible, against the other background conditions, for the glass's breaking. How would we show that it was not, after all, causally relevant to the glass's breaking? The first sort of test we employ is to try to show that, among the background conditions  $C$ , in addition to the

<sup>4</sup> The term 'screening off' was originally used, I believe, by Reichenbach. He and also Salmon formulate a probabilistic condition, LePore and Loewer a condition in terms of counterfactuals, for the causal relevance of one particular to another. My condition is articulated in terms of strict laws, and is for types rather than particulars. I think causation is to be understood in terms of strict necessitation.

<sup>5</sup> I borrow the example from Dretske 1988. I argue that Dretske's account of how reasons explain behaviour is false, and cannot solve the problem it sets out to, in 'Dretske on Explaining Behavior', forthcoming in *Acta Analytica*.

note's having a meaning that we have kept fixed, there are other factors, such as the pitch and volume of the note, which even in the absence of the note's having a meaning are sufficient for the glass's breaking. This would show that the meaning was not necessary, relative to the rest of the conditions we hold fixed. Call this test 1:

Holding other conditions in C fixed, see whether the soprano's singing a note without singing a note with the meaning of the word she actually sings is sufficient for the glass to break.

A second test we employ is to try to show that singing a note with that meaning was not by itself sufficient, without some other factors also being present in the conditions we had so far held fixed, again e.g., the note's having a certain volume and pitch. Call this test 2:

Varying other conditions in C, see whether the soprano's singing a note with that meaning fails to be sufficient for the glass's breaking.

While we often make decisions about causal relevance after employing either of these tests, neither is by itself sufficient to show that the meaning of the word sung is not causally relevant to the glass's breaking. The first test is not sufficient to show failure of causal relevance, because of the possibility that the effect type is overdetermined in the circumstances. Consider someone who swallows a dose of cyanide, in conditions in which he also swallows a dose of strychnine. His death would have occurred without his swallowing cyanide, because swallowing the strychnine was independently sufficient, and so swallowing cyanide was not necessary. But that does not mean that it was not causally relevant, because if we evaluate whether swallowing a dose of strychnine is necessary, relative to conditions which include the swallowing of cyanide, it is not necessary for the swallower's death either. Neither is necessary, but both are independently sufficient. That the occurrence of an event of a given type is not necessary we take as evidence that it is not sufficient, because we assume that typically we are not faced with a case of causal overdetermination. The second test is not sufficient to show that the meaning is causally irrelevant either, because, while the meaning might not be sufficient for the glass's breaking, it might be a necessary part of a sufficient condition. Thus, e.g., singing a high C may not be sufficient for a glass's breaking, but it may be a necessary part of a sufficient condition which includes singing sufficiently loudly.

Notice, however, that each test can be used to show that the condition in which the other test fails does not obtain. Thus if an event type fails test 2, it is not sufficient for the effect type, and so cannot



causally overdetermine it. If an event type fails test 1, it is not necessary, and so it cannot be a necessary part of a sufficient condition. Thus we have a sufficient condition for one event type's being screened off from another if we know that it fails both tests, that is, if we know that it is neither necessary nor sufficient, relative to the background conditions. Call this test 3:

See whether the soprano's singing a note with that meaning fails both tests 1 and 2, i.e., is neither necessary nor sufficient relative to the background conditions.

It suffices to show that the occurrence of an event of a given type is neither necessary nor sufficient for the occurrence of a subsequent event of a given type, if the occurrence of an event falling under another type in the circumstances is sufficient without the occurrence of an event of the type we are evaluating, and the occurrence of an event of the type we are evaluating is not independently sufficient. For convenience, let me write 'the occurrence of an event of type X' as 'X'. Then, if M is a note's being sung with a certain meaning, C is the set of background conditions we keep fixed, and B is the glass's breaking, there is an event type E which screens off M from B relative to C, if E is included in C and the conditions C without E but with M are not sufficient, but with E are sufficient independently of M. There is a plausible candidate in this case, namely, the pitch and volume of the note the soprano sings.

Now let us state our condition canonically.

[7] An event type  $E_1$  in conditions C is screened off from an effect type  $E_3$  by another event type  $E_2$  iff  $E_2$  is included in C and there are conditions C minus  $E_2$ ,  $C^-$ , which together with  $E_1$  are not nomically sufficient for  $E_3$ , and  $E_2$  together with  $C^-$  is nomically sufficient for  $E_3$ .

By 'there are conditions C minus X' we mean conditions which include everything C does except X and whatever nomically or logically requires X or is required nomically or logically only by X. We can abbreviate 'E<sub>1</sub> is screened off by E<sub>2</sub> from E<sub>3</sub> in conditions C' as 'S(E<sub>1</sub>, E<sub>2</sub>, E<sub>3</sub>, C)'. This then gives us the following three necessary conditions on causal relevance:

- [8] CR(E<sub>1</sub>, E<sub>3</sub>, C) only if
- (i) L(E<sub>1</sub>, E<sub>3</sub>, C)
  - (ii) SLI(E<sub>1</sub>, E<sub>3</sub>, C)
  - (iii) there is no event type E<sub>2</sub> such that S(E<sub>1</sub>, E<sub>2</sub>, E<sub>3</sub>, C).

## III

I turn now to an application of these conditions to the causal relevance of thought content to our behaviour. I am assuming that our ordinary practices in explaining behaviour commit us to content properties' being causally relevant to such things as the movement of my hand as I gesture, and the motion of my feet when I walk across the floor. For example, we can cite my desire that my hand move as part of the cause of (what causally explains) its moving, and my desire to reach the door as a part of the cause of (what causally explains) the movement of my feet. I now argue that content properties cannot enter into the sorts of causal relevance relations our ordinary practices commit us to positing if either functionalism or externalism about thought content is correct. I begin with functionalism.

(a) *Functionalism*

The argument against functionalism rests on the logical independence requirement for causal relevance. A functionalist theory of content characterizes contentful states as states which enter into certain causal relations with other states and with input and output. To illustrate the difficulty, let us take a simple functional state, the state S that a system is in, such that, when it is in S and receives input I, it produces output O. Suppose that a system comes to be in state S. Later it receives input I. Can the event of coming to be in state S be causally relevant to O's occurrence? Let us include in the background conditions that input I was received; this is necessary if S is to be sufficient for O. Given the definition of S, it is not logically possible that it occur and, relative to these background conditions, O not occur. Thus, the event of S's occurring, relative to those background conditions, is not logically independent of O. It therefore cannot stand in the causal relevance relation to O.

While we have no workable explicit functional characterizations of mental states, we know enough about how the story is supposed to go to see that functionalism will violate the logical independence condition. Its two main varieties are causal role functionalism, and machine table functionalism.<sup>6</sup> I consider each in turn.

<sup>6</sup> The *locus classicus* for machine table functionalism is Putnam 1967, and for causal role functionalism is Lewis 1972.

Causal role functionalism characterizes attitudes as states which play certain causal roles with respect to other states, input, and output. Thus that I desire to move my finger requires that I be in a state such that, when such-and-such circumstances obtain, my finger moves. The difficulty is immediate. Since I am not in that state unless it causes finger movement in those circumstances, finger movement is not logically independent of having a desire to move my finger in those circumstances. Therefore having that desire cannot be causally relevant to my finger's moving, if causal role functionalism is correct. Since causal role functionalism will be plausible only if the output in terms of which states are characterized is what we ordinarily use them to explain, this will mean that if causal role functionalism is correct, our reasons are not causally relevant to that behaviour to which we ordinarily take them to be causally relevant.

According to machine table functionalism, having a mind is having a certain functional organization, which can be spelled out in terms of states characterized by sets of conditionals relating input and present state with future states and output. So a given machine table state is characterized by a set of conditionals relating input to output and future state. To be in a given state S would be to be in a state such that, when input I is received, the system produces output O, and changes to state S\* (which may be identical with S), and so on. It would not be plausible to represent a given attitude as a state in the machine table, since presumably what behaviour I display in response to a given input is a result of a large number of attitudes. Thus having a certain desire, say, the desire to move my finger, would be characterized in terms of a disjunction of all of the machine table states in which the input-output relations would be appropriate for me to have that desire. Suppose, then, that being in state S is sufficient for desiring to move my finger, and suppose that upon input I, the appropriate output is that my finger moves. Is being in S causally relevant to my finger moving? No, because in order for S to be causally relevant there must be conditions relative to which it is sufficient to bring about my finger's moving, but those conditions will have to include receiving input I. Relative to those conditions, however, that my finger moves is not a causal but a logical requirement of being in state S. If my finger does not move, then I am not, after all, in state S. Thus machine table states are not causally relevant to the output in terms of which they are defined. What about the desire to move my finger? For this is not identical with the state S. The desire cannot be causally relevant either, because it also fails the logical independence

requirement [5] formulated in the previous section, for being a desire to move my finger is not weakly logically independent of being in state *S*, and so is not strongly logically independent of my finger's moving. Thus, if machine table functionalism is correct, content properties are not causally relevant to the output in terms of which they are defined. Since this output will presumably be those movements which we regularly call upon reasons to explain, this will mean that machine table functionalism will be unable to accommodate the causal relevance of reasons to that to which we take them to be causally relevant.

Thus both causal role functionalism and machine table functionalism make content properties causally irrelevant to those ordinary physical events which we take content properties to be able to explain causally. This constitutes, I think, a sufficient reason to reject both of them, at least as conceptual accounts of the nature of thought content.

It is an interesting additional consequence of the logical independence condition on causal relevance that syntactical states of mind of the sort appealed to by Language of Thought theorists, if individuated functionally, fail to be causally relevant to behaviour. Thus the image of the mind as essentially a syntactic engine, driven not by the content properties of our states, but by their syntax, turns out to rest on a confusion.

### (b) *Externalism*

I now turn to a criticism of externalist theories of thought content. I shall call a theory of content 'externalist' if it holds that having a propositional attitude with a certain content is a relational property of the individual who has it. *P* is a relational property iff, necessarily, *x* has *P* only if there is a *y* such that *y* is not identical with *x* or with any part of *x*, and *y* is not a necessary existent. My claim is that, if externalism is true, then content properties are not causally relevant to most of the event types to which we are committed to taking them as causally relevant. The argument relies on the following two assumptions:

1. Whenever my body moves there is some non-relational physical property of me, *P*, which relative to conditions present, *C<sub>p</sub>*, is necessary and sufficient for that bodily movement.

2. For any non-relational property  $P$  of a body, and background conditions  $C_p$  relative to which  $P$  is necessary and sufficient for some bodily movement, the properties that any plausible externalist theory will select as those that determine thought contents will be nomically and logically independent of  $P$  and  $C_p$ .

To see why these assumptions, together with our conditions on causal relevance, entail that relational theories cannot make content properties causally relevant to the movements of our bodies, let us suppose that I want to move my hand, and that  $R$  is the relational property sufficient for an object to have the desire to move its hand. Let  $M$  be the event type of my hand's moving, to which we want my desire to move my hand to turn out to be causally relevant. Suppose that  $R$  is sufficient for  $M$  relative to circumstances  $C_r$ . (I continue to use the convention that writing 'X' is to be understood as 'the occurrence of an event of type X' or 'the instantiation of X'.) The argument can be laid out as follows:

- (i) By hypothesis,  $R$  relative to  $C_r$  is sufficient for  $M$ .
- (ii) By 1 and (i), when  $R$  and  $C_r$  obtain, there is some non-relational physical property of me,  $P$ , which is necessary and sufficient for  $M$ , relative to certain background conditions present,  $C_p$ .
- (iii) Therefore, by (i) and (ii),  $R$  and  $C_r$  are sufficient for  $P$  and  $C_p$ .
- (iv) By 2,  $R$  is nomically and logically independent of  $P$  and  $C_p$ .
- (v) Therefore, by (iii) and (iv),  $C_r$  is sufficient for  $P$  and  $C_p$ .
- (vi) From (ii) and (v), there are conditions less inclusive than  $C_r$ , which do not include  $P$ ,  $C_{r,p}$ , which are such that those conditions together with  $P$  are sufficient for  $M$  without  $R$ , and relative to which  $R$  is not sufficient for  $M$ .
- (vii) From (vi) and [7] (the definition of screening off) it follows that  $R$  is screened off from  $M$  by  $P$  relative to  $C_r$ .
- (viii) From (vii) and [8] (the necessary conditions on one property's being causally relevant to another) it follows that  $R$  is causally irrelevant to  $M$ .

The conclusion of the argument is perfectly general, since we used no facts about the particular bodily movement type or content property we used in illustration.

Both assumptions 1 and 2 seem to me to be overwhelmingly plausible. Given these two assumptions, and our characterization of causal relevance, it follows that externalist theories of thought content cannot accommodate the causal relevance of content properties to the movements of our bodies, and what these in turn cause. If, as I think, our ordinary practices of explanation commit us to treating content properties as causally relevant to our behaviour, we should reject externalist theories of thought content.

(c) *Relations to causally relevant properties*

What can we say more generally about the relation of content properties to the physical properties of our bodies, if content properties are to be causally relevant to our movements? In general, for content properties to be causally relevant to movements of our bodies, they must be sufficient for these, relative to the circumstances in which they are produced. Since we know that some physical properties of our bodies are sufficient and necessary for our bodies' movements, it must be the case that content properties 'line up' with these properties. Let us assume that our content properties are causally relevant to the movements of our bodies. Then there are circumstances C relative to which they are sufficient for those movements. But we also know that there is a physical story to be told about those movements, relative to those same background conditions. Let us assume materialism is true,<sup>7</sup> and that mental states are states of our bodies. Suppose D is my desire to move my hand, and M is my hand's moving. Suppose P is the physical property necessary and sufficient for M relative to C. Then D will be sufficient for M relative to C only if it is sufficient for P. However, it is implausible to suppose that P is dependent on D. We think instead that mental properties and states should be determined in some sense (weaker than entailment) by physical properties and states. So if D is sufficient for P, it must be that, relative to the circumstances C, P is the only physical property which is sufficient for D. Thus, relative to the circumstances, if D is causally relevant to M, it must be nomically correlated with P.

Thus, the requirement that content properties be causally relevant to the movements of our bodies leads us to a kind of relativized type-type identity theory of content properties with non-relational physical

<sup>7</sup> In the sense articulated by Hellman and Thompson: every object is composed of physical parts except for basic physical constituents.

properties of our bodies. By a type–type identity theory of properties F and G relative to conditions C, I mean only that it is nomically necessary that, given C, for all objects  $x$ ,  $x$  is F iff  $x$  is G. Note, however, that this is a weak version of type–type identity, for it is only relative to the conditions in which the physical properties are causally sufficient for the movements of our bodies to which our content properties are causally relevant. If other physical properties, in other conditions, could be causally necessary and sufficient for those movements, then all our result would tell us is that, if those movements could be causally explained by reasons in those conditions, the content properties must be, relative to those conditions in turn, nomically type-correlated with those physical properties. So whether we get more general correlations between physical and mental types cannot be deduced from general considerations about causal relevance.

None the less it seems to me quite plausible, once we have determined that content properties supervene on non-relational physical properties, and that functionalism is false, that they are the result only of certain quite specific physical properties. But I confess that my only reason for thinking this is that the alternative would be a much messier picture, and one for which, once we have abandoned externalism and functionalism, there would seem to be no explanation. In the end, whether there are universal type–type correlations between our mental and physical states must remain an empirical question, even if we grant that our mental states are causally relevant to the movements of our bodies.

#### IV

In this section, I consider a number of objections to the preceding argument.

(i) In response to my criticism of functionalism, it might be said that the criticism works only against versions of functionalism which require that functional states be defined in terms of strict causal necessitation between being in a certain state, receiving an input, and producing an output. But functional states can be characterized also in terms of probabilistic transitions. For all such functionalist theories, the criticism above fails, because the functionally defined states will be strongly logically independent in the circumstances of the output in terms of which they are defined.

*Response.* The objection is right that my conditions on logical independence will not rule out functionally defined properties from being causally relevant to bodily movements, if they are defined in terms of probabilities (less than one) of various outputs given a certain input. But the conditions I have given require nomic necessitation relative to the circumstances, and for functional states defined in terms of output probabilities less than one, this condition will not be met. It might be replied that nomic necessitation is too strong a requirement on causal relevance, and that we can make sense of causal relevance also when we have only probabilistic relations among event types. This I think is so only if those relations are underlain by strict laws. But I will not insist on this. If we allow nomic probabilistic relations to ground relations of causal relevance, we shall also require conditions parallel to the strong logical independence condition above, for reasons similar to those given above, which will preclude functional states defined in terms of output probabilities less than one from standing in the relation of causal relevance to the output in terms of which they are defined.

(ii) It might be objected that some functionalist positions can escape the argument in the previous section by defining functional states in terms not of bodily input or output, but instead (a) in terms of *neurophysiological* input and output conditions, or (b) in terms of *causes and effects* of bodily movements. In case (a) it would be argued that, even if functional states so defined could not be causally relevant to the neurophysiological output conditions in terms of which they are defined, they could be causally relevant to features of what those outputs in turn cause, and so they could be causally relevant to behaviour non-intentionally described. In case (b) it would be argued that, if functional properties are defined in terms of causes and effects of bodily movement, that is, input and output characterized as distal events, the functional properties will not be logically relevant to *behaviour*, and so can be causally relevant to it.

*Response.* I shall attempt to meet each of these objections directly, but it should be remarked first that for theories which hold that propositional attitudes can be analysed functionally, it seems unlikely that any plausible account can be given in terms of neurophysiological input and output conditions, for one can possess the concepts of belief, desire, and so on, without possessing the concepts needed for describing any neurophysiological input or output conditions. Furthermore, option (b) is not much help, even if it could secure the



causal relevance of functional states to behaviour, since we think content properties are also causally relevant to the typical effects of behaviour, those states of affairs we desire and which our actions often bring about, and which are the only plausible candidates for distal output in terms of which to characterize content functionally. But there are further difficulties for each suggestion.

First, objection (a) requires a property of an event  $c$ , which is not causally relevant to *any* property of an event  $e$  which  $c$  causes, to be causally relevant to an effect  $f$  of  $e$ . But for a property of an event  $c$  to be causally relevant to a property of another event  $f$ , when  $c$  is not the proximate cause of  $f$ ,  $c$ 's causal influence must be transmitted through a causal chain to  $f$ . But if that property is not causally relevant to any property of some event in the causal chain leading up to  $f$ , then its causal influence terminates at that link, and it cannot be causally relevant to any property of  $f$ . It could be causally relevant to some property of  $f$  *only by* being a property of its proximate cause and being causally relevant to some property of  $e$  or by being causally relevant to some property of the proximate cause  $d$  of  $f$ . Then in turn it would have to be either a property of the proximate cause of  $d$  and be causally relevant to some property of  $d$ , or be causally relevant to some property of the proximate cause of  $d$ , and so on. Thus if a property of an event is not causally relevant to any property of an effect of that event, it is not causally relevant to any of that event's effects in turn.

Second, in the case of objection (b), if the functional properties are defined in terms of causes and effects of bodily movements, they will be irrelevant to the specific mechanisms that produce those effects. The functional property will then be *neither* logically *nor* causally sufficient or necessary for any particular intervening mechanism, and so neither logically nor causally relevant to it.

(iii) It might be objected to my criticism of functionalism that functional states are clearly causally relevant to the sort of effect types they are introduced in terms of because their method of introduction guarantees it. Thus, we might introduce a term 'M' to denote or express the property which is *whatever* property of our central nervous systems is causally responsible for the production of a certain bodily movement in certain circumstances. The property  $M$  then, by definition, is causally relevant to the movement type in terms of which it is introduced.

*Response.* This objection confuses functional properties with functional descriptions of properties. I have been objecting to the view that

mental properties are functional properties, not to the view that they are introduced using functional descriptions, though I believe this latter view also to be false. To see the distinction, consider the following two ways of introducing a predicate into our language:

[9]  $(x)(x$  is F iff  $x$  causes R in circumstances C)

[10]  $(x)(x$  is F iff  $x$  has whatever property is causally responsible for the production of R in circumstances C)

If we introduce 'is F' by [9], then the property which it expresses is the same as the property which 'causes R in circumstances C' expresses. In contrast, if we introduce 'is F' by [10], then the property expressed by 'is F' is not expressed by the predicate on the right hand side of [10]. Instead, the right hand side of [10] describes what property the expression 'is F' is to express, without expressing it. To discover what property 'is F' expresses, we have to discover what property is causally responsible for the production of R in circumstances C. That property will not itself be a property which anything has (logically) in virtue of causing or being causally responsible for anything, and so will not be a functional property. In illustration, we can imagine that we introduce a term, call it 'water', in the following way:

[11]  $(x)(x$  is water iff  $x$  has whatever microstructural property is responsible for such and such phenomenal features of these samples of liquid)

Upon empirical investigation, we discover that being  $H_2O$  is the property causally responsible for the cited phenomenal features. The predicate 'is water' then expresses the property of being  $H_2O$ , which is clearly not a functional property (or at least not in virtue of the way in which we fix it as the property expressed by 'is water' here).<sup>8</sup>

(iv) Finally, it might be objected that externalism does not fail the screening off condition with respect to actions. For no bodily movement *is* an action unless it has the right aetiology, that is, is caused by reasons that rationalize it. This will secure for us that, no matter what our account of thought content, there cannot be any actions that do not have reasons in their aetiology that rationalize them. Thus with

<sup>8</sup> It might still be said that mental terms are introduced or controlled by such descriptions. I think this is false, but I do not have anything to add to extant arguments against this position. For a good recent treatment of this issue, see Bealer.

respect to action types, e.g., throwing a ball, certain reason properties cannot be shown to be screened off because that they are present in its aetiology is a necessary condition for the instantiation of that action type.

*Response.* The first thing to note is that this does not contradict my conclusion, which is that if externalism is true, reason properties are not causally relevant to our bodily movements described as such. The second point to note is that while this may secure for us the *relevance* of reasons to actions described as actions, it does not secure their causal relevance, which is what we are interested in. Rationalizing reasons required by a given action type are not causally relevant to actions *qua* actions because the relevant reason properties and action type fail the logical independence condition on causal relevance.<sup>9</sup> We want reasons to be causally relevant to behaviour non-intentionally described, not just to my moving my hand, but to my hand's movement, and to my moving my hand in so far as that is my hand's movement. This is all reasons could be *causally* relevant to.<sup>10</sup>

*University of Florida, Gainesville*

#### REFERENCES

- Antony, L. 1991: 'The Causal Relevance of the Mental', *Mind & Language* 6, pp. 295–327.  
 Bealer, G. 1994: 'Mental Properties', *Journal of Philosophy* 91, pp. 185–208.  
 Block, N. (ed.) 1980: *Readings in the Philosophy of Psychology* (Harvard UP).  
 Dardis, A. 1993: 'Sunburn: Independence Conditions on Causal Relevance', *Philosophy and Phenomenological Research* 53, pp. 577–98.  
 Davidson, D. 1963: 'Actions, Reasons and Causes', *Journal of Philosophy* 60; reprinted in Davidson 1984.  
 ——— 1984: *Essays on Actions and Events* (Oxford UP).  
 Dretske, F. 1988: *Explaining Behavior* (Cambridge, Mass.: MIT Press).  
 Fodor, J. 1989: 'Making Mind Matter More', *Philosophical Topics* 17, pp. 59–79.  
 ——— 1991: 'A Modal Argument for Narrow Content', *Journal of Philosophy* 88, pp. 5–26.  
 Hellman, F. and Thompson, G. 1977: 'Physicalist Materialism', *Nous* 11, pp. 309–45.  
 Kim, J. 1990: 'Explanatory Exclusion and the Problem of Mental Causation', in E. Villanueva (ed.), *Information, Semantics and Epistemology* (Oxford: Basil Blackwell), pp. 36–56.  
 LePore, E. and Loewer, B. 1987: 'Mind Matters', *Journal of Philosophy* 84, pp. 630–41.

<sup>9</sup> Dardis also makes this point, at greater length, and with a different emphasis. See also Fodor 1991, though here I think that, although the starting point is roughly right, and the conclusion is right, there are troubles in getting from the one to the other.

<sup>10</sup> I thank the following for helpful comments, criticism and advice: John Biro, Tony Dardis, Brian McLaughlin, Piers Rawling, the participants in the 1993 IUC Conference on Connectionism and the Philosophy of Mind, and an anonymous referee for this journal.

- Lewis, D. 1972: 'Psychophysical and Theoretical Identification', *Australasian Journal of Philosophy* 50, pp. 249–58; reprinted in Block 1980, pp. 207–15.
- Putnam, H. 1967: 'The Nature of Mental States', in W.H. Capitan and D.D. Merrill (eds), *Art, Mind, and Religion* (Univ. of Pittsburgh Press), pp. 37–48; reprinted in Block 1980, pp. 223–31.
- Reichenbach, H. 1956: *The Direction of Time* (Univ. of California Press).
- Salmon, W. 1984: *Scientific Explanation and the Causal Structure of the World* (Princeton UP).
- Sosa, E. 1984: 'Mind–body Interaction and Supervenient Causation', *Midwest Studies in Philosophy* 9, pp. 271–82.