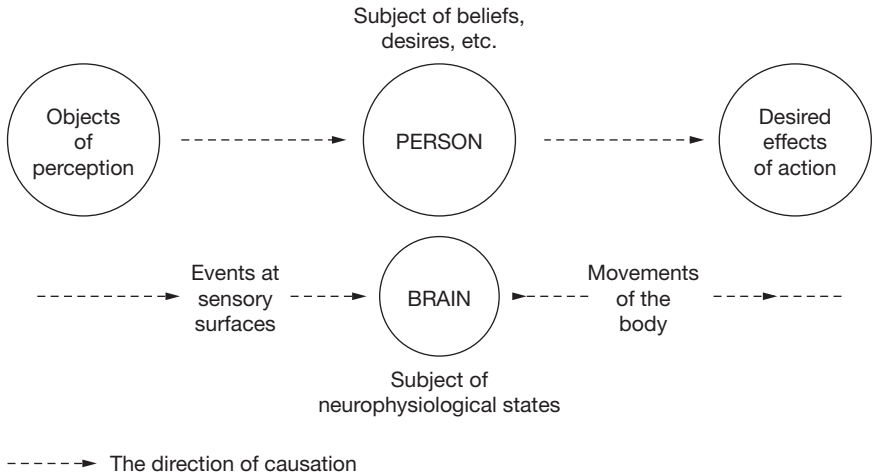


3

J. Hornsby, “Physicalist Thinking and Conceptions of Behaviour”

1. I start from two pictures. One presents a view of what is involved when we ascribe propositional attitudes to one another. The other presents a view of what is involved when the scientist treats a human being as a physical thing – of what a neurophysiologist sees as going on when he concerns himself with the stimulations of sense organs, with the motor responses in a person’s body, and with events and states that intervene between such stimulations and responses.



One has only to look at these two picture to be tempted to make a superimposition. Two considerations may combine to make the temptation irresistible. First, the brain and central nervous system is a part of a person whose proper functioning is a necessary condition of that person’s having the effects on the world she desires to have. Second, the causal chains that lead up to and away from a person’s psychological states apparently pass through the events depicted in the area that circumscribes the neurophysiologist’s study.

If you extend the causal chains of the representation of the brain backwards and forwards, what you reach is the elements standing at the left and right of the representation of the person. The dependence of the person's functioning on the functioning of her brain may make one think of the brain as a mechanism inside the person which is responsible for producing the effects in virtue of which she has her distinctive effects on the world. But then the common properties of the brain's states and of the person's mental states – states of each sort being seen as causal intermediaries – may make one think that in placing the brain inside the person one locates the propositional–attitude states there. Many will therefore feel compelled to say that particular beliefs and desires *are* the neurophysiological states of a person.

This line of thought gives a very quick argument for a version of physicalism. Perhaps no one wishes to acknowledge that he takes such a direct or simple route. But I think that there is a widespread presumption that if beliefs and desires have any place in the physical world, then they are internal states of persons, or of their brains; and I think that this presumption can be created by the sort of high-level comparison of pictures I have just imagined. My project in this paper is to question certain versions of physicalism the quick argument may seem to recommend, by challenging the envisaged superimposition of the two pictures. More particularly, I shall challenge the use to which a certain conception of *behaviour* is put. According to this conception, *behaviour* subsumes both a brain's outputs and a person's outputs, and thus provides an area common to both pictures.

2. Naively we think that we can become informed about people's answers to such questions as 'Why did she keep to the edge of the pond?' or 'Why did she turn on the burner?'. And we suppose that such answers give psychological explanations of behaviour. But it is often said nowadays that any account of psychological explanatory states is bound to use a purely bodily notion of behaviour.

Consider, for example, Kim's claims:¹

[An] action of turning on the burner, insofar as this is thought to involve the burner going on, is not an action that it is the proper business of psychological theory to explain or predict. . . . It is not part of the object of *psychological* explanation to explain why the burner went on. . . . The job of psychological explanation is done once [psychological theory] has explained the bodily action of turning a knob; whether or not this action results in my also turning on the stove, my starting cooking the dinner . . . is dependent on facts quite outside the province of psychology, [which] are not the proper concern of psychological theory.

Kim and others believe, then, that we ought to recognize psychology's proper business to be much narrower than we naively take it to be.

Kim’s claim that psychological states cannot serve to explain (for example) why Kim turned on the burner is rested on the premiss that such states do not serve to explain why the burner went on. Both the premiss and the argument here may be questioned. In order to question the premiss, one must take a relaxed view about psychological explanation. Then it will seem that it can be psychologically explained (for example) why a burner went on. ‘Why did the burner go on? Is the switch faulty?’ ‘No: Jane turned it on, she wanted to make some tea’.

To question the argument, one may take a less relaxed view and start with the assumption that any psychological explanation has as its explanandum why some person did what she did. The principle underlying the argument would then seem to be something like this:

Even if the explanation why p appears to be the fact that z , still if q and r are necessary for p and the fact that z does not explain why q , then the fact that z can only really explain why r .

But such a principle is surely unacceptable. Suppose that we thought that we could explain why the window broke by saying that a heavy stone hit it at speed. We then notice that the window’s breaking required that the window be situated at p and that p be on the stone’s trajectory, and that the stone’s hitting the window at speed does not explain why the window is situated at p . We do not conclude that after all the stone’s hitting the window at high speed cannot really explain why the window broke.²

It is a question how narrow the province of what is psychologically explained would become if one endorsed Kim’s argument wholeheartedly. Kim himself speaks as if *turning a knob*, unlike *turning on the burner*, were an admissible object of psychological explanation. Yet *turning a knob* is surely proscribed for him: it seems no more to be ‘part of the object of psychological explanation’ to explain why a knob turned than it is to explain why the burner went on. And we may wonder whether in fact Kim’s principle does not rule out psychological explanations even of ‘bodily actions’ – of why someone moved her finger, say; for it is by no means obvious that someone’s moving her finger is not ‘dependent on facts which are not the proper concern of psychological theory’. (I return to this at the end of section 3.)

Of course Kim’s conclusion about the objects of psychological explanation may not be meant to rely on the principle alone. It may rely on a prior view of psychological states – as internal states of people which are the immediate causal ancestors of movements of their bodies. This view is certainly held by functionalists. And the functionalists’ conception of behaviour may be supposed to recommend itself on the merits of functionalism. So it will be worth discovering whether the attractions of functionalism can survive scrutiny of the particular notion of behaviour that that doctrine employs.

3. Functionalists think that the defining feature of any type of mental state is given by describing the causal relations that its instances bear (a) to the

environment's effects on a person, (b) to mental states of other types, and (c) to a person's effects on the environment. And they think that mental terms can be simultaneously implicitly defined in a total psychological theory of all the types of mental states. Such a theory contains terms of two sorts, which David Lewis has called the T-terms and the O-terms. The T-terms are, intuitively, mental terms, to be thought of as receiving implicit definition; in a functional theory, their denotations are accorded functional roles that are specified using only the non-mental O-terms. The functionalist thinks of the functional theory (abstracted, as it were, from the psychological theory) as true of, or realized by, the physical states of individuals: physical states occupy the functional roles of mental states.³

Functionalism is to be understood here as a thesis in the philosophy of mind, which treats of those states and events that in the ordinary way we attribute to one another, for example in explaining action. We can ask then 'What does the functionalist have to say about the role of the propositional-attitude states in producing action?'. Put in the functionalist's own terms, this is a question about output generalizations ((c) above), which are meant to give an account of the systematic ways in which such states as beliefs lead, as it is said, to behaviour.

We are told that behaviourism is the ancestor of functionalism, and that functionalism inherits the virtues of behaviourism. But the functionalist's notion of behaviour is very much more restrictive than that which some of the behaviourists employed. When functionalists speak of behaviour, they speak, like Kim, of bodily movements, or else they speak of motor responses.⁴ When Ryle spoke about behaviour, he meant such characterizations of people's actions as these: 'telling oneself and others that the ice is thin, skating warily, shuddering, dwelling in imagination on possible disasters, warning other skaters, keeping to the edge of the pond'.⁵ (It is true that some behaviourists were reductionists, and that they used a narrower conception of behaviour than Ryle. But if one is allowed to think of functionalism as inheriting its attractions from a non-reductionist position, then Ryle's everyday use of 'behaviour' ought not to be legislated into invisibility.)

There are two important differences here between the (Rylean) behaviourist and the functionalist. The behaviourist makes allusion to things beyond the agent's body in his specifications of behaviour, but the functionalist does not. And the behaviourist's behavioural items are actions (that is, events of people doing things such as moving their bodies), whereas the functionalist's behavioural items are apparently not actions, but movements of people's bodies (which are either effects of actions, or proper parts of actions, depending upon your views).⁶ We need to understand why the functionalists should depart from the behaviourists in these two ways and employ the particular conception that they do.

The functionalist's stated objection to behaviourism is familiar enough: the behaviourist said that to believe something (for instance) is to be disposed to certain behaviour, whereas the functionalist insists that belief cannot be

defined in terms of behaviour alone, because allowance has to be made for the simultaneous determination of behaviour by many different mental states. In this point alone, however, there is nothing that evidently constrains one to used a bodily conception of behaviour. And we need to notice something else, which is seldom stated very explicitly by functionalists; the behaviourists’ neglect of the interdependencies between mental things was not in fact the only defect of behaviourism that the functionalist needed to correct for. Certainly, if your belief that it is going to rain is to lead you to take your umbrella, then you need (for example) to want not to get wet and to believe that umbrellas keep the rain off, and to have no other countervailing desires or interfering beliefs. But equally certainly, if your belief that it is going to rain is to lead you to take your umbrella, then you need to believe of something that is your umbrella *that* it is your umbrella. Not only can it not be left out of account what desires a person has (as the behaviourist seemed to suppose), it also cannot be taken for granted that what people believe is true. Avoiding taking this for granted, one might say that someone who believes that p is (very roughly) someone who would, *given that p* , realize such desires as prevailed given her other desires and beliefs. But there is a problem about incorporating this into a functionalist psychological theory as it stands. For this does not tell us, in behavioural terms however broadly construed, what someone with a certain belief would ever unconditionally *do*.

There are two ways in which the functionalist might try to make allowance for the fact that it is only where other relevant beliefs of the agent are true that behaviour as we naturally and widely conceive it is predictably matched with particular desires and beliefs. First, he might settle for using what one could call a world-conditioned notion of behaviour, saying, at the behavioural end of an output generalization, that a person would do things of this sort: such-and-such-if-the-world-is-as-it-would-be-if-relevant-beliefs-of-the-agent-were-true. (I return to this idea in section 5.) Second, he might restrict the notion of behaviour, so that something counts as a description of behaviour only if an agent can be expected to satisfy it irrespective of whether her beliefs are true. In talking about behaviour, he then confines himself to those things that an agent would do no matter whether the world were as she believed it to be – things, one might say, that she is *simply able* to do.

This provides the real explanation of why the functionalist should go back to the body in describing behaviour. And perhaps we can now also understand the functionalist’s other deviation from the behaviourist – his not treating actions themselves as behavioural items. Even if a person’s beliefs about what it is to ϕ are false, she will at least *try* to ϕ if she has overwhelming reason to ϕ (or so a functionalist may say). It seems then that the notion of *trying* or *attempting* can be introduced if one wants a means of saying in ‘purely psychological’ terms what someone’s beliefs and desires in conjunction do produce. One then arrives at a two-stage account of action production, such as can be found in some functionalist writings.⁷ At the first stage, one says how beliefs and desires modify one another and mediate the production of

attempts; so much is ‘pure psychology’, in the language of the T-terms. At the second stage, one says what attempts to do things would actually bring about, whatever the truth values of the beliefs that led to those attempts; the idea is that a sufficiently motivated agent who is *simply able* to do something will do that thing. This second stage takes one from the T-terms to the O-terms; and it is here that one is constrained to use the bodily movement vocabulary for describing behaviour, and to speak (not of actions themselves, but) only of things that are the most immediate, bodily effects of a person’s attempts.

Something like this functionalist view of action production is presumably shared by Kim (section 2). But what one now gets at the second stage of the account of action production will be instances of

a tries to ϕ & *a* is *simply able* to $\phi \rightarrow$ there occurs a ϕ -type movement

(where to be a ‘ ϕ -type movement’ is, intuitively, to be a movement of the type associated with actions of ϕ -ing). And this means that, unless we are prepared to say that an agent’s being *simply able* to do something is a ‘proper concern of psychological theory’, the argument that Kim gave in order to encourage us to suppose that only ‘bodily action’ is genuinely psychologically explicable could be used again now to show that even movements of bodies are not things ‘that it is the proper business of psychological theory to explain or predict’. But an agent’s being *simply able* to move a part of her body is constituted by the integrity and functioning of the relevant bits of her motor system and the absence of constraints on her body itself, and such things are in no obvious or intuitive sense psychological and would seem to be quite on a par with (for instance) the burner’s being such as to light when the knob is turned. Following through on Kim’s argument, then, the province of psychological explanation would become even more circumscribed than Kim allowed: the proper objects of psychological explanation could only be events described as agents’ *trying* to do things.⁸

4. Functionalists for their part will probably be happy to allow that conditions relating to agents’ *simple abilities* have to be specified using T-terms and be caught up with psychological theory. Their aim is to show how some of the brain’s complexity can be seen to mirror the complexity of the propositional–attitude scheme. And it might seem that the use of purely bodily O-terms for describing people’s outputs is in no way inimical to that aim. Although a bodily motion of behaviour is more restricted than an everyday one, there is much that can be said about people’s bodies’ movements, and it may seem that functionalists can avail themselves of anything that can be said about them and proceed to an interesting psychological theory.

But it must not be forgotten that functionalist output generalizations are still meant to be got from what we all know about action-explanation in knowing common-sense psychology. One thing that we know is that ϕ -ing is a proper explanandum of the common-sense psychological scheme only if

agents have some beliefs in the ascription of which ϕ -ing could be mentioned. So functionalists are not in fact entitled to use whatever bodily movement terms they like; their resources can include only such terms as could be used in giving the contents of agents' mental states. It seems, then, that they must refrain from using any very detailed bodily movement terms.

In fact it will be controversial exactly how much detail can enter into the bodily movement descriptions of common-sense psychology. If someone turns on a light (say), how detailed can a bodily description be of what she intentionally does, or tries to do? What beliefs about the movements of their bodies do people in practice employ? My own view is that hardly any detail can enter. When we engage in the practice of skills that require the manipulation of objects, for instance, it is unclear that we employ any beliefs which concern purely and simply the movements of our hands. It seems that a person can act as a result of having beliefs and desires, while having next to nothing in the way of beliefs about how her body moves when she acts.⁹ And if this is right, then the functionalist, in confining himself to bodily movements, confines himself to an extremely impoverished notion of behaviour indeed.

What is certain is that functionalists don't in fact envisage using a notion of behaviour that would strike us as at all impoverished. Even if common-sense bodily movement descriptions can be richer than I have just suggested, we may still doubt that they can be as rich as those that functionalists actually want to employ. In functionalist writings, one often finds what appear to be gestures towards great complication in accounts of behaviour. Armstrong spoke of 'making certain motions with the hand and so on'; he remarked that this was vague, and said that 'the matter might be investigated in a time-and-motion study for instance'.¹⁰ Lewis speaks of 'Karl's fingers moving on certain trajectories and exerting certain forces'.¹¹ It can seem as if the functionalists, feeling that the complexity of the propositional-attitude scheme must indeed demand some richness in the specification of behaviour, simply ignore the common-sense character of the truths about propositional attitudes that they represent themselves as beginning from.

It will be no good saying that, since every bodily movement does have some detailed description of which a student of time-and-motion or a physiologist could become apprised, any detail that the functionalist's aims require can always be introduced into functionalist theories. For one thing, the student of time-and-motion may discover that the sorts of bodily movements that agents think of themselves as going in for are not connected in systematic ways with the sorts of motions his studies concern.¹² And although it is surely right that there are, occasion by occasion, identities between the (coarse) bodily movement effects of actions and the (refined) bodily movement effects of finely discriminated states of the nervous system, it is unclear that this can help the functionalist who is trying to avoid a notion of behaviour that strikes him as too crude for the use to which he wants to put it. Someone who hoped to use physiological knowledge occasion by occasion to pin down the neurophysiological states that caused some effect of some action would have lost sight of

one of the functionalists' aspirations – to use our knowledge of interpersonal psychology to define types of mental state.

5. I suggest that some of the allure of functionalism has resulted from failure to keep track of the use of the simple term 'behaviour'. The elements of common sense that give rise to the idea of a psychological theory seem correct when 'behaviour' is understood in Ryle's way, as including all the many things an agent does. The idea of a functional theory realized in neurophysiological states seems correct when 'behaviour' is understood in (say) the physiologist's way, as an agent's moving her body in all kinds of complex fashions. These two notions of behaviour overlap, and when 'bodily movements' is used to catch them both, they are made to appear to coincide. But the two notions do not coincide. And if one wants to preserve both common sense and the idea about functional theories, then one can only conclude that there is a complexity in propositional–attitude psychology that does not derive from any complexity in people's bodily movements conceived in ways available to common-sense psychology.¹³

No doubt many functionalists will say that theoretical psychology has to be enlisted in the service of common-sense psychology. They would make proposals about how theoretical psychological findings could be brought to bear on common-sense psychological states, and they would claim that the proposals will enable us to discover states that must be counted as beliefs and desires even though common-sense psychology unaided would never have recognized them as such.¹⁴ It is as if common-sense psychology had a hidden complexity that the theoretical psychologist could uncover experimentally; as if the superimposition of the picture of the person on the picture of the brain could reveal a sort of complexity in the picture of the person which ordinarily goes unheeded. But why should we think that common-sense psychology, in order to achieve what we can all achieve using it, must really be capable of achieving a great deal more that non-theoreticians will never know about? If common-sense psychology has no concern with how exactly we move our fingers when we turn on lights (say), then this is because we do not have to try to move our fingers in the exact way in which we actually move them in order to turn on a light when we want to. But where the details of bodily movements are not within common-sense psychology's province, how can that which bears on the details have a bearing on common-sense psychological states? How can theoretical psychology dictate to common sense answers to questions that it is in the nature of common sense not to ask?

Instead of resorting to theoretical psychology, we could suppose that the picture of the brain cannot be superimposed on the picture of the person because the picture of the person has its own fine points which are not such as to be exposed in the structure of the brain. What we should then have to exploit in understanding the felt complexity of propositional–attitude psychology is not the brain's complexity, but our knowledge that common-sense psychology enables us to explain so much more than why there are the

movements of people’s bodies that there are. The step from a Rylean sort of behaviourism to functionalism will then seem to have been, in a way, a retrograde step. If mental states are to be thought of as dispositions of any sort (or, if you prefer, as states that are parts of systems that exhibit an overall structure), then, to the extent that they are dispositions to behave (or states connected systematically with ways of behaving), the relevant notion of behaviour is the broad one that the philosopher behaviourists used and the functionalists left behind.

If we do employ the ordinary and richer conception of behaviour in specifying the upshots of mental states, we cannot hope to circumscribe mental states in anything like the way that the functionalist envisages. Recall what was wrong with the old behaviourist’s conception from the functionalist’s point of view. Using that conception, one cannot leave the truth or falsity of agents’ beliefs out of account. We imagined that this point might be accommodated by using a ‘world-conditioned’ notion of behaviour, but left this suggestion rather vague (section 3). Now the ramified character of the interdependencies between mental states, which the functionalist is so anxious to take account of, ensures that any worldly conditions incorporated in a notion of behaviour would ramify in any theory that attempted to accommodate that notion. A person can be expected to do what she tries to do on occasion only if certain beliefs that explain her then trying to do that are true. But the interdependencies between mental things ensure that for any desire or belief whose causal role we might think to define, it is possible that almost any belief might interact with it in the production of some possible event of trying.¹⁵ Thus if the world-conditioned notion of behaviour is introduced by the functionalist, and from case to case he makes it explicit which beliefs are such that their truth or falsity on occasion is relevant to what behaviour is produced, his task turns into the project of giving an account of the structure of rational thought and practice, any exemplification of which is conditioned by a simultaneous view of the world as a subject confronts it. This is not the project of providing descriptions, however abstract, of the brain.

6. It is not a novel claim that explanation in the rational mode cannot be converted into science. As Davidson has said:

Any effort at increasing the accuracy and power of a theory of behaviour forces us to bring more and more of the whole system of the agent’s beliefs and motives directly into account. But in inferring this system from the evidence, we necessarily impose conditions of coherence, rationality, and consistency. These conditions have no echo in physical theory.¹⁶

I take the mismatch Davidson sees between the mental and the scientific physical to show up in the fact that an attempt to incorporate conditions of rationality in a physicalist theory, using a conception of behaviour that is

bodily but constrained by common-sense psychology, seems to leave something out. Even to its proponents it seems to leave something out, and they proceed by injecting some extra detail into bodily behavioural descriptions (cf. Armstrong and Lewis). But there is no warrant for the extra detail.¹⁷

Davidson himself thinks that the mismatch between the mental and the scientific physical shows up in two particular ways at the level of what can be said about people's 'outputs'. First, 'Practical reasoning . . . may simply fail to occur'. Second, 'Wanting to do something . . . may cause someone to do [the] thing, and yet the causal chain may operate in such a manner that the act is not intentional'.¹⁸ These two claims surely reveal an immediate and insuperable obstacle to constructing functionalists' output generalizations. But I have allowed the argument to progress, believing that the superficial plausibility of the functionalists' contrary claims derives in large part from their free use of a quite schematic notion of *behaviour*.

The idea upon which the arguments here have traded is present in Davidson too, of course – in the claim that the mental is not a closed system. The felt complexity of propositional–attitude psychology will be accommodated only when 'the constitutive role of rationality' is properly acknowledged, and the attempt to see the patterns in a person's mental states embodied in the states of physical science is duly abandoned.¹⁹

7. The fundamental assumption that has been in dispute is that, in stating the causal powers of mental states, one can prescind from all but the most immediate effects of the actions they produce, and ignore almost everything under the head of 'desired effects of actions' in the picture of the person (section 1). This assumption underlay the physicalist line of thought sketched at the outset. And we shall see now that it is the same assumption which leads people to accept the supervenience of the mental on the neurophysiological, and which gives rise to another physicalist view of intentional states of mind.

To many it seems (a) that a difference of mental state between two people requires some difference in their behavioural dispositions, and (b) that a difference in the behavioural dispositions of two animal bodies requires some difference in their internal physical machinery. They think, then, that if one were to allow that there could be a mental difference without a difference in brain state, one would be denying that the brain was responsible for the production of behaviour.²⁰

Their argument is guilty of the same equivocation on 'behaviour' as the functionalists rely on. Premiss (a) requires for its truth a broad and everyday conception of a behavioural disposition; (b) requires a narrow one. (a) is true if we take it to mean that a change in mental state affect the proper explananda of psychological explanations; (b) is true if we take it to mean that only a change in brain state could affect how a creature moves itself. Nothing in the argument holds these two conceptions of a behavioural disposition together.²¹

Of course it is well known that there are counterexamples to a thesis of the supervenience of the psychological on the neurophysiological. Putnam's Twin

Earth examples show that there can be variations in the objects of *de re* states of mind that are not reflected in any dispositions to move the body one way rather than another.²²

Some proponents of the supervenience thesis try to show that these examples do their thesis only negligible damage, as if the existence of *de re* states posed some special, local problem. But we saw that the so-called holism of the mental is apt to embrace all of those worldly facts which a person's attitudes concern and which her bodily movements confront: the problem for supervenience is not a problem specifically about *de re* states of mind.

In some physicalist writings, this last point is acknowledged, and it is agreed that propositional–attitude states cannot be characterized as the functionalist envisages; but it is then said that these states must nevertheless have causal–explanatory *components*, which components may be seen to coincide with brain states. According to this new view, the picture of a person from which we begin is not itself something upon which any picture of the brain can be superimposed; but the picture of the person can, as it were, be split into two, and one of the resultant parts – the ‘internal side’ of a person, which is supposed to incorporate explanatory states – is suited to having some picture of the brain fitted on to it.²³

Yet it is hard to see how anyone is in a position to claim that there are states whose ascription to people is explanatory of their behaviour unless he can demonstrate that the ascription of such states does, or would, cast light upon behaviour. We know of course that such states as we ascribe – beliefs having contents, for instance – do cast light. But it is no help then to be told that there must be states which lurk behind the states we ascribe and which carry their explanatory force. It is a strange idea that the satisfaction yielded by common-sense explanations has its source in something of which the parties to the explanation are quite ignorant – as if light had been cast through a medium that we cannot yet see through. But it would be a quite baffling idea that the explanatory force of an explanation resides in something that is not capable of illuminating anything for us – as if we could be sure that light will one day pass through a medium that is always opaque.

Not only can the picture of the brain not be superimposed upon the person, then: we have no reason to believe in any picture of a person's non-worldly aspect for it to be superimposed upon instead.

8. These conclusions ought not to surprise anyone who accepts that our reason for believing that mental states are occupants of causal roles is given by pointing to the place of mental states in causal explanation. For nothing in the argument here is hostile to the thought that causal roles are constitutive of at least some mental concepts. It can be true that the explanatory task that propositional–attitude ascription serves is a causal one; and it can be true that we cast all the light we can on propositional–attitude concepts by saying (not in functionalist theories, but in the available ways) what explanatory task their ascriptions serve.

It will be said that there is a puzzle here, however. How can the propositional–attitude states be thought of as mediating causally between inputs to and outputs from persons, although nothing with the appropriate causal powers of mental states can be found by scrutiny of a person’s interior? Does not our conception of causality compel us to see the states which are cited in causal explanations of (*inter alia*) movements of a body as states which are located on causal chains that can be traced through space and time and that run through space–time volumes incorporating movements of that body? But then are we not obliged to see bodily movements as somehow primary among the explananda of action explanations? (Some line of thought such as this must be what lends plausibility to arguments like Kim’s in section 2.)

One will feel tremendous pressure to accept this if one adopts a paradigm of causal explanations, got (say) from the picture of the brain, and takes it that the causal explanations obtained in viewing a person as a person must also conform to that paradigm. But if the causal–explanatory powers of mental states cannot be specified in such a way that a scientist could be led to recognize states that are the subjects of his studies as having those powers, then the belief that common-sense psychological states conform to the paradigm is undermined. The impossibility of specifying the causal–explanatory powers of mental states in ways that would suit a scientist is revealed in the difficulty of finding a notion of behaviour which is both available to common-sense psychology and rich enough to define states that are explanatory according to the paradigm even while they share in the complexities of common-sense psychological states.

9. Why is the idea that propositional–attitude states can be fitted to the scientific paradigm so compelling? I suspect that an unacknowledged allegiance to principles of positivist epistemology must take a share of the blame.

If one begins with a distinction between psychological terms and non-psychological terms, and dresses this up in a distinction between T-terms and O-terms such as Lewis’s (section 3), then one comes to think of the O-terms as conveying all the data from which psychological theories could be constructed. Application of the O-terms seems then to be independent of anything one knows about people (*per se*), and common-sense psychology begins to seem to be a theory of such observables as the O-terms describe, a theory distinctive only in its particular concepts.

But reflection on the practice of psychological explanation shows what an extraordinary myth this is.²⁴ Someone required to explain why some agent has done something has to show how the psychological facts about the agent are consistent with what she ostensibly did. This may require him to become clearer about what went on in the world even as he speculates about her mental states. (Equally of course he may learn about the world by learning of her states of mind.²⁵) It is not, as the model of theory and observation might suggest, that he has to arrive at a view about what went on in the agent’s head

which coheres with some prior account of what happened at the place where her body meets the world.

We ought not then to expect to find any notion of behaviour (‘the observable’) that is suited to reductionist claims. Certainly Rylean talk of skaters’ dispositions to warn other skaters seems laughable if it is read as offering any reduction of believing that the ice is thin. And there is reason to suppose that the features of Ryle’s behavioural term which contaminate them psychologically must in fact be present equally in any terms that figure in any account of mental things. There are two (related) ways in which the application of everyday behavioural terms is caught up with the application of psychological terms. First, bits, or items, of behaviour, as described by behavioural concepts, are the effects of mental states; but it is impossible to divide behaviour up into bits in such a way that the bits correspond to things that have a psychological history unless we know something about the mental states that actually produce the behaviour. Second, Ryle’s descriptions tell us of things that the agent intentionally did; and one is not in a position to take a view of which things are intentionally done by people unless one has some view of their mental states. These two features are bound to be inherited by any behavioural descriptions that are fit for inclusion in an account of mental states: their application must presuppose (a) a psychologically informed method for articulating the events that flow from a person, and (b) a sense of what is psychologically relevant among the events thus articulated.²⁶

If one rests content with a naive conception of psychology’s province, one cannot then construe ‘observable’ in positivist spirit and think of behaviour, the objects of psychological explanation, as ‘observable’. Helping oneself to ‘observable’ behaviour, on the other hand, is only a way of ensuring that one ignores the truth about psychological explanation.

10. Section 9 is no doubt inadequate as a diagnosis of the attractions of present-day physicalist accounts of intentional phenomena. Here I suspect that the whole history of the subject is to the point. Philosophers of mind have come to see Cartesian dualism as the great enemy, but have underestimated what they have to contend with. Taking the putatively immaterial character of minds to create the only problem that there is for Descartes’ account, they marry up the picture of the person with the picture of her brain, and settle for a view of mind which, though material in its (cranial) substance, is Cartesian in its essence.²⁷

Of course the acceptance of immaterial substance *was* one of Descartes’ errors. But it does not take a scientific materialism to remedy that error. After all, it has not been said that there are elements of the person picture that science fails to deal with because they are ethereal and unnatural, but only that we have to look for the source of common-sense psychology’s complexity elsewhere than at the junction between the central nervous system and the world.

Notes

- 1 Jaegwon Kim, 'Psychophysical Supervenience', *Philosophical Studies*, xli (1982), 64. (I have changed the order of Kim's sentences.) For a similar argument, given in the course of a defence of functionalism, see Brian Loar, *Mind and Meaning* (CUP, Cambridge, 1981), 88.
- 2 It may be said that this counterexample is importantly different from the examples that were Kim's concern, because in the counterexample it is only in the presence of the explanans that we come to be able to separate necessary conditions for the obtaining of the explanandum. But this feature may also be present in psychological cases. We can know that someone turned on the burner (and that there is a psychological explanation of that) without knowing what sort of bodily movement on her part resulted in the burner's being on.
- 3 David Lewis, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 1(1972), 249–58. (Where Lewis uses 'state' I use 'type of state', because, unlike Lewis, I reserve the word 'state' for particulars; and I make no assumption that functionalists are committed to any type identities of Lewis's sort.)
- 4 'Bodily movements' is used by Loar and others; 'motor responses' by Lewis in 'Psychophysical and Theoretical Identifications'.
- 5 *The Concept of Mind* (Hutchinson, London, 1949), 129.
- 6 There is little here that is uncontroversial. (For an account of the controversy, see my 'Bodily Movements, Actions and Mental Epistemology', in *Midwest Studies in Philosophy*, vol. IX (1985).) What I assume now is that we can distinguish the denotations of descriptions such as 'a's raising her arm', which are actions, from the denotations of such descriptions as 'a's arm's rising', which are not actions.
- 7 See *Mind and Meaning*, 86–91. Loar uses a technical, theoretical psychological notion of *willing*, where my exposition uses *trying*. Loar's view of theoretical psychology is discussed in section 5. For arguments (in effect) that 'try' has many of the properties needed for the two-stage account of action production, see Chapters 3 and 4 of my *Actions* (Routledge and Kegan Paul, London, 1980).
- 8 Loar's argument for narrowing down psychological explanation (see n. 1) introduces *basic actions*, which Loar calls 'primary explananda'. He says that non-basic things are explained by 'independent facts', that is facts that are not themselves psychologically explained. The question for Loar is quite parallel to that for Kim: is the reason offered for extricating 'independent facts' from non-basic things not also a reason for extricating non-psychological facts even from basic things?

There are two points about Loar's terminology. (a) I have spoken of *basic things* (*done*), rather than *basic actions*, because actions themselves, assuming these are particulars, do not stand to one another in relations of relative basicness. (See *Actions*, Chapter 5, where the relation *more basic than* is taken to hold between descriptions of actions, or, better, my 'Action and Ability', in R Haller (ed.), *Language, Logic and Philosophy* (Reidel, Dordrecht, 1981), where the relation is taken to hold between the things a person does.) (b) It may be that Loar himself does not distinguish actions from bodies' movements, as my exposition of functionalism has suggested functionalists do.

I state the second stage of a functionalist's account of output as I do because I think that (on occasion) an agent's ϕ -ing is the same as her trying to ϕ and not the same as a movement of the ϕ -type. Still the consequent of the conditional

“PHYSICALIST THINKING AND BEHAVIOUR”

might read ‘a ϕ s’ – only then the conditional cannot be supposed to take one from cause to effect (unless one denies that, on occasion, events of trying simply are actions). At any rate, the argument addressed to Kim and Loar requires only that some condition relating to the agent’s body’s functioning, as well as some condition relating to the agent’s mind, is necessary for the occurrence of an action. To accept this much, no stand needs to be taken on the controversy mentioned in n. 6, or on the details of a correct formulation of the two-stage account of action production.

- 9 See my discussion of what I there called *teleologically* basic descriptions of actions in Chapter 6 of *Actions*. And notice that I make no assumption to the effect that all beliefs are linguistically expressible by their possessors.

If bodily movement descriptions of actions are not basic, then there will be some *non-bodily* basic descriptions for functionalists to use in describing outputs. But if the functionalist does use output O-terms which touch on regions beyond the agent’s body, then (a) it will become more implausible that his ‘simple ability’ conditions are ‘purely psychological’, and (b) this will do nothing to supply the richness that functionalists seem to want to find in their notion of behaviour (see below).

- 10 D. M. Armstrong, *A Materialist Theory of The Mind* (Routledge and Kegan Paul, London, 1968), 147.
- 11 David Lewis, ‘Radical Interpretation’, at p. 114 in the reprinting in his *Philosophical Papers*, vol. I (Clarendon Press, Oxford, 1983), quoted more fully at n. 26. Many functionalists speak of behaviour in a purely schematic way.
- 12 I am relying here on such claims as are made for example by P. F. Strawson and G. J. Warnock in D. F. Pears (ed), *Freedom and the Will* (Macmillan, London, 1963): our system of classifying actions is grounded in quite different interests from any system of bodily movement classification. I do not conclude, as they did, that actions are not bodily movements; but I do take their claim to show the irreducibility of movements-classified-by-someone-interested-in-action to movements-classified-by-the-scientist-interested-in-movements-*per-se*.
- 13 It may be pointed out that complexity could be derived from complexity in descriptions of *inputs*. (Functionalists insist that stimulus terms as well as behavioural ones are needed to characterize the mental ones: this is a respect in which functionalists are thought to differ from behaviourists which I have not singled out for attention.) But I think that there are things to be said about *perception* which lead in the same direction as the things I have said about *action* and which would show that the functionalist cannot make anything of this point. And the argument I actually give is meant to rely on the fact that my opponents themselves believe that more detail is needed in *output* specifications than (so I say) they are entitled to.
- 14 This explains the technicalities in Loar’s account (see n. 7). Loar’s suggestion is that if a theory well confirmed by experiments in theoretical psychology and neurophysiology were true of a person, and certain of the functional states of the theory satisfied the full complement of the constraints imposed by common sense as necessary conditions of having beliefs and desires, then it would be correct to count those functional states as beliefs and desires.
- 15 The ‘certain’ beliefs are (intuitively, and leaving out modifications that would be required to accommodate fortuitously false beliefs) those whose truth is required for the agent’s trying to ϕ to result in the agents ϕ -ing.

It is easy to miss the point about the ramification of worldly conditions if one thinks about explanation as we know it and forgets about the predictive aspirations of functionalists' output generalizations. When we know the explanation of an agent's doing something, we are in a position to specify a small number of beliefs which enter the explanation of her doing that, and only some of these are beliefs whose truth values bear on whether she has actually done the thing. But this is not the position of someone who hopes for a general, predictive theory of what agents would do. Consider a functionalist who hoped to derive actual mental/physical identities for the case of a particular person at a particular time. He might start with a list of types of state which were instantiated in the person at that time. For any type of state on that list, he will say (counterfactually) that its instance would interact thus and so *if* . . . ; for any type of state not on that list, he will say (doubly counterfactually, as it were) that *if* there were an instance of it, then it would interact thus and so *if* . . . Thus *possible* events of trying must be seen (at *t*) as such as to be produced in a person (at a time later than *t*) by way of beliefs that that person may lack (at *t*). (Some people seem to forget that a functionalist's psychological theory must make mention of all the types of mental state there are. I have put the matter as I have here only because it may help to remind one of this.)

- 16 This is from 'Psychology as Philosophy'; see Donald Davidson, *Essays on Actions and Events* (Clarendon Press, Oxford, 1980), 231. This volume, 24.
- 17 I say that there is no warrant for the detail. But I should acknowledge that I have an argument only in so far as I am in a position to ask (rhetorically) 'What reason could there be for supposing that the detail is warranted?' ('What reason could there be for supposing that theoretical psychology can dictate to common sense?'). Of course the committed proponent of scientific materialism thinks that there are reasons where I see none: he supposes that metaphysical principles provide the warrant. I try to engage with his position in sections 8–10.
- 18 The quotations are from 'Freedom to Act', at pp. 77 and 78 in *Essays on Actions and Events*.
- 19 John McDowell also argues that if expositions such as Loar's seem to undermine Davidson's claims about the mental's anomalousness, then that could only be an illusion. He focuses attention on 'internal constraints', where I have focused on 'output generalizations'. See 'Functionalism and Anomalous Monism', in Ernest Le Pore and Brian McLaughlin (eds.), *The Philosophy of Donald Davidson: Perspectives on Actions and Events* (Blackwell, Oxford, 1986).
- 20 The argument is in Colin McGinn, *The Character of Mind* (Oxford University Press, Oxford, 1982), 29; though McGinn would not endorse the argument just as it stands (see n. 23).
- 21 Of course anyone can if he wants, put common-sense descriptions of behaviour together with scientific descriptions of behaviour, and call what he arrives at descriptions of behaviour. What cannot be guaranteed, however, is that, having assembled a notion of behaviour by reference to two explanatory schemes, one has then accorded some stable sense to 'explanation of behaviour' or 'behavioural disposition'. (I have not said that no one is entitled to the functionalists' conception of behaviour; I have only questioned whether one is entitled to suppose that it can be put to the use to which functionalists put it.)
- 22 See Hilary Putnam 'The Meaning of "Meaning"', in his *Mind, Language and Reality* (CUP, Cambridge, 1975).

“PHYSICALIST THINKING AND BEHAVIOUR”

- 23 See for example Colin McGinn, ‘The Structure of Content’ in Andrew Woodfield (ed.), *Thought and Object* (Clarendon Press, Oxford, 1982). McGinn himself argues that the notion of *content* is decomposable into explanatory and truth-conditional aspects. The explanatory states he envisages do not have semantic content by virtue of their explanatory role, and the arguments below apply to his position inasmuch as this is so.
- 24 The story that mental terms were actually introduced as theoretical terms is called a myth by David Lewis (op. cit. n. 3). In support of the idea that it is a good myth (sc. that our terms for mental states mean just what they would if the myth were true), Lewis says that if it were a good myth, it would explain the appeal of Rylean behaviourism. But the appeal of Rylean behaviourism cannot be separated from its use of a broad and everyday notion of behaviour; and this is a notion which ought not to be available to one who tells the mythical story. (See below, and n. 26 on Lewis on behaviour.)
- 25 That this is a point about psychological explanation, rather than about ‘head/world correlations’, will be apparent only if one takes a relaxed view of psychological explanation; see section 2.
- 26 Lewis is sensitive to the fact that intentional notions must be precluded from accounts of behaviour if they are to be put to a theoretical use. He wrote (loc. cit. n. 11):

There is an ambiguity in the term ‘behaviour’. . . . I am using it to refer to raw behaviour – body movements and the like – . . . ; not to behaviour specified partly in terms of the agent’s intentions. . . . That Karl’s fingers move on certain trajectories and exert certain forces is what I call ‘behaviour’; that he signs a cheque is not.

What Lewis seems to want to rule out from ‘raw’ behavioural descriptions is only instances of ‘ ϕ ’ such that any event of someone’s ϕ -ing is an event of her intentionally ϕ -ing. There is no need to deny that this ruling can give us a notion of behaviour. What I do deny is that such a ruling provides us with a notion that might have been used by someone ignorant of all common-sense psychological truths. I deny, then, that Lewis has found a notion fit for his ‘good myth’. (Of course there *is* a notion of behaviour (or anyway of *output*) that we can imagine applied in utter independence of any interest in persons; but with this (very raw) notion, we return to one on which we get no real purchase when we state common-sense psychological accounts.)

- 27 What Descartes’s commentators are typically most anxious to remind us of is that Descartes should not have held both that mind and body are substances whose essences are distinct, and that mind and body causally interact. Bernard Williams is an exception: he describes a difficulty which is independent of Descartes’s treating the mind as soul-like, and depends only upon Descartes’s thinking that all the transactions between the mental and the physical happen, as it were, at an interface. See the discussion of *terminal* interactionism (as Williams says we might call it) in *Descartes: The Project of Pure Enquiry* (Penguin, Harmondsworth, 1978), 288–92.

