Journal of Consciousness Studies
www.imprint-academic.com/jcs

E.J. Lowe

Self, Agency and Mental Causation

A self or person does not appear to be identifiable with his or her organic body, nor with any part of it, such as the brain; and yet selves seem to be agents, capable of bringing about physical events (such as bodily movements) as causal consequences of certain of their conscious mental states. How is this possible in a universe in which, it appears, every physical event has a sufficient cause which is wholly physical? The answer is that this is possible if a certain kind of naturalistic dualism is true, according to which the conscious mental states of selves, although not identifiable with physical states of their brains, are emergent effects of prior physical causes. Moreover, mental causation on this model promises to explain certain aspects of physical behaviour which may appear arbitrary and coincidental from a purely physical point of view.

I: Introduction

The following claims all seem fairly compelling, upon reflection, and yet they appear not to form a consistent set:

(1) The self, although physically embodied, is not to be identified with any physical body nor with any part of such a body.

(2) The self is by its very nature an agent, something that is naturally capable of performing intentional actions, some of them with physical results.

(3) Every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event (and rarely if ever is a physical event causally overdetermined).

The apparent inconsistency of this set of claims has led many philosophers to reject one or more of them. Some reject (1), either denying that there is any such thing as the self, or else identifying it with something bodily, such as an animal organism or brain. Some reject (2), holding that our experience of volitional control over our bodies is merely illusory. And some reject (3), maintaining that the self's intentional states are non-physical causes of certain physical events which lack sufficient wholly physical causes. (This appears to have been Descartes' view.) Instead, I shall argue that claims (1), (2) and (3) are in fact perfectly consistent. Whether all of those claims are *true* is

another matter — though, clearly, if they are all not only fairly compelling but also consistent, something is to be said in favour of their all being true. I should add, though, that elsewhere I have argued in defence of claims (1) and (2) (see Lowe, 1996). Consequently — in view of the widespread acceptance of claim (3) — I have a vested interest in establishing the consistency of the three claims.[1] So, before proceeding, let me briefly explain why I think that claims (1) and (2) are true.

II: The Self Is Not Its Body

I believe, first of all, that selves exist, not least because I believe that *I* exist and consider myself to be a 'self'. I use the term 'self' interchangeably with the term 'person'. I take it, however, that the term 'self' is a particularly appropriate synonym for 'person' because it reflects the fact that a necessary condition of personhood is a capacity for *self-reference* — a capacity which is manifested linguistically by use of the first-person pronoun, 'I'. A person or self, in short, is a being which can have thoughts *about itself*, of the sort that are appropriately expressed (in English) by sentences containing the first-person pronoun, 'I', as their grammatical subject — sentences such as 'I feel hot' and 'I am six feet tall'. But I also believe that a person or self, even though physically embodied, is never to be *identified* with its physical body nor with any part of it, such as that body's brain. This is claim (1) above.

Our ordinary self-conception seems to involve a commitment to claim (1). For example, when I have a conscious first-person thought — such as the thought that I feel hot — I regard *myself* as being the subject of this thought, both in the sense of being the thing having the thought and in the sense of being the thing that the thought is about. But I am not at all inclined to regard *my body* or *my brain* as being its subject, in either sense. Since I am the subject of the thought but neither my body nor any part of it is, it follows that I am not identical with my body or any part of it. Of course, with the benefit of a little scientific knowledge, I may well be prepared to concede that, but for the existence and normal functioning of my brain, I could not so much as have this or any other thought: but that doesn't (or shouldn't) persuade me to believe that *my brain* is, after all, the subject of my thoughts. That would be like inferring that my feet run from the fact that I could not run without having feet. Anyway, quite apart from anything else, it seems clear that, even granted that I need *a* brain in order to be able to think, I don't need to have the particular brain that I do have. I find nothing inconceivable in the thought that I might wake up one morning to be told (truly) that, overnight, I had undergone an operation in which my old organic brain was somehow replaced by a new inorganic one.

Here it may be objected that, if I am not to be identified with my physical body nor any part of it, then it only remains for me to be identified with something altogether *non*-physical, such as a spirit or soul or 'Cartesian ego' — and this, it will be said, is a view wholly at odds with a naturalistically acceptable conception of persons. However, it is a simple mistake to suppose that if I am not to be identified with my physical body or any part of it, I must therefore be identified with something non-physical, that is, with something possessing no physical characteristics whatever. And, indeed,

---

[1] For a recent example of a philosopher who endorses claim (3) — and espouses a thoroughgoing physicalism as a consequence — see David Papineau (1993), p. 22.

identifying oneself with something non-physical is quite as counterintuitive as identifying oneself with one's physical body or brain. It seems to me no less literally true that *I* have a certain height than it seems literally false that *my brain* has certain thoughts.

The self can be a 'physical' thing — possess physical characteristics such as height — even though it has different identity-conditions from those of the body or brain. Somewhat analogously, a statue can be a physical thing — possess physical characteristics such as shape — even though it has different identity-conditions from those of the piece of matter which composes it. The analogy isn't perfect, however, for I don't want to say that the relation of embodiment is simply one of composition: I am not composed by my body, in the way that the statue is composed by bits of matter. Indeed, I don't believe that the self is a composite entity at all: I don't believe that it is literally *made up of* distinct and separable parts. The self, I want to say, possesses a strong kind of unity which is incompatible with its being a composite thing. I don't have space to argue for this view here (but see further Lowe, 1996, Ch. 2). All I want to stress at present is that claim (1) above is not only plausible, but is perfectly consistent with the equally plausible claim that the self is a physical thing, in the sense of being a thing which possesses physical characteristics or states.

III: Mental States Are Not Physical States

However, it doesn't follow from this that *mental* states of the self can intelligibly be thought of as being *physical* states of it, akin to such physical states as height and weight. Indeed, I very much want to *deny* that mental states are physical states, even though they are states of something physical — the self. This is because I can make no clearer sense of the idea that a conscious mental state might just *be* a physical state than I can of the idea that a physical object might just *be* a natural number (cf. Geach, 1979, p. 134). Consider a typical mental state, such as this one: consciously thinking of Paris. I know what it means to be in such a state, at least as clearly as I know what it means to be in the physical state of sitting in a chair. But I cannot at all understand what it would mean to say that the state of consciously thinking of Paris just *is* a 'physical' state. This is because — as I understand it — a physical state is, by its very nature, one whose possession by a thing makes some real difference to at least part of the space which that thing occupies. Thus, my sitting qualifies as a physical state of me because, in virtue of possessing it, I fill out a part of space in a certain way, rendering that part of space relatively impenetrable by my presence. But my consciously thinking of Paris has no spatial connotations of this sort whatsoever, so far as I can see (cf. McGinn, 1995). In fact, the identity-conditions of mental states would appear to be thoroughly unlike those of physical states — as unlike them as the identity-conditions of physical objects are unlike those of the natural numbers (see further Lowe, 1989, pp. 131–3 and Lowe, 1996, pp. 25–30). And consequently the thesis that mental states 'just are' (identical with) physical states is simply unintelligible.

A whole generation of philosophers has, alas, mistaken this unintelligible thesis for something much more exciting, namely, a profound truth which has only now begun to be revealed to us through the advance of science. (I don't expect to be able to shake their faith, however, any more than one could hope to shake the faith of a dedicated Pythagorean.) Truths of identity simply *cannot* be exciting in the way such metaphy-

sicians fondly imagine, because it can only be intelligible to identify items *of the same kind* (that is, kinds importing the same identity-criteria for their instances), and the 'exciting' identifications — of physical objects with mathematical objects, or of mental states with physical states — all violate this principle by trying to identify items of quite *different* kinds.

IV: Selfhood Requires Agency

A word or two is now needed in defence of claim (2) — that the self is by its very nature an agent, something that is naturally capable of performing intentional actions, some of them with physical results. Since I have already characterized the self as something necessarily capable of self-reference, I have already implicitly characterised it as something necessarily possessing agency, since self-reference is a species of intentional action. To refer to oneself as 'I', whether in speech or merely in thought, is to perform a kind of intentional act. If done merely in thought, this act may perhaps have no physical results, though if done in speech it clearly must. However, the idea that there might be a self which, throughout its life, was *only* capable of engaging in intentional actions of a purely mental kind — never, thus, in actions having physical results — is one that is hard to credit.[2] Such a self would be constitutionally incapable of communicating with other selves. It is strongly arguable, however, that the development of self-awareness is necessarily linked to the development of other-awareness and that both are necessarily linked to the development of powers of communication, whether through language or merely through various kinds of non-verbal behaviour. If that is so, then there couldn't be a self which was constitutionally incapable of communicating with other selves throughout its life — though there might, conceivably, be a self which *lost* this capability having once developed it, as is suggested by cases of so-called 'locked-in syndrome' (people who seem to remain self-aware even though they have lost all control over their bodies through complete paralysis of the non-autonomic nervous system).

Another reason for thinking that a self must be capable — at least at some stage during its existence — of performing intentional actions which have physical results is that it is strongly arguable that only a being capable of such actions can develop a concept of *causation* and that possessing such a concept is a necessary condition of self-reference and thus of selfhood itself. (It is a necessary condition of self-reference because to self-refer is to perform an intentional action, to perform an intentional action is to act in a certain way *knowing* that one is so acting, the concept of intentional action is a causal concept, and knowledge is possible only for one who possesses the requisite concepts.) The thought here, then, is that a being that was condemned from birth to complete physical passivity, even though endowed with powers of sensation and perception, would be incapable of distinguishing between causal and non-causal sequences of events, because an ability to make this distinction depends upon an ability to *intervene actively* in the course of nature, with a view to discovering by means of experimental manipulation which events do or do not depend upon which other events (cf. von Wright, 1971, pp. 69–74). One's own inner

---

[2] Thus I find Galen Strawson's imaginary example of the 'Weather Watchers' highly implausible (see Strawson, 1994, Ch. 9).

mental life does not present a sufficiently independent arena in which this capacity could be developed, it seems: one needs to be able, as it were, to get to grips with things *outside* oneself in order to get any purchase on the thought that some events stand in causal relations of dependence to one another whereas others are only accidentally conjoined.[3] This line of reasoning is, I confess, only very sketchily presented here, but that is because its full articulation would require much more space than I have available.

V: Are the Three Claims Inconsistent?

Why should the three claims stated at the beginning of the paper be *thought* to be inconsistent? For the following reasons, I imagine. First of all, claim (2) seems to imply — indeed, I agree that it *does* imply — that intentional states of the self can be causes of physical events. This is because the concept of an intentional action is a causal one: when an agent acts intentionally, an intentional state of that agent plays a causal role in the production of some event — an event which, in the case of an intentional action which has a physical result, will obviously be a physical one.

Next, claim (1) seems to imply that intentional states of the self are states of something non-physical and are therefore themselves non-physical states. Now, of course, we have just seen that claim (1) does *not*, in fact, imply that the self is something non-physical. But we have also seen that there is, all the same, good reason to think that even though the self is physical, inasmuch as it possesses physical states, *mental* states of the self — including its intentional states — are *not* physical states of it and so are indeed non-physical states. So, although claim (1) does not strictly have the implication it might seem to have — that intentional states of the self are non-physical states — I think that any adherent of claim (1) ought nonetheless to *accept* the thesis that intentional states of the self are non-physical states.

Finally, claim (3) seems to imply that no physical event can have a non-physical state amongst its causes. (We shall examine this alleged implication in a moment.) Together, then, claims (1), (2) and (3) — or, more accurately, claims (2) and (3) together with the thesis, consistent with claim (1), that intentional states of the self are non-physical states — seem to imply that non-physical states *both are and are not* causes of physical events: a contradiction. However, even if we grant the alleged implications of claims (1) and (2), this reasoning is incorrect, because it ignores the *transitivity of causation*, as we shall now see. (The key point to appreciate here is the very simple one that if x is causally sufficient for y and y is causally sufficient for z, then, by transitivity, x is causally sufficient for z, but that this doesn't imply that z is *causally overdetermined* by both x and y.)

---

[3] Against me here it might be urged that a capacity to discriminate perceptually between (at least some) causal and non-causal sequences of events could be *innate*, even in a completely passive creature, and indeed that there is some empirical evidence for such an innate capacity in human infants. However, it could still be argued that such a capacity would inevitably be destined to lie dormant or atrophy in any creature incapable of engaging in active exploration of its perceptual environment (including here as 'active exploration' a creature's voluntary direction of its sense organs, such as its eyes, towards stimuli selected by it for attention).

VI: Naturalistic Dualism Is Possible

It is possible for claim (3) to be true — that every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event — and yet for it also to be true that some physical event, *P*, has a non-physical event or state, *M*, amongst its causes (without envisaging this as involving the causal overdetermination of *P*). This is because *M* itself may have a set of wholly physical causes which are collectively causally sufficient for *its* occurrence. If *M* is a cause of *P*, then, by the transitivity of causation, all of those physical causes of *M* are also causes of *P* — and, clearly, they may form a subset of a set of wholly physical causes which are collectively causally sufficient for the occurrence of *P*. Hence, claims (1) and (2) are not inconsistent with claim (3), but only with something much stronger, namely

(4) No physical event has a non-physical cause.

Obviously, however, no one can pretend that claim (4) is strongly confirmed by empirical evidence, however much it may be an article of faith with some philosophers. Even claim (3), although significantly weaker than claim (4), is not exactly strongly confirmed empirically. A presumption in its favour, however, is that modern science encourages us to believe that the universe is a causally closed system whose origins were wholly physical. At the time of the 'big bang', we suppose, all events were wholly physical — and all subsequent physical events have been and will continue to be long-term effects of those initial events. (Here I am assuming a thoroughgoing causal determinism, but not much is affected by assuming instead that a good deal of causation is irreducibly probabilistic.) But this presumption in no way rules out the possibility that, at some stage during the evolution of the universe, *non*-physical events or states have come into existence, along with subjects of those events or states (that is, selves, conceived of in accordance with claim (1)). There is no reason to disparage this idea as 'spooky', since it need involve no element of supernaturalism — taking 'supernaturalism' to be the view that some events are brought about by agents (such as a divine being) which do not exist within the space–time universe. (Such an agent would, of course, be a non-physical thing, quite unlike human selves as I conceive of them.)

Even if it is conceded that this is a genuine possibility and that claims (1), (2) and (3) are not logically inconsistent, it may nonetheless be thought that the suggestion that this is how things *actually are* is an extravagant one which somehow violates canons of parsimony or simplicity in matters metaphysical. On the contrary, I shall now attempt to show how the invocation of mental states, conceived of as non-physical causes of physical events, has the potential to strengthen our causal explanations of certain physical events. This is because such non-physical causes can be represented as rendering *non-coincidental* certain physical events which, from the perspective of purely physical causation, may appear to occur merely by coincidence.

VII: On Coincidental Events

An event *occurs by coincidence*, or *coincidentally*, in the sense I now have in mind, when two or more events co-occur and jointly cause that event, but those causes are themselves causally independent, in the sense of having no common cause amongst

their various causes. (Some philosophers describe the *co-occurrence* of two or more events which have no common cause as being a 'coincidence', and I have no quarrel with this usage: but my concern now is with the notion of a *single* event which occurs *by* coincidence, in the sense just explained. I am not concerned, then, with the question, which exercises some of those philosophers, of whether 'coincidences', in their sense, have causal explanations. See e.g. Owens, 1992, ch. 1, and Sorabji, 1980, ch. 1.)

Here is a familiar example. A man walks past a house just as a gust of wind dislodges a slate from the roof, causing it to fall, with the result that he is hit by it and killed. The man's walking there and the slate's falling there co-occur and jointly cause his death, but, we assume, there was no common cause of the man's walking there and the slate's falling there. Consequently, his death occurred by coincidence. But if, say, the man's approaching the house had set off a trip-wire attached to the slate, causing it to fall just as he passed underneath it, then his walking there and its falling there *would* have had a common cause and so his death would not have been coincidental.

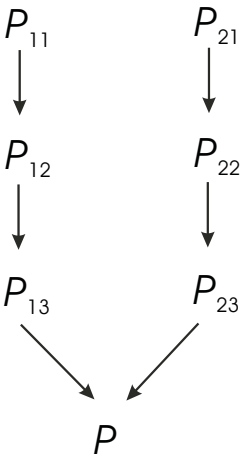
An event which occurs by coincidence is not an *uncaused* event: it has causes, which themselves have causes, which likewise have causes, and so on — what makes it coincidental is the fact that its immediate causes have independent causal histories. An event which does not occur by coincidence is one whose immediate causes share a common cause, rendering the causal histories of its immediate causes non-independent. At least, this will do, to a first approximation, as an account of the distinction between an event which occurs by coincidence and one which does not. (We might need to refine this account in order to avoid having to describe as 'non-coincidental' certain events whose immediate causes do share some common cause, but only a relatively insignificant one lying in the remote past of their respective causal histories. After all, we have already conceded that all current physical events are ultimately effects of events which occurred at the time of the 'big bang', but we don't want *this* to count as a reason for denying that certain current physical events are 'coincidental'.)

In our foregoing illustration of the distinction between coincidental and non-coincidental events, it is clear that the two different cases in which the man is killed by the falling slate differ, not only in respect of some of the physical events which occur and are causally responsible for the man's death in each case, but also in respect of some of the relations of physical causation which obtain between various physical events which occur in both cases. Thus, in the non-coincidental case, but not in the coincidental case, the physical event of the man's setting off the trip-wire occurs and is one of the causes of his death. And in the non-coincidental case, but not in the coincidental case, the physical event of the man's approaching the house is related by physical causation — via the movement of the trip-wire — to the physical event of the slate's falling. This is because the common cause which makes the difference between the coincidental and non-coincidental cases in our illustration is not only a physical event itself, but also one which links the causal histories of the immediate causes of the man's death by means of a chain of purely *physical* causation. But matters may be otherwise if what links the causal histories of the immediate physical causes of some non-coincidental physical event is a causal chain involving *non*-physical events, as I shall now demonstrate.

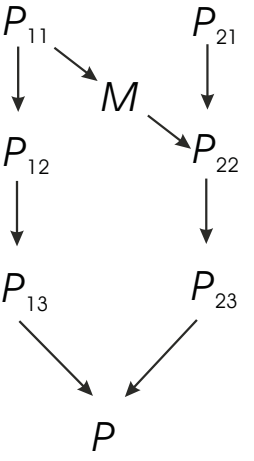
VIII: A Comparison Between Two Possible Worlds

Suppose that two independent causal chains of physical events, P_{11}, P_{12}, P_{13} and P_{21}, P_{22}, P_{23} jointly give rise to a physical event P as the immediate effect of P_{13} and P_{23}. Here the occurrence of P is coincidental. But, I submit, it is metaphysically possible for P to have (in a sense explained below) exactly the same *physical* causal history and yet *not* to occur by coincidence, because it is metaphysically possible for the immediate physical causes of P — namely P_{13} and P_{23} — to share a common cause which links them by a *non*-physical causal chain, thereby rendering their causal histories non-independent. It might be the case, for instance, that in this alternative scenario P_{11} is a cause of a mental event M which is in turn a cause of P_{22}: see the diagram below.

A note on how to read this diagram: each node, marked by a letter, represents a particular event and a line drawn between two nodes — whether or not it passes through other nodes — signifies that the event represented by the upper of those two nodes *is a cause of* the event represented by the lower of those two nodes. I should perhaps emphasise that to say that one event *is a cause of* another event is by no means to rule out the possibility that a third event, also, *is a cause of* that second event: that is to say, in the sense of 'cause' now in play, an event may have many different causes, without thereby being causally overdetermined. I am taking it that to say that one event *is a cause of* another event is — barring the possibility of causal overdetermination — at least to imply that if that first event had not occurred, then that second event would not have occurred either.



World 1



World 2

World 1 and world 2 are the same in the following respects: (i) the same physical events occur in both (in the space–time region with which we are concerned) and (ii) those events bear the same relations of purely physical causation to one another. By (ii) I mean that wherever two physical events in one of the worlds are linked by a certain chain of purely physical causation (causation not involving any non-physical event), they are linked in the same way in the other world. That is to say, wherever, in

one of the worlds, a certain physical event *is a cause of* another physical event, either directly or via certain other intervening physical events, those events stand in that same relation in the other world as well: in the other world, too, the first physical event *is a cause of* the second physical event, again either directly or via the same intervening physical events.[4] Of course — assuming that causation is law-governed — the two worlds are *not* the same in respect of the causal laws which obtain in them, because in world 2 certain psychophysical laws obtain which do not obtain in world 1. I shall return to this point in a moment. But what is of special interest to us now is that in world 1 the occurrence of P is *coincidental* whereas in world 2 it is *not* coincidental.

So we see that two worlds could contain the same physical events standing in the same relations of purely physical causation to one another and yet it be the case that in one of the worlds a certain physical event was coincidental whereas in the other it was not, because in the second world a certain non-physical (mental) event rendered the causal histories of that physical event's immediate physical causes non-independent. Incidentally, the existence of this possibility does much, in my view, to undermine the popular — though rather obscure — thesis that mental events 'supervene' upon physical events: if the thesis is taken to be that worlds which are the same in respect of what physical events occur in them and what relations of purely physical causation obtain between those events are worlds which are the same in respect of what mental events occur in them and what causal relations those events stand in, then we see, first, that this could at best be true of a restricted range of worlds and, second, that there is no positive reason to suppose that our world is one of those worlds. Of course, not all true counterfactual conditionals concerning physical events in one of our two depicted worlds could also be truths in the other of those worlds. For instance, in world 2 of our diagram — where P_{11} is depicted as being a cause of P_{22}, albeit only via the mental event M and not, thus, via any chain of purely physical causation — it is true that if P_{11} had not occurred, then P_{22} would not have occurred (barring causal overdetermination, which is not at issue here), whereas in world 1 it is *not* true that if P_{11} had not occurred, then P_{22} would not have occurred. But that just reflects the fact that different causal laws are operative in our two worlds and that different events (though not different *physical* events) occur in them. I should perhaps stress here, if it isn't sufficiently obvious already, that I am by no means suggesting that the situations depicted in the two worlds of our diagram are *compossible*: thus, the *actual* world could not simultaneously be *both* as depicted in world 1 *and* as depicted in world 2. However, it *is* implicit in what I have said that if one were to know, concerning the physical events depicted in our diagram, *only* which events they were and what relations of *purely physical* causation they bore to one another, one would not be in a position to decide on that basis whether the actual world was world 1 or world 2, for the simple reason that world 1 and world 2 do not differ in these respects. The significance of this fact will emerge in a moment.

---

[4] Since, manifestly, some events do not have exactly the same causes and effects in the two worlds as I have represented them, I am assuming that it is not an implication of any acceptable principle of transworld identity for events that an event has the same causes and effects in any world in which it occurs. But I take it that this is uncontroversial, since to say that it is metaphysically impossible for an event to have had causes and effects other than those which it actually has is to violate Hume's principle that there is no metaphysically necessary connection between cause and effect.

Another important point to observe concerning our worlds 1 and 2 is that *both* of them — not just world 1 — can be worlds in which claim (3) is true, that is, in which every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event. (Note here that the diagrams representing worlds 1 and 2 are not meant to be *complete* representations of those worlds, so that, for instance, it is not implied that P_{11} and P_{21} *lack* causes in those worlds.) For, as was pointed out earlier, it is perfectly possible that the non-physical event *M* of world 2 should itself have a set of wholly physical causes which are collectively causally sufficient for its occurrence. This can be so even if, as is plausible, *M* also has some other *non*-physical events amongst its causes (provided that each of these likewise has a set of wholly physical causes which are collectively causally sufficient for its occurrence).[5]

IX: The Significance of these Findings

What is the significance of these findings? Just this: they show that even if one has identified all the physical causes of a certain physical event within a certain space–time region — for instance, even if one has identified all the neural events causally responsible for a certain bodily movement — this doesn't preclude the possibility that the existence of a *non-physical* (mental) event or state, such as a person's belief or desire or intention, might serve to explain why that movement is *non-coincidental*, in a way in which the purely physical causal history of that movement does not. One might, thus, have truly discovered all the *physical* causes of the movement and this discovery might be consistent with the possibility that these were indeed *all* the causes of the movement — in which case the movement would be coincidental — and yet the discovery would also be consistent with the possibility that there were, in addition to these physical causes, certain *non*-physical causes which served to render the movement non-coincidental. So, merely to have satisfied ourselves that we have discovered all the physical causes of such a movement and to have satisfied ourselves that such causes *could* provide a complete causal explanation for the occurrence of the movement, is not yet to have ruled out the possibility that they *don't* in fact provide a complete causal explanation — because the question may still be left open as to whether or not the movement occurred by coincidence.

Positing certain non-physical (mental) causes of a physical event, in addition to the physical causes which have already been discovered, may serve an explanatory purpose which cannot be served by appeal to the physical causes alone. The non-physical part of the explanation need not deny anything which has been discovered about the identity of the physical causes and their purely physical causal relations and is in this

---

[5] Incidentally, it is important to distinguish both world 1 and world 2 from yet another possible world — call it world 3 — which is just like world 1 save that in it there is an additional relation of *purely physical* causation between event P_{11} and event P_{22}. In world 3, as in world 2, event *P* is not coincidental. But, of course, worlds 1 and 3 are *not* the same in *both* of the respects (i) and (ii) stated above — unlike worlds 1 and 2 — since worlds 1 and 3 differ from each other in respect (ii): the physical events in them do not bear exactly the same relations of purely physical causation to one another. In consequence, *P* does not have exactly the same *physical* causal history in worlds 1 and 3, whereas it does in worlds 1 and 2. Naturally, I have no wish to deny that non-coincidental events may *sometimes* have a purely physical causal explanation, as in world 3: I am only concerned to show that, and how, they may sometimes have a causal explanation which is at least partly *non*-physical and which is nonetheless consistent with the truth of claim (3).

sense perfectly compatible with the purely physical part of the explanation — though, as has been noted, accepting the non-physical part of the explanation *as well as* the physical part, rather than just accepting the physical part alone, will require adopting a different view as to what causal laws and what counterfactual conditional truths obtain.

Of course, if it should turn out, in a particular case, that a *physical* event can be discovered which renders a bodily movement (say) non-coincidental, then appeal to a non-physical event will be otiose in that particular case: the case will be like that of the non-coincidental death caused by the falling slate, where the triggering of a trip-wire rendered the death non-coincidental. But what if it were to be discovered that this can be done in *all* cases in which we now see reason to invoke items such as beliefs in our causal explanations? That, I suppose, might be taken by some to be a reason for holding that beliefs and so forth just *are* physical items (neural states or events, perhaps). However, my own view, on the contrary, is that it would instead be (at best) a reason for holding either that beliefs and so forth do not really exist at all (eliminativism) or else that they are causally inefficacious (epiphenomenalism). I take this view of the matter because, as I explained earlier, I do not consider that the thesis that mental states 'just *are*' physical states is even an intelligible one (though, to be fair, I don't regard either eliminativism or epiphenomenalism as being in much better shape, conceptually, so that I am pretty much committed to denying that we *could*, even in principle, make the discovery that has just been contemplated).

What I am chiefly concerned to point out for the moment, however, is that it is perfectly conceivable that we *should* discover, in the case of some bodily movement, that all of the physical (say, neural) events which we can implicate in its occurrence are such that, as far as their purely physical causal relations to one another are concerned, they do nothing to show that that movement is anything other than coincidental. And in such a case, my claim is, we could quite consistently and plausibly invoke a *non*-physical (mental) event as rendering that movement *non*-coincidental, without denying anything that had hitherto been claimed about the identities of that movement's physical causes and their purely physical causal relationships to one another and to the movement in question.

X: Intentionality and Mental Causation

So far I have said nothing about *how* mental events and states might cause physical events and states. For that matter, of course, neither have I said how *physical* events and states might cause physical events and states. But there is good reason to suppose that mental causation has some distinctive features which relate to the intrinsic natures of mental causes. We have been taking mental causes to be items such as beliefs, desires and intentions — in short, *intentional* states of the self. (The *onsets* of such states are events, but beliefs, desires and intentions are states rather than events — not that very much turns on the distinction between events and states in what follows.) Of course, some mental states — such as 'pure' sensations (if such there be) — are not intentional states, since they lack any intentional content (they are not 'about' anything, in the way that beliefs and desires are always 'about' something). But I am not really concerned with such non-intentional mental states at present. One distinctive feature of mental causation by intentional states is that *what* is caused by such

states is intimately related to the intentional content of those states. In the case of normal voluntary action, movements of the agent's body have amongst their causes intentional states of that agent which are 'about' just such movements. For instance, when I try to raise my arm and succeed in doing so, my arm goes up — and amongst the causes of its going up are such items as a desire of mine *that my arm should go up*. The intentional causes of physical events are always 'directed' upon the occurrence of just such events, at least where normal voluntary action is concerned (see further Lowe, 1996, Ch. 5). Nothing like this seems to be the case when physical events or states cause other physical events or states: such purely physical causation always appears to be 'undirected' or 'blind'.

Notice, however, that although, in normal voluntary action, an intentional state of the agent is 'directed' upon an event of the kind which it causes, it is not 'directed' upon the *particular* event which it causes. When I try to raise my arm and succeed in doing so, my desire that my arm should go up is amongst the causes of the event of my arm's going up: but my desire is not that *that particular event of arm-rising* should occur, but merely that *an* event of arm-rising of the appropriate kind should occur at a certain time, or during a certain interval of time. This has to be so because, even if I know that my attempt to raise my arm will succeed, I cannot know in advance *which* particular event of arm-rising will occur as a result of my success, since this will depend on factors outside my knowledge and control, such as the speed with which my nervous system reacts at the time of the attempt. Consequently, when my arm goes up as a result of my successfully trying to raise it, what is causally explained by my desire that my arm should go up is not merely the occurrence of this particular event of arm-rising, but the obtaining of the *general* state of affairs of an event of that kind's occurring during a certain interval of time — a state of affairs which happens to be 'realized' on this occasion by this particular event of arm-rising, but which could equally well have been 'realized' by a different particular event of arm-rising, provided it had been one of a suitable kind and had occurred at the right time.[6]

There is, I believe, a connection between this feature of intentional causal explanation and the already proposed role of mental causes in rendering certain of their physical effects non-coincidental. As we have seen, what qualifies an event as being 'non-coincidental' is a fact about the causal history of that event: the fact that its immediate causes have a common cause, that is, the fact that its immediate causes do not have independent causal histories. And I have suggested that when a mental state causes some physical event, its causal role may be one of rendering that event non-coincidental, which it can do by rendering non-independent the causal histories of that event's immediate physical causes. My further suggestion, now, is that this feature of the causal role of mental states is intimately related to the way in which they serve to provide causal explanations of certain *general* physical states of affairs and not merely of particular physical events. By causally connecting what would otherwise be independent chains of physical causation, I suggest, a mental cause can render the common effect of those chains non-coincidental and in so doing explain why an event *of that kind* occurred, not merely why *that particular event* occurred. For it

---

[6] Not all philosophers like to include 'states of affairs' in their ontology — but see Armstrong (1997) — and those who do are not necessarily in favour of including *general* states of affairs. But that is a debate for another occasion.

seems that the way in which a mental cause interconnects chains of physical causation is such as to ensure that the common effect of such chains is, in the following sense, *robust*: in all relatively 'close' possible worlds in which some of the physical events in those chains are different from those of the actual world but the interconnecting mental cause is still present, their common effect is nonetheless still *of the same kind* as that of the actual world — namely, the kind specified by the intentional content of that mental cause. This suggestion can perhaps best be elucidated by means of an example.

XI: An Illustrative Example

An assumption behind the following example will be that 'fusions' of events are, in general, themselves events: for instance, that if five battles occur over a certain period of time, then there is an event which occurs over that period of time and contains those five battles as parts (we might call this event a 'campaign', perhaps).[7] Most of the macroscopic events we normally talk about are event-fusions in this way. For example, even so 'simple' an event as the rising of a person's arm consists of many sub-events, such as the flexing of certain muscles and the movements of various parts of the arm.

Suppose that, over a period of several minutes, the following series of purely physical events is observed to occur: one after another, all of the coloured balls remaining on a snooker table are struck by the cue ball and as a result fall into pockets. The fusion of all these events of a snooker ball falling into a pocket is itself an event, call it event *E*. Suppose, now, we ask why event *E* occurred. Clearly, in one sense, *E* occurred because each of the sub-events of which *E* is the fusion occurred — but this is not a *causal* explanation of *E*. However, each of the sub-events — each event of a snooker ball falling into a pocket — has a causal explanation and one might suppose that the causal explanation of *E* is simply the conjunction of all those causal explanations (though we shall see in a moment that there could be good reason to challenge this supposition). Moreover, one might suppose that each event of a snooker ball falling into a pocket has a *wholly physical* causal explanation, adverting solely to prior physical events, such as movements of the snooker cue, movements of the player's hand (we are assuming here that there is just one player involved), neuronal events in the player's efferent nerves and motor cortex, and so forth. However, I suggest, if that *is* all there is to the explanation of event *E*, then we shall have to regard event *E* as having happened merely 'by coincidence'. Moreover, this explanation will not serve to explain, in any interesting sense, why an event *of this kind* occurred: we shall only be able to say that an event of this kind occurred 'because' this particular event occurred and was an event of this kind (and such a 'because' is not causal in force).

Now, of course, most of us would be extremely surprised if this were all there were to the explanation of event *E*. Most of us would surmise that *E* occurred because the snooker player had *formed and acted upon a desire* to pot all the coloured balls remaining on the table (and possessed the skill needed to achieve this). That desire would not be a desire specifically for event *E* to occur, but only a desire for the occurrence of an event of a certain kind, namely, an event consisting in the potting of all the remaining balls. We would surmise that, even if some of the balls had moved some-

---

[7] On the notion of an 'event-fusion', see Thomson (1977), pp. 78–9.

what differently from the ways in which they actually moved — or, indeed, even if there had been more or fewer balls remaining on the table — the player would have adjusted his action so as to ensure the same *kind* of result, even though event *E* itself would not have occurred in those circumstances. Citing the player's desire as an explanation of event *E*, then, explains not merely why *E* occurred, but, more interestingly, why an event of *that kind* occurred. And, it seems, no purely physical explanation of all of the sub-events of which *E* is the fusion can provide an interesting explanation of this sort. Such a purely physical explanation makes *E* appear to be a merely coincidental event and a 'fluke', in the sense that it provides us with no rational expectation that an event of *this kind* would still have occurred even if many of the individual movements of the balls had been rather different. Such a rational expectation can only be provided, it seems, by an explanation in terms of the player's *intentional state*. And such an explanation requires us to assign a *causal* role to that intentional state, a role which no purely physical state seems apt to occupy.

But what I have just said about the explanation of event *E* very often applies equally to the explanation of other event-fusions, such as the sort of event-fusion which constitutes an arm-rising. The advantage of focusing on event *E*, for our purposes, is simply that it is very much less likely, in its case, than it is in the case of an arm-rising, that such an event *could* be provided with a wholly physical causal explanation other than one which made it appear to be a mere 'fluke' which happened purely by coincidence.

XII: An Objection and a Reply

At this point I anticipate the following sort of objection. Surely, it may be said, someone could design a snooker-playing robot which could be pretty well relied upon to pot all of the snooker balls on a table — and when it did this there would occur an event of the same kind as *E*, which would be neither a 'fluke' nor 'coincidental': and yet, clearly, this event *would* have a purely physical causal explanation, since the robot would be a purely physical device possessing no mental states whatever, let alone a 'desire' to pot all of the snooker balls. So how can we be at all confident in denying that a wholly physical causal explanation of event *E* is available in the case of the *human* snooker-player?

My response is as follows. I agree that such a snooker-playing robot could, in principle, be designed and constructed. But note, first of all, that this is not to abandon appeal to *intentional causation* in our explanation of its feats, since we are quite explicitly appealing to the intentional states of the robot's designer and maker. It is not remotely plausible to suppose that a device like this could come into being *completely without* any causal contribution from the intentional states of any thinking being. Secondly, note that, in describing the workings of the robot, we have conceded that *it* possesses no intentional states whatever, such as a desire to pot all of the snooker balls on a table. Indeed, the force of the objection that has been raised — that the robot provides an example of how an event of the same kind as event *E* could have a purely physical causal explanation — rests upon the presumption that the robot does *not* have any intentional states. But then it follows that if we are to see the example of the robot as providing an alternative paradigm for the explanation of event *E* in the case of the *human* snooker-player, we have to regard that paradigm as an *eliminativist*

one, in which appeal to the player's 'desire' to pot all of the balls has no genuine explanatory force.

Now, I concede that it is *possible* for every physical action which is performed by a human being to be caused in a wholly 'robotic' way, without any causal contribution from intentional states. But the point is that this possibility is an *extremely remote* one: we have no good reason whatever to suppose that it has been realized in this, the *actual* world. On the contrary, we have every reason to think that people's beliefs and desires *do* contribute causally to their physical behaviour and help to explain it. What we *cannot* do, however, is try to combine this conviction with the thought that, somehow, human behaviour does, in principle, have a purely physical causal explanation along the lines of robotic behaviour, in the hope of reconciling intentional causation with a thoroughgoing physicalism. We can either take intentional causation seriously, in which case we must abandon physicalism, or else we can cleave to physicalism, in which case we must be eliminativists (or, perhaps, epiphenomenalists) about the mental. There is no middle ground which allows us to have it both ways. However, as I have tried to show, we *can* espouse a version of 'dualism' (for want of a better word) which preserves one central tenet of physicalism, namely, claim (3): that *every physical event has a set of wholly physical causes which are collectively causally sufficient for the occurrence of that event*. If it is only a concern that this claim is denied by dualism which persuades some philosophers to reject dualism in favour of physicalism, then I hope I have shown them why that concern is quite misplaced. If, on the other hand, their physicalism is motivated by a faith in claim (4) — that *no physical event has a non-physical cause* — then I can only say that it seems to me that their doctrine is an unwarranted dogma which commits them, whether they like it or not, to eliminativism or epiphenomenalism regarding the mental. The kind of 'dualism' that I am defending is fully deserving of the title 'naturalistic' — provided that the term 'naturalism' is not hijacked as a mere synonym for 'physicalism', but is accorded its proper meaning as denoting a repudiation of the *super*natural.

Acknowledgements

Many thanks to John Heil for making me think about many of these matters, though we disagree about a lot of them. Thanks to him also, to Jim Edwards, and to members of audiences at the Universities of Durham and Stirling for their comments on an earlier draft of the paper. I am also grateful to an anonymous referee.

References

Armstrong, D.M. (1997), *A World of States of Affairs* (Cambridge: Cambridge University Press).

Geach, P.T. (1979), *Truth, Love and Immortality: An Introduction to McTaggart's Philosophy* (London: Hutchinson).

Lowe, E.J. (1989), *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms* (Oxford: Blackwell).

Lowe, E.J. (1996*), Subjects of Experience* (Cambridge: Cambridge University Press).

McGinn, C. (1995), 'Consciousness and space', *Journal of Consciousness Studies*, **2** (3), pp. 220–30.

Owens, D. (1992*), Causes and Coincidences* (Cambridge: Cambridge University Press).

Papineau, D. (1993), *Philosophical Naturalism* (Oxford: Blackwell).

Sorabji, R. (1980*), Necessity, Cause and Blame: Perspectives on Aristotle's Theory* (London: Duckworth).

Strawson, G. (1994), *Mental Reality* (Cambridge, MA: MIT Press).

Thomson, J.J. (1977), *Acts and Other Events* (Ithaca, NY: Cornell University Press).

Von Wright, G.H. (1971), *Explanation and Understanding* (London: Routledge and Kegan Paul).