

Non-Cartesian Substance Dualism and the Problem of Mental Causation

Author(s): E. J. Lowe

Source: *Erkenntnis* (1975-), Vol. 65, No. 1, Prospects for Dualism: Interdisciplinary Perspectives (2006), pp. 5-23

Published by: [Springer](#)

Stable URL: <http://www.jstor.org/stable/27667849>

Accessed: 06-03-2016 20:53 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Erkenntnis* (1975-).

<http://www.jstor.org>

E. J. LOWE

NON-CARTESIAN SUBSTANCE DUALISM AND THE PROBLEM OF MENTAL CAUSATION

ABSTRACT. Non-Cartesian substance dualism (NCSD) maintains that *persons* or *selves* are distinct from their organic physical bodies and any parts of those bodies. It regards persons as ‘substances’ in their own right, but does not maintain that persons are necessarily separable from their bodies, in the sense of being capable of disembodied existence. In this paper, it is urged that NCSD is better equipped than either Cartesian dualism or standard forms of physicalism to explain the possibility of mental causation. A model of mental causation adopting the NCSD perspective is proposed which, it is argued, is consistent with all that is currently known about the operations of the human central nervous system, including the brain. Physicalism, by contrast, seems ill-equipped to explain the distinctively intentional or teleological character of mental causation, because it effectively reduces all such causation to ‘blind’ physical causation at a neurological level.

1 INTRODUCTION

Non-Cartesian substance dualism (NCSD) maintains that *persons* or *selves* – that is to say, self-conscious subjects of experience and agents of intentional actions – are distinct from their organic physical bodies and any parts of those bodies, such as their brains or central nervous systems. It regards persons as substances in their own right, in the sense of ‘substance’ in which this denotes a persisting entity and bearer of properties which does not depend for its identity on anything other than itself.¹ However, NCSD does not maintain that persons are necessarily separable from their bodies, in the sense of being capable of disembodied existence. It allows, indeed, that persons are themselves physical beings, in the sense that they genuinely possess physical characteristics, such as shape, height, and weight. An early proponent of this sort of view was Strawson, although he himself would almost certainly want to resist using the term ‘dualist’ in this context.²

NCSD may be defended on a number of grounds, one of which is that it is better equipped than either Cartesian dualism or standard

forms of physicalism to explain the possibility of mental causation. Cartesian dualism, by insisting that the self or soul, as a purely mental substance, possesses no physical characteristics and hence lacks spatial location, is apparently faced with a difficulty of explaining how any soul is causally related to its 'own' body in particular, as Jaegwon Kim has recently urged.³ Physicalism, on the other hand, seems ill-equipped to explain the distinctively intentional or teleological character of mental causation, because it effectively reduces all such causation to 'blind' physical causation at a neurological level. NCSA, recognizing as it does both the physical aspects of the self and the autonomous nature of mental causation, is well-positioned to avoid both of these failings. In the present paper, a model of mental causation adopting the NCSA perspective will be proposed and I shall argue that it is consistent with all that is currently known about the operations of the human central nervous system, including the brain.

2 CARTESIAN DUALISM

There are many reasons for doubting the identity of the human *self* with the human *body*, or any part of it, such as the brain, some of which I shall discuss shortly. But, if we deny any such identity, how can we account for the apparent fact that we have *causal control* over parts of our bodies – in short, how can we accommodate the possibility of *mental causation*? In the remainder of this paper, I shall first look at some reasons for denying self–body identity – that is, for accepting a form of *substance dualism* – and then at some ways of understanding how the notion of mental causation might be rendered consistent with such a view of the self.

René Descartes, of course, was the most famous of all substance dualists.⁴ On his version of the doctrine, the *mind* (or *self*, or *soul*) and the *body* are two distinct and separable substances, which exist together in a 'substantial union'. (Note that by a 'substance' in this context Descartes meant an individual being that *exists independently* of other such beings – save, of course, God – and *bears properties*, which may change over time. There is no implication, however, in Descartes's conception of substance that a substance must consist of 'stuff' of any kind, so that it would be a gross caricature of his position to allege that he regarded the soul as being composed of some sort of 'spiritual matter', such as the 'ectoplasm' postulated by 19th-century spiritualists.) Now, each of these two substances has,

according to Descartes, a principal attribute which the other lacks – *thought*, or consciousness, in the case of the mind, and spatial *extension* in the case of the body – and all of its properties are modes (that is, particular determinations) of that attribute. Thus, modes of thought include particular beliefs, desires, and volitions, whereas modes of extension include particular shapes, sizes, and velocities. So, according to Descartes, whereas the mind has beliefs, desires, and volitions, but no shape, size, or velocity, the body has shape, size, and velocity, but no beliefs, desires, or volitions.

Descartes has two main arguments in favour of this view. First, an argument from the *conceivability* of the separate existence of mind and body, and second an argument from the supposed *indivisibility*, or simplicity, of the soul, which he contrasts with the (more evident) divisibility of the body. Let us look briefly at those arguments.

The *conceivability argument* may be reconstructed as follows.

- (1) It is *conceivable* that I should exist without a body.
- (2) What is conceivable is *possible*, since at least God can bring it about.
- (3) So, it is *possible* that I should exist without a body.
- (4) If it is possible that I should exist without a body, then I and my body must be distinct and separable substances.
- (5) Therefore, I and my body are distinct and separable substances.

This argument is open to challenge in various places, notably at step (2) and at step (4). Step (2) presumes that conceivability is a reliable guide to possibility – that is, to real or *metaphysical* possibility. Step (4) presumes that my body is *essentially* a body or, in other words, that in every possible world in which my body exists, it is a body (for if my body is *not* essentially a body, then it is possible that I *am* my body, and the possible worlds in which I exist without a body are simply those in which my body exists but is not a body).⁵ That, however, is a very plausible presumption, so the main difficulty lies with step (2), the supposition that what is conceivable is really possible.⁶ But step (1) is also controversial, especially if it is supposed to imply – via step (2) – step (3). The argument as a whole is, I think, simply too questionable to carry much persuasive force, and that certainly seems to be the verdict of most philosophers since Descartes's time.

The *divisibility argument* is rather simpler, and may be reconstructed in the following fashion.

- (1) The self or soul lacks any parts into which it is divisible.

- (2) The body, being spatially extended, is divisible and so composed of parts.
- (3) Hence, the self and the body are distinct substances and the self is, unlike the body, unextended.

(Note that, whereas the conceivability argument delivers the conclusion that the self and the body are *distinct and separable substances*, the divisibility argument delivers the conclusion that the self and the body are *distinct substances and only the latter is extended*.) Here, it is premise (1) that is most likely to be challenged, but it may also be questioned whether, as (2) presumes, being extended is logically sufficient for being divisible into parts. For it is possible to maintain that even a *simple* substance may, in principle, be spatially extended.⁷ I should perhaps reveal at this stage that my own view is that the self is indeed a simple, but spatially extended substance: so that, in fact, I *agree* with the two premises of the divisibility argument (setting aside the presumption, implicit in (2), that being extended implies being divisible). However, that does not mean that I think that the argument is a good one. Indeed, I could not myself employ that argument without circularity, since part of my reason for maintaining that the self is a simple substance is precisely that I believe it to be distinct from the body.⁸ But this is already to presuppose something that is part of the conclusion, (3), of the divisibility argument. My own reasons for holding the self to be distinct from the body (or any part of it) will be outlined shortly.

To sum up: I think that it is clear enough that Descartes's two arguments for his version of substance dualism are simply not sufficiently compelling to warrant belief in his doctrine, so I shall say no more about them here.

3 NON-CARTESIAN SUBSTANCE DUALISM

Let me now explain more fully what I mean by *non-Cartesian substance dualism* – NCS D. By this I mean a position which holds, with Descartes, that the self is *distinct* from the body or any part of it, but does not insist either that the self is *separable* from anything bodily or that it is spatially *unextended*. It allows, that is, that the self may not be able to exist without a body and that it may be extended in space, thus possessing spatial properties such as shape, size, and spatial location. It may also allow – indeed, I think that it should insist – that the self is *simple*, or not composed of parts, although I shall not dwell

on that point now. One way to think of the self-body relation – the relation of *embodiment* – according to NCSD is by analogy with the relation between a bronze statue and the lump of bronze composing it at any given time.⁹ The statue and the lump are, very plausibly, not *identical*, because each could continue to exist in the absence of the other (for example, the lump could survive squashing into a thin disc, but the statue could not, whereas the statue could survive the replacement of one of its arms, but would then be composed of a different lump of matter). Even so, the statue clearly does need to be composed of *something* material: it could not exist in an entirely ‘disembodied’ state. Of course, if the self really is *simple*, as I believe, then this analogy is imperfect, but it still suffices to let us see how two individual ‘substances’ may be *distinct* – that is, non-identical – and yet be so intimately related that they exactly coincide spatially at a given time and necessarily share, at that time, many of their physical properties, such as their shape, size, and mass.

What can be said in favour of NCSD, and why should it be preferred to Descartes’s version of substance dualism? As for the second question, it may be urged that NCSD is a less extreme and intuitively more plausible doctrine – less extreme because it is not committed to the possibility of disembodied existence and more plausible because it respects our intuition that we *ourselves*, not just *our bodies*, occupy space and have properties of shape, size, mass, and spatial location. One may also point out, as I have already done, that the standard arguments for Cartesian dualism are not very compelling. As for the first question, however, I answer as follows. It seems clear, upon reflection, that the *identity-conditions* of selves and bodies – including parts of bodies – differ radically, quite as much as do those of statues and lumps of bronze. (By the ‘identity-conditions’ of entities of any given kind, *K*, I simply mean the conditions logically necessary and sufficient for the truth any identity-statement of the form ‘*a* is the same *K* as *b*’, where ‘*a*’ and ‘*b*’ are names for entities of kind *K*.¹⁰) In evidence of this, it is very plausible to suppose, for example, that I could survive the gradual replacement of every cell in my body by inorganic parts of appropriate kinds, so that I would end up possessing a wholly ‘bionic’ body, distinct in all of its parts from my existing biological body.¹¹ The implication, of course, is that I cannot have the same identity-conditions as those of my existing biological body or any part of it, because I can survive a change which it and all of its parts cannot survive – and consequently I cannot be *identical* with that body or any part of it.

Another – and in my view even more compelling – consideration is the following. It does not appear that either my whole body, or any particular part of it, can qualify – in the way that *I* do – as the unique subject of all of *my* thoughts and other conscious mental states, for no *one* bodily entity is necessary for the existence of *all* of those mental states of mine, even if each of them depends for its existence upon *some* bodily entity. Suppose, for instance, that it were proposed that I am identical with *my brain*, as many physicalists believe. Now, even if it is granted that each of my conscious thoughts and feelings depends on some particular neuronal activity going on in some part of my brain and could not exist in the absence of such activity, it seems clear that there is no *one* part of my brain such that *all* of my conscious thoughts and feelings depend upon neuronal activity going on precisely *there*. Nor, of course, need *all* of my brain exist in order for me to enjoy each and every one of my conscious thoughts and feelings, even if it is true that without a brain I would enjoy no conscious thoughts and feelings whatsoever. For, clearly, I could still have many conscious thoughts and feelings even if my brain were reduced in various different ways by the loss or destruction of various different parts of it. But all of this means that the relationship between my conscious thoughts and feelings and my brain or any particular part of it is very different from the relationship between my conscious thoughts and feelings and *me*, as their subject. For it seems clear that each and every one of these thoughts and feelings requires *me* to exist – in other words, *could not exist without me* – since, after all, none of them could exist as *someone else's* conscious thoughts and feelings, and still less as *no one else's*.¹² By contrast, as we have just seen, it is *not* true of each and every one of my conscious thoughts and feelings that it could not exist without *the whole* of my brain existing, since many of them could still exist even if it were reduced in various ways. Nor, as we have also seen, is there is any specific *part* of my brain such that it is true of each and every one of my conscious thoughts and feelings that it could not exist without *that part* existing. Hence, I conclude that I cannot be identical either with my brain as a whole or with any specific part of it. In my view, the ultimate reason for this is that I, as a self or subject of experience, possess a strong kind of *unity*, in virtue of being a *simple* substance – a kind of unity that is not possessed by my brain, composed as it is of millions of distinct and separable material parts (all of the neurons and other cells that collectively make it up).

I simply cannot see any remotely plausible way for a physicalist to respond to the foregoing argument. I think, indeed, that such a

philosopher is ultimately committed to denying the very existence of the self or 'I', as a subject of experience. Of course, some philosophers have bravely adopted this position, notably David Hume¹³ – though even he conceded that doing so is intolerably paradoxical. I prefer to see in it the *reductio ad absurdum* of physicalism.

4 THE PROBLEM OF MENTAL CAUSATION

But let me turn now to *the problem of mental causation*. This has always been thought to present a particular difficulty for Cartesian dualism – but is that really so, and if so, does it equally present a difficulty for NCS? Let us look at Cartesian dualism first. Here it is often complained that it is completely mysterious how an unextended, non-physical substance could have any causal impact upon the body – the presumption being, perhaps, that any cause of a physical event must either be located where that event is, or at least be related to it by a chain of events connecting the location of the cause to the location of the effect.¹⁴ This presumption is closely related to another, namely, that the domain of physical events is *causally closed*, in the sense that no chain of causation can lead backwards from a purely physical effect to antecedent causes some of which are *non-physical* in character.¹⁵ It may certainly seem that Cartesian dualism cannot meet this sort of objection, if the presumptions just mentioned are correct – because this form of dualism seems to imply that the mental causes of bodily effects, being changes in a supposedly unextended soul, could have no spatial location. Be that as it may, NCS, in contrast, does *not* insist that the self is unextended and lacking in spatial properties, and hence it is not nearly so apparent that the preceding sort of objection applies to it. In any case, we need to think afresh about mental causation in order to see whether, and if so how, NCS can accommodate it.

5 NCS AND THE NON-IDENTITY OF MENTAL AND PHYSICAL CAUSES

What seems plausible is that if we were to trace the purely *bodily* causes of any peripheral bodily event, such as the movement of my arm on a given occasion, backwards in time indefinitely far, we would find that those causes *ramify*, like the branches of a tree, into a complex maze of antecedent events in my nervous system and brain – these neural events being widely distributed across large areas of

those parts of my body and having no single focus anywhere, the causal chains to which they belong possessing, moreover, no distinct *beginnings*.¹⁶ See Figure 1 below.¹⁷ And yet, my mental act of *decision* or *choice* to move my arm seems, from an introspective point of view, to be a *singular* and *unitary* occurrence which somehow *initiated* my action of raising my arm. How, if at all, can we reconcile these two apparent facts? It seems impossible to *identify* my act of choice with any individual neural event, nor even with any combination of individual neural events, because it and they seem to have such different causal features or profiles. The act of choice seems to be unitary and to have, all by itself, an ‘initiating’ role, whereas the neural events seem to be thoroughly *disunified* and merely to contribute in different ways to a host of different ongoing causal chains, many of which lead independently of one another to the eventual arm-movement.

NCSD can, I believe, enable us to see how *both* of these causal perspectives on physical action can be correct, without one being reducible to the other and without any sort of rivalry between the two. The act of choice is attributable to the *person* or *self* – to *me*, in this case – whereas the neural events are attributable to parts of the *body*: and self and body are *distinct* things, even if they are not *separable* things. Moreover, the act of choice *causally explains* the bodily movement – the movement of my arm – in a different way from the way in which the neural events explain it. The neural events explain why the arm moved *in the particular way* that it did – at such-and-such a speed and in such-and-such a direction at a certain precise time. By contrast, the act of choice explains why a movement *of that general kind* – in this case, a rising of my arm – occurred around about the time that it did. It did so because shortly beforehand I decided to raise my arm. My decision certainly did not determine the precise speed, direction, and timing of my arm’s movement, only *that* a movement of that general sort would occur around about then. The difference between the two kinds of causal explanation reveals itself

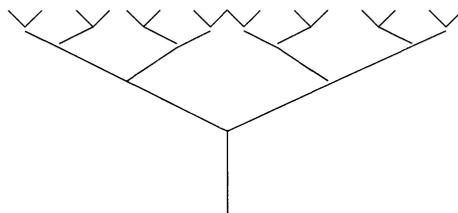


Figure 1.

clearly when one contemplates their respective *counterfactual* implications. If I had not decided to raise my arm, there wouldn't have been an arm-movement of that kind *at all* – my arm would either have remained at rest or, if I had decided to make another movement instead, it would have moved in a quite different way. It doesn't seem, however, that one can isolate any neural event, or any set of neural events, whose non-occurrence would have had *exactly the same consequences* as the non-occurrence of my decision.¹⁸ Rather, the most that one can say is that if this or that neural event, or set of neural events, had not occurred, the arm-movement might have proceeded in a somewhat different manner – more jerkily, perhaps, or more quickly – *not* that my arm would have remained at rest, or would instead have moved in a quite different kind of way.

6 COUNTERFACTUALS, CAUSATION, AND THE NON-IDENTITY THESIS

This last point is an extremely important one and warrants further elucidation. It is now standard practice amongst philosophers of logic and language to interpret counterfactual conditionals in terms of possible worlds, as follows.¹⁹ A counterfactual of the form 'If it were the case that *p*, then it would be the case that *q*' is said to be true if and only if, in the *closest* possible world in which *p* is the case, *q* is also the case – where the 'closest' possible world in question is the one in which *p* is the case but otherwise *differs minimally* from the actual world. Now, suppose that a physicalist were to propose that my decision, *D*, to raise my arm on a given occasion – my mental act of choice – is identical with a certain neural event, *N*, which is correctly identifiable as a *cause* of the subsequent bodily event, *B*, of my arm's rising. (Here I must stress that *D*, *N*, and *B* are, each of them, *particular events*, each occurring at a particular moment of time, with *B* occurring at least an appreciable fraction of a second later than *D* and *N*, since our decisions to act do not take effect immediately – and the physicalist must suppose, of course, that *D* and *N* occur at the *same* time, since he holds them to be identical.²⁰) Let us concede, consequently, that the following counterfactual is true: 'If *N* had not occurred, then *B* would not have occurred'.²¹ What I am interested in is the following question: what sort of event *would* have occurred, instead of *B*, if *N* had not occurred? In other words: in the closest possible world in which *N* does not occur, what sort of event occurs instead of *B*? My contention is that what occurs in this world is an

event of the same sort as *B*, differing from *B* only very slightly. The reason for this is as follows.

It seems evident, from what we know about the neural causes of an event such as *B*, that *N* must be an *immensely complex* neural event: it must be, in fact, the sum (or ‘fusion’) of a very large number of individual neural events, each of them consisting in some particular neuron’s firing in a particular way. (Recall, here, that *N* must be supposed to occur an appreciable amount of time *before B*, at a time at which the neural antecedents of *B* are many and quite widely distributed across my cerebral cortex.) It would be utterly implausible for the physicalist to maintain, for example, that my decision *D* is identical with the firing of just a *single* neuron, or even of a small number of neurons. If *D* is identical with any neural event *at all*, it can surely only be identical with an extremely complex one, consisting in the firing of many neurons distributed over quite a large region of my cerebral cortex. However, it seems indisputable that if *N* is, thus, the sum of a very large number of individual neural events, then the *closest* world in which *N* itself does not occur is a world in which *another* highly complex neural event, *N**, occurs, differing *only very slightly* from *N* in respect of the individual neural events of which it is the sum. In other words, *N** will consist of *almost exactly the same* individual neural events as *N*, plus or minus one or two. Any possible world in which a neural event occurs that differs from *N* in *more* than this minimal way simply will not qualify as the *closest* possible world in which *N* does not occur. This is evidently what the standard semantics for counterfactuals requires us to say in this case. But, given what we know about the functioning of the brain and nervous system, it seems clear that, in the possible world in which *N** occurs, it causes a bodily event *very similar* to *B*, because such a small difference between *N* and *N** in respect of the individual neural events of which they are respectively the sums cannot be expected to make a very big difference between their bodily effects. There is, we know, a good deal of redundancy in the functioning of neural systems, so that the failure to fire of one or two motor neurons, or the abnormal firing of one or two others, will typically make at most only a minimal difference with regard to the peripheral bodily behaviour than ensues. Thus, the answer to the question posed earlier – what sort of bodily event would have occurred instead of *B*, if *N* had not occurred? – is this: a bodily event *very similar* to *B*. In other words, if *N* had not occurred, *my arm would still have risen in almost exactly the same way as it actually did*.

Now, I hope, we can see the importance of this conclusion. For, if we ask what sort of bodily event would have occurred instead of *B* if *my decision, D*, to raise my arm had not occurred, then we plausibly get a very different answer. Very plausibly, if *D* had not occurred – if I had not made the very act of choice that I did to raise my arm – then my arm *would not have risen at all*. It is, I suggest, quite incredible to suppose that if I had not made *that* very decision, *D*, I would have made a decision virtually indistinguishable from *D* – in other words, *another* decision to raise my arm in the same, or virtually the same, way. On the contrary, if I had not made *that* decision, then I would either have made a quite different decision or else no decision at all. Either way – assuming that there is nothing defective in my nervous system – my arm *would not* have risen almost exactly as it did.

If all of this reasoning is correct, then it follows unavoidably that my decision *D* cannot be identical with the neural event *N* with which the physicalist proposes to identify it, for the counterfactual implications of the non-occurrence of these two events are quite different. If *D* had not occurred, my arm would not have risen at all, but if *N* had not occurred, it would have risen almost exactly as it did. The ultimate reason for this parallels the reason given earlier for the non-identity of the self with the brain or any part of it. It is that a mental act of choice or decision is, in a strong sense, a *singular* and *unitary* event, unlike a highly complex sum or fusion of independent neural events, such as *N*. *N** differs from *N* only in excluding one or two of the individual neural events composing *N* or including one or two others. That is why *N* and *N** can be so similar and thus have such similar effects. But *D* cannot intelligibly be thought of, in like manner, as being *composed* of myriads of little events and that is why, in the closest possible world in which *D* itself does not occur, there does *not* occur another decision *D** which differs from *D* as little as *N** differs from *N*.

7 THE DISTINCTIVE CHARACTER OF MENTAL CAUSATION

So far, I have tried to explain why the mental and neural causes of voluntary bodily movements must be distinct, consistently with allowing, as I do, that such movements have *both* mental *and* neural causes. Now I want to say a little more about the respects in which mental causation is distinctively different from bodily or physical causation. Most importantly, then, mental causation is *intentional* causation – it is the causation of an *intended* effect of *a certain kind*.

Bodily causation is not like this. All physical causation is ‘blind’, in the sense that physical causes are not ‘directed towards’ their effects in the way that mental causes are. *Both* sorts of causation need to be invoked in order to give a full explanation of human action and NCSD seems best equipped to accommodate this fact. The very *logic* of intentional causation differs from the *logic* of bodily causation. Intentional causation is *fact* causation, while bodily causation is *event* causation.²² That is to say, a choice or decision to move one’s body in a certain way is causally responsible for the *fact* that a bodily movement of *a certain kind* occurs, whereas a neural event, or set of neural events, is causally responsible for a *particular* bodily movement, which is a particular *event*. The decision, unlike the neural event, doesn’t causally explain why that *particular* bodily movement occurs, not least because one cannot *intend* to cause a particular event, only to bring it about that an event of a certain kind occurs. (One can only intend something if one can make it *an object of thought*: but I cannot make an *as yet non-existent* future event the object of my thought – I can at most think of the future as including an event of a certain kind, such as a rising of my arm.)

As I have just remarked, the two species of explanation, mental and physical, are both required and are mutually complementary, for the following reason. Merely to know why a *particular* event of a certain kind occurred is not necessarily yet to know why an event of *that* kind occurred, as opposed to an event of some other kind. Intentional causation can provide the latter type of explanation in cases in which bodily causation cannot. More specifically: an event, such as a particular bodily movement, which may appear to be merely *coincidental* from a purely *physiological* point of view – inasmuch as it is the upshot of a host of independent neural events preceding it – will by no means appear to be merely coincidental from an *intentional* point of view, since it was an event of *a kind* that the agent intended to produce.²³

Notice, here, that the foregoing fact – that a mental decision, *D*, to perform a certain kind of bodily movement, cannot be said to cause the *particular* bodily event, *B*, of that kind whose occurrence renders that decision successful – is already implied by the arguments of the preceding section of this paper. For, given that *D* is *not identical* with the actual neural cause, *N*, of *B*, the closest possible world in which *N* does not occur *is still a world in which D occurs* – but in that world a slightly different bodily movement, *B**, ensues, being caused there by a slightly different neural cause, *N**. (Clearly, if *D* is not identical with *N*, then there is no reason to suppose that the closest world in

which *N* does not occur is also one in which *D* does not occur, for a world in which *both* of these events do not occur evidently differs more from the actual world than a world in which just *one* of them does not occur, other things being equal.) However, this means that the occurrence of *D* is compatible with the occurrence of two *numerically different* bodily movements of the same kind, *B* and *B**, and hence does not causally determine *which* of these occurs, but only that *some* bodily movement of their kind occurs.

At this point, I anticipate the following possible objection on the part of the physicalist. Couldn't the physicalist simply *concede* that the complex neural event *N*, in our example, is not identical with the mental decision *D* – and thereby concede that *D* does not cause the *particular* bodily movement, *B*, that is caused by *N* – while still insisting that *D* is identical with *some* neural event, call it *M*, which has precisely the causal role that I am attributing to *D*? According to this view, *D* is identical with a neural event, *M*, which causally explains why *some* bodily movement of *B*'s kind occurred, but not why *B* in particular occurred. No – such a position is not tenable, for reasons which we have already encountered. Recall that I argued that the following counterfactual conditional is true: 'If *D* had not occurred, then *no* bodily movement of *B*'s kind would have occurred'. That is to say, if I had not performed that decision to raise my arm, my arm would not have risen in anything like the way that it did – it would either have moved in some quite different way, or not at all, because if I had not made *that* decision, I would either have decided to do something quite different or else not have decided to do anything. Can the same thing be said with regard to the putative neural event *M*? No, it can't. This is because, once again, plausibility demands that the physicalist takes *M* to be an *extremely complex* neural event, composed of the firings of very many individual neurons, so that the closest possible world in which *M* itself does not occur will be one in which a neural event, *M**, occurs which differs from *M* only in respect of the firing of one or two individual neurons. And it simply isn't credible to suppose that this very small difference between *M* and *M** should make all the difference between my arm rising and some quite different kind of bodily movement occurring. Consequently, the counterfactual conditional that is true of *M* is this: 'If *M* had not occurred, then a bodily movement of *B*'s kind would still have occurred'. So, once more, because different counterfactuals are true of *D* and *M*, *D* and *M* cannot be identical. The physicalist's new proposal encounters exactly the same difficulty as did his original proposal. The difficulty is that mental causes like *D* have a *strong*

unity which fails to characterize extremely complex neural events such as *N* and *M*. Because of this lack of strong unity, the closest worlds in which events like *N* and *M* do not occur are worlds in which the vast majority of *their parts* still occur, with the consequence that similar bodily effects still ensue.

Much more can and should be said on these matters, but since I have discussed many of them extensively elsewhere, I shall rest content with the foregoing remarks for present purposes.²⁴

8 CAUSAL CLOSURE AND CAUSAL DETERMINISM

Here it may be asked: *But what about the causes of my acts of decision or choice? Are these* bodily, or mental, or both? My own opinion is that an act of decision or choice is *free*, in the ‘libertarian’ sense – that is to say, it is *uncaused*.²⁵ This is not to say that decisions are simply *inexplicable*, only that they demand explanations of a non-causal sort. Decisions are explicable in terms of *reasons*, not causes. That is to say, if we want to know why an agent *decided* to act as he did, we need to inquire into *the reasons in the light of which* he chose so to act.²⁶ Since decisions are, according to NCSD, attributable to the *self* and not to the body or any part of it, there is no implication here that any *bodily* event is uncaused.

But now it may be wondered: How is it really possible for mental acts of decision to explain anything in the physical domain, if that domain is *causally closed*, in the sense defined earlier? Let us recall how, precisely, we defined the causal closure of the physical domain. According to the thesis of physical causal closure, I said, no chain of causation can lead backwards from a purely physical effect to antecedent causes some of which are *non-physical* in character. But intentional causation on the NCSD model, as I have described it, does not violate the thesis of physical causal closure, since it does not postulate that mental acts of decision or choice are events *mediating between bodily events* in chains of causation leading to purely physical effects: it does not postulate that there are ‘gaps’ in chains of physical causation that are ‘filled’ by mental events.²⁷ As we have seen, on the NCSD model, a decision can explain the fact that a bodily movement *of a certain kind* occurred on a given occasion, but not the *particular* movement that occurred.²⁸

Even so, it may be protested, if physical causation is *deterministic*, then there is really no scope for intentional causation on the NCSD model to explain anything physical, because the relevant

counterfactuals will all be *false*. It will be *false*, for instance, to say that if I had not decided to raise my arm, a rising of my arm would not have occurred: rather, precisely the same bodily movement *would* still have occurred, caused by precisely the same physical events that actually did cause it – for if physical determinism is true, there was never any real possibility that those physical events should not have occurred, nor that they should have had different effects. Maybe so. But, in view of the developments in quantum physics during the 20th century, we now know that physical causation is *not* in fact deterministic, so the objection is an idle one and can safely be ignored.

The NCS model of intentional causation may nonetheless seem puzzling to many philosophers. I suggest that that is because they are still in the grip of an unduly simple conception of causation – one which admits only of the causation of one event by one or more antecedent events belonging to one or more chains of causation which stretch back indefinitely far in time. Since this is the only sort of causation recognized by the physical sciences, intentional causation on the NCS model is bound to be *invisible* from the perspective of such a science.²⁹ To a physicalist, this invisibility will seem like a reason to dismiss the notion of intentional causation as spurious, because ‘non-scientific’. To more broad-minded philosophers, I hope, it will seem more like a reason to perceive no genuine conflict between explanation in the physical sciences and another, more humanistic way of explaining our intentional actions, by reference to our choices or decisions and the reasons for which we make them.

ACKNOWLEDGMENTS

I am grateful for comments received when an earlier version of this paper was delivered as a special lecture at the University of München in July 2004. I am also grateful to two referees for their comments on and criticisms of a previous draft and to Wolfram Hinzen for his recommendations for amendments to the penultimate draft.

NOTES

¹ For more on this conception of substance, see Lowe (1998), Chapter 6.

² See Strawson (1959), especially Chapter 3.

³ See Kim (2001). I do not want to imply that I myself fully endorse this sort of objection. As I have remarked elsewhere, the idea that causation must be ‘local’ was

effectively abandoned by the Newtonian theory of gravitation some 300 years ago: Lowe (2000b), pp. 22–23. Similar observations are made concerning post-Newtonian physics by Wolfram Hinzen and Juan Uriagereka in their contribution to the present issue of *Erkenntnis*. However, since many philosophers still seem wedded to a localization constraint on causation, NCSD can at least claim a dialectical advantage over Cartesian dualism in philosophical debate.

⁴ See, especially, Descartes (1984).

⁵ For more on this point, see Trenton Merricks, (1994), pp. 80–85. Although Merricks's objection is technically correct, I find it extremely implausible, because I regard an entity's ontological category as being one of its essential features – and 'body' surely qualifies as an ontological category *par excellence*. For more on ontological categories, see Lowe (2006), especially Chapter 1.

⁶ For extensive debate on the controversial relationship between conceivability and possibility, see Gendler and Hawthorne (2002).

⁷ See further Lowe (1998), pp. 202–203.

⁸ See, especially, Lowe (2001).

⁹ Compare Baker (2000). I explain why I myself don't accept a 'constitution' view of embodiment like Baker's, and thus see the foregoing analogy as imperfect, in Lowe (2001). See also Lowe (2000b), pp. 15–21. My own view is that the relationship of embodiment is a *sui generis* one, but none the less intelligible on that account.

¹⁰ For more details, see Lowe (1989), or Lowe (1997).

¹¹ Compare Baker (2000), pp.122–123.

¹² See further Lowe (1996), pp. 25ff.

¹³ See Hume (1978), p. 252 and also the Appendix.

¹⁴ See again Kim (1993). And again, for reasons mentioned in n. 3 above, I am not implying that I entirely sympathize with this sort of complaint. I merely recognize it as one that seems, rightly or wrongly, to have some purchase in the minds of many contemporary philosophers.

¹⁵ See, for example, Kim (1993). Once more, I do not mean to imply that I myself wish to endorse any such principle of causal closure, only that something like it is widely championed by many philosophers today. That contemporary physics does not in fact subscribe to or support any such principle is urged by Henry Stapp and by Peter Molenaar in their respective contributions to the present issue of *Erkenntnis*.

¹⁶ There is indeed ample empirical evidence that the pattern of brain-activity which characteristically precedes voluntary movement is (until very shortly before the movement) non-specific, widely distributed over the cortex, and gradual in build-up. See Deecke et al. (1969). See also. Popper and Eccles (1977), pp. 282ff and pp. 293f.

¹⁷ In Figure 1, the direction of time runs from the top to the bottom of the diagram, each node of the 'tree' represents a particular bodily event – with the foot of the tree representing the arm-movement and the nodes above it representing particular neural events – and each line between nodes represents a relation of causation, with the neural event that is represented by the *upper* of any two such connected nodes being represented as a cause of the neural event that is represented by the *lower* one. The diagram is, of course, very schematic and vastly simpler than a fully realistic one would be. But for our purposes it correctly represents the fact that, the further back in time we trace the neural antecedents of any particular bodily movement, the greater in number and more independent of one another we may expect them to be.

¹⁸ Note that, at this stage, I say only that it doesn't *seem* that one can *isolate* any such neural event, or set of neural events. In the next section of the paper, I shall

advance reasons for thinking that it is very unlikely that *there is* any such neural event, or set of neural events – one whose non-occurrence would have had exactly the same consequences as the non-occurrence of my decision. But, notoriously, it is very difficult if not impossible to ‘prove a negative’, if by a ‘proof’ one means an absolutely compelling argument. In matters like this, the most that one can hope to establish is a strong probability. Of course, if my conclusion is correct, there then remains the question of how my decision *is* related to neural events going on at about the same time, given that it is not related to any of them by *identity*. I shall say something about this matter in the final section of the paper, but a full account would require another paper.

¹⁹ See, especially, Lewis (1973). I simplify his account somewhat, but not in any way that materially affects the argument that I am developing.

²⁰ It may perhaps be thought that doubt is cast on the proposed identity of *D* with *N* by Benjamin Libet’s celebrated experiments concerning the timing of volitions: see Libet (1985). His findings are, to say the least, controversial, however, and I shall not consider them here: but, for further discussion, see Lowe (2000b) pp. 252–256.

²¹ All that I am presupposing here is that if *N* was indeed a cause of *B*, then this counterfactual is true. The physicalist cannot, I think, have any quarrel with me on this account. I am not taking any advantage, then, of the various reasons that have been advanced for doubting, at least in some cases, whether causal statements entail the corresponding counterfactuals. For discussion of some of these reasons, see Lowe (2002), Chapter 10.

²² For very full discussion of the distinction between event causation and fact causation, see Bennett (1988), pp. 21ff. I don’t mean to suggest, however, that I endorse every aspect of Bennett’s characterization of the distinction.

²³ I discuss this point more fully in Lowe (1999).

²⁴ See again, in particular, Lowe (1999).

²⁵ See, for example, Lowe (2003a).

²⁶ Compare Dancy (2000).

²⁷ For more on the consistency of interactive dualism with physical causal closure, see Lowe (2000a).

²⁸ Here it may be asked: But what if a physicalist were to urge not only a principle of physical causal closure for *event* causation, but also one for *fact* causation – and thus claim, in effect, that *any* cause of *anything* physical must itself be physical? Would this not rule out anything like my NCS model of mental causation? Yes, of course it would. But such a physicalist could not appeal to this claim to *argue* against an interactive dualist position like mine, because any such ‘argument’ would be blatantly question-begging. To assert that any cause of anything physical must itself be physical is equivalent to asserting that no cause of anything physical can be non-physical, which directly contradicts the interactive dualist’s claim that something physical may have a non-physical cause. A ‘causal closure argument’ that appeals to a principle of causal closure which is itself *inconsistent* with interactive dualism amounts, in effect, to nothing more than this: *P*, therefore not not-*P*. Hence, it is in the physicalist’s own interest not to appeal to a causal closure principle that is so overwhelmingly strong as this. See again Lowe (2000a) for further elaboration of this point.

²⁹ In this connection, see Lowe (2003b).

REFERENCES

- Baker, L. R.: 2000, *Persons and Bodies*, Cambridge University Press, Cambridge
- Bennett, J.: 1988, *Events and their Names*, Clarendon Press, Oxford
- Dancy, J.: 2000, *Practical Rationality*, Clarendon Press, Oxford
- Deecke, L., P. Scheid and H. H. Kornhuber: 1969, 'Distribution of Readiness Potential, Pre-Motion Positivity and Motor Potential of the Human Cerebral Cortex Preceding Voluntary Finger Movements', *Experimental Brain Research* 7, 158–168.
- Descartes, R.: 1641, 'Meditations on First Philosophy', in J. Cottingham, R. Stoothoof, and D. Murdoch (eds.): 1984, *The Philosophical Writings of Descartes*, Cambridge University Press, Cambridge.
- T. S. Gendler and J. Hawthorne (eds.): 2002, *Conceivability and Possibility*, Clarendon Press, Oxford.
- Hume, D.: 1978, *A Treatise of Human Nature (1739–40)*, L. A. Selby-Bigge and P. H. Nidditch, (eds.) Clarendon Press, Oxford.
- Kim, J.: 2001, 'Lonely Souls: Causality and Substance Dualism', in K. J. Corcoran (ed.), *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, Cornell University Press, Ithaca, New York.
- Kim, J.: 1993, 'The Non-Reductivist's Troubles with Mental Causation', in J. Heil & A. Mele (eds.), *Mental Causation*, Clarendon Press, Oxford 189–210.
- Lewis, D. K.: 1973, *Counterfactuals*, Blackwell, Oxford
- Libet, B.: 1985, 'Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action', *Behavioral and Brain Sciences* 8, 529–566.
- Lowe, E. J.: 1989, 'What is a Criterion of Identity?', *Philosophical Quarterly* 39, 1–21.
- Lowe, E. J.: 1996, *Subjects of Experience*, Cambridge University Press, Cambridge
- Lowe, E. J.: 1997, 'Objects and Criteria of Identity', in B. Hale & C. Wright (eds.), *A Companion to the Philosophy of Language*, Blackwell, Oxford 613–33.
- Lowe, E. J.: 1998, *The Possibility of Metaphysics: Substance, Identity, and Time*, Clarendon Press, Oxford
- Lowe, E. J.: 1999, 'Self, Agency, and Mental Causation', *Journal of Consciousness Studies* 6, pp. 225–39, reprinted in B. Libet, A. Freeman and K. Sutherland (eds.): 1999, *The Volitional Brain: Towards a Neuroscience of Free Will*, Imprint Academic, Thorverton, pp. 225–239.
- Lowe, E. J.: 2000a, 'Causal Closure Principles and Emergentism', *Philosophy* 75(2000), 571–585.
- Lowe, E. J.: 2000b, *An Introduction to the Philosophy of Mind*, Cambridge University Press, Cambridge
- Lowe, E. J.: 2001, 'Identity, Composition, and the Simplicity of the Self', in K. Corcoran, (ed.), *Soul, Body, and Survival*, pp. 139–158.
- Lowe, E. J.: 2002, *A Survey of Metaphysics*, Oxford University Press, Oxford
- Lowe, E. J.: 2003a, 'Personal Agency', in A. O'Hear (ed.), *Minds and Persons*, Cambridge University Press, Cambridge 211–227.
- Lowe, E. J.: 2003b, 'Physical Causal Closure and the Invisibility of Mental Causation', in S. Walter & H. D. Heckmann (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, Imprint Academic, Exeter, 137–154.
- Lowe, E. J.: 2006, *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Clarendon Press, Oxford

- Merricks, T.: 1994, 'A New Objection to A Priori Arguments for Dualism', *American Philosophical Quarterly* **31**, 80–85.
- Popper, K. R. and J. C. Eccles: 1977, *The Self and its Brain: An Argument for Interactionism*, Springer, Berlin.
- Strawson, P. F.: 1959, *Individuals: An Essay in Descriptive Metaphysics*, Methuen, London

Department of Philosophy
University of Durham, 50 Old Elvet,
DH1, 3HN, Durham, UK
E-mail: e.j.lowe@durham.ac.uk