

4 Substance Dualism

T. J. Mawson

Introduction

Substance dualism could not have a more venerable lineage, being traceable back through Descartes at least as far as Plato and Socrates. However, the respect with which people treat the view has declined to such an extent in the last few hundred years that it has recently been described as, not so much a position to be argued against, as a cliff over which to push one's opponent. Certainly within contemporary educated circles, were one to venture the opinion that we have souls, one should expect to find oneself held to have propounded an extravagance only slightly less great than had one ventured the opinion that visiting extra-terrestrial life was in part responsible for the construction of the pyramids or that Elvis may be seen working in the local chip shop. The most favourable response one could realistically hope for would be the concession that perhaps, before the development of such things as computer science and neuroscience, such a whimsy might have been excusable, but even so, *now* souls must surely go the way of phlogiston and light-carrying ether: onto the intellectual scrap heap.¹

Here I shall advance the claim that, despite the near universality of the assumption that the theory may be easily cast aside, within the structure of a hylomorphic substance/property metaphysic, the only reason to suppose that we do not have souls is that provided by Ockham's razor and even that reason is conditional upon an assumption, albeit an assumption that it is no more my intention to cast doubt upon here than it is my intention to cast doubt upon the substance/property metaphysical structure within which I shall be framing this debate.² The assumption is that there is physical stuff.

Given that there is physical stuff, it would indeed be simplest to suppose that the mind is ontologically reducible to that or to processes going on in that. But given that, as I shall also argue, there are some reasons to suppose that we *do* have souls – that is, that such a reduction cannot be accomplished, so one might find oneself, probably idly, reflecting on the fact that idealist substance monism would offer one all the advantages of simplicity offered by physicalist substance monism while in addition accommodating these reasons for supposing we have souls. This reflection would probably be idle as there is little

danger of idealist substance monism emerging, on balance, as the preferable theory of the mind for us: the assumption that there is physical stuff is held by most of us so deeply as to be near immovable by argument. (For similar reasons, I shall ignore neutral monism.) Thus it is that most of us will find ourselves weighing the reasons in favour of the claim that we have souls in the balance against the rational attractiveness of the simpler metaphysic that a physicalist substance monism offers. Where this balance ultimately settles is something on which opinions will divide. All I can hope to secure consensus on by what follows is that no substance/property metaphysic will give us everything we want, which in itself of course is a reason to re-examine the substance/property starting point. If one refrains from doing that however (as I shall), one must conclude that either several of our assumptions concerning the nature of persons (assumptions which are not held significantly less deeply than is our assumption that there is physical stuff) are in error or the world is more complex than physicalist substance monism allows, for we do have souls after all.

What Substance Dualism is

We have to start somewhere and time is pressing, so let us put onto the table without offering argument in its favour a certain commonsense realism about the physical world and our knowledge of it as gained through the natural sciences. First then, let us assume that there is physical stuff. This may be characterized as stuff of the sort that we suppose ourselves to encounter with our five senses in everyday life; that our folk science describes more or less adequately for our everyday purposes; and that our natural sciences describe with increasing accuracy as they develop. We may define the sort of stuff we have in mind by paradigm examples of things which are made of it: this desk, here; that star, there; and so on. In a previous century, we might have called this physical stuff simply 'matter', but now we know that matter may be converted into energy and vice versa and we hear scientists speculate concerning quarks, hyper-dimensional strings, and so forth as making up the more commonplace objects that we encounter in everyday life. These are things which, while striking us as no doubt physical, do not strike us as in any obvious way material, so, instead of 'matter', we call this stuff 'physical stuff'. We shall call the view that this physical stuff is all the stuff that there is 'physicalist substance monism' or 'physicalism' for short. Obviously one might hold that in addition to this sort of stuff, there is another type of stuff as well. We shall call this second view 'substance dualism'.

Substance dualism is committed then to there being a type of stuff that resists full integration into the natural sciences. What we might call 'partial

integration' will need to be allowed for to take into account psycho-physical causal interactions, which – as we shall see – the most plausible substance dualist view will wish to maintain occur. Incapacity for full integration is not however by itself enough to characterize this second type of stuff adequately. There might turn out to be a sort of stuff that resisted the sort of integration that the substance dualist will wish to claim for his or her souls yet which it was obvious to commonsense was nevertheless purely physical; the unity of the natural sciences is a hope or perhaps something stronger, a regulative idea. But it could have turned out, or could even yet turn out, to be misguided.

Similar problems would beset attempts to characterize this second type of stuff in terms of its failure to fit into the existing categories of the natural sciences. We will not wish now to draw up a list of what properties might be made mention of in a completed science, conscious as we are that some of the properties of physical stuff as quantum physics describes them are very different from the properties of it as we encounter it in everyday life or as would have been supposed to be primary and fundamental in the days of the corpuscularians; they are, it has been said with some understatement, spooky. But we can issue a promissory note here and that is sufficient for our current purposes: the second type of stuff in which the substance dualist believes is a type of stuff the nature of which will not be fully integrated into a completed science of objects such as our paradigm physical objects – tables, stars, et cetera – because it has properties that will not feature in that completed science and will not be reducible in any way to properties that do so feature. The obvious contender for the fundamental property here is the capacity to have mental properties such as beliefs, desires, emotions, and so on. We might think of intentionality and/or first-person privilege as their hallmarks.

The substance dualist will maintain that the essence of soul substance is that it is capable of thought in the broadest sense of the term, and this is a property which would, in a completed science, turn out *not* to be a property of stuff of the sort that makes up tables, stars and so forth and turn out not to be reducible to any such properties. Nevertheless, the substance dualist maintains, contra the eliminative materialist, thinking is definitely going on and, given the substance/property metaphysic within which this debate takes place, he or she validly concludes from all of this that it must thus be going on in a substance other than a physical one – soul substance, as we have been calling it. We might say then, more or less following Descartes as we do so, that, according to the substance dualist, it is of the essence of physical substance to have properties of the sort that our paradigm examples of physical objects – tables, stars and so on – have, which will not include thinking. (Descartes settled on spatial extension as the essential property of physical stuff; we have left this more open; perhaps spatial extension as we ordinarily understand it will turn out in a completed science to be a property of only some physical stuff

[e.g. tables], a property constructed out of more basic elements.) It is of the essence of soul substance that it is capable of thought, where thought is taken in the broadest of senses to include all mental happenings – beliefs, desires, sensations, emotions, acts of the will, and so on. Belief in the existence of these two types of substance is what is definitional of substance dualism.

There are various views within substance dualism about the relationship between this soul stuff and us as persons. Are we as persons simply our souls, or do we persons have souls as one part and bodies as another? Descartes entirely identified the person with his or her soul, and thus would have had no objection to our talking of disembodied souls – were they to continue on after bodily death – as fully the people they had been when earlier embodied. An alternative view is possible. Arguably it is that held by Aquinas. This is the view that persons are to be identified with the conjunction of body and soul and thus that where these two cease to be conjoined in the right sort of way – most obviously, perhaps, if they cease to be conjoined in any way at all as one of the conjuncts entirely ceases to exist (e.g. the body is vaporized by an exploding nuclear device) – what survives, if anything, is not the person in his or her entirety, but merely a part of the person. And it may be that a disembodied soul part (of a former person) would not, as a matter of causal fact, be able to do any thinking once separated in this way from the body part with which it had previously formed a person. One might go even further down this track and think that the destruction of the body part would inevitably cause the destruction of the previously associated soul part and thus the entirety of the person. But if any of these things are so, then, according to substance dualism, they are so as a matter of metaphysical contingency, not necessity. substance dualism makes it metaphysically possible for the person (Descartes) or a part of the person (Aquinas) to survive the complete and final destruction of his or her body, but it does not entail that this actually ever happen. It makes it metaphysically possible that any disembodied soul would be able to have a mental life as rich in what we might call ‘pure’ mental properties (not, for example, suffering from toothache – a concept which spans the ontological gap between body and soul) as an embodied soul, but it does not entail that this actually ever happens.³

There is a third view, interior to substance dualism, although it has not in fact ever been propounded by anyone who believes we do have souls; this is the view which would identify us entirely with our bodily parts. The most plausible variant of this view would, it strikes me, have to give up on the idea that we are fundamentally persons and thus may be pictured as having something akin to animalism: we are our animal selves; at the moment, these animal selves happen to be in causal contact with souls and thus happen to be able to think (and through doing so become persons), but, were such souls destroyed, no element of our animal selves, and thus no element of us, would be destroyed;

we would just cease to be able to be persons. The soul is entirely inessential to what makes us *us* even though it is not inessential to what makes us persons. As it has in fact never been propounded by a substance dualist, despite its being a potential variant of the view, I shall ignore this view in what follows.

As well as providing the materials with which different views of the nature of the self and us-as-persons may be constructed, substance dualism also allows for various views on the causal commerce between souls and bodies. For various reasons which will become apparent as we progress, the most plausible variant of substance dualism is interactionist substance dualism' (the body causally affects the soul and vice versa). The alternative views are psycho-physical parallelist substance dualism (the two have no causal interchange whatsoever); epiphenomenalist substance dualism (the body causally affects the soul, but not vice versa); and the view – again a 'neglected alternative' in that no-one actually holds it – that the soul causally affects the body, but not vice versa.

Reasons to Suppose Substance Dualism False

As mentioned in the introduction, as a theory about the nature of the mind, substance dualism is more ontologically extravagant than substance monism. Given that there is physical stuff, it would be simplest to suppose that the mind is somehow reducible to that stuff, or, more plausibly, to processes going on in certain bits of that stuff: brains, presumably; mind is to brain – *mutatis mutandis* – what digestion is to the digestive system. Given that the properties of physical stuff are by no means obvious to us – and recent scientific developments have indicated that some seem to be spookier than earlier generations of scientists would have found even imaginable – and given that, from what we already know, the brain is the most complex structure in the universe, it is not unreasonable for us to hold out hope that a completed science would be able to fill in the *mutatis mutandis* here. Of course it cannot do so yet, but these are early days. This is, it must be conceded, a reason to suppose substance dualism false. What reasons might we have to suppose it false beyond its complexity relative to physicalism?

I shall consider two areas from which it is often suggested additional reasons for supposing substance dualism false emerge.

The first area centres on supposed problems in identifying souls, both ontologically and epistemically. What is it that makes one soul different from another and how can we ever know of souls that they are the same over time or know of souls other than our own that they exist at all? In short, my analysis here will be as follows: firstly, insofar as substance dualism faces problems that parallel those faced by physicalism (as it does in addressing the issue of what

makes one fundamental unit of substance different from another and how we know of such units that they are the same over time), these problems cannot be reasons to favour physicalism over substance dualism and so are not properly construed as objections to it, rather than perhaps as objections to the wider substance/property metaphysic within which this debate is taking place. Secondly, insofar as substance dualism faces problems not faced by physicalism (as it does in addressing the issue of how we can ever know that units of substance other than our own exist at all), the fact that it commits one to a certain sort of scepticism here is a reason to suppose it true, not false.

The second area of concern centres around supposed problems in explaining the causal interaction between the two sorts of substance the substance dualist posits. How can mind and body act on one another? Does not any answer to this question run into insuperable problems from what we already know of physics, for example concerning the causal closure of the physical world and the conservation of energy? In short, my analysis here will be that the interactionist substance dualist is not beholden to answer the question of how mind and body act on one another, rather than merely assert that they do, as it is not a commitment of interactionist substance dualism that this question will be answerable by us. Positing that there is an interaction of this kind does not in fact require one to contradict things which we already know of physics, although there is potential for physics (were it to move back into a deterministic mode) to put pressure on the claim that there is in fact interaction of the sort posited. At the moment then, there is no reason from science to suppose interactionist substance dualism false.

Let us go into these objections in more detail.

Problems of identification

We may sensibly ask the substance dualist what it is that makes one soul distinct from another and predict that he or she will have little informative to say by way of reply. Obviously, he or she may maintain that it is extremely unlikely that two souls will have all the same properties as one another, so – he or she may point out – any two souls will in fact differ in this fashion. One will be thinking about strawberries, another, about cream, and so on. But exact qualitative identity between two souls is not a metaphysical impossibility generated by the nature of souls per se and even if it were somehow impossible for two souls to have exactly the same properties, this impossibility would not ground the numerical difference between two souls, but rather presuppose it. In any case, it looks as if the substance dualist should agree that there is nothing in the nature of souls per se that prevents there being two qualitatively identical yet numerically distinct souls, for it seems that there's nothing in the nature of

souls per se that prevents there being an exact duplicate of this universe. In that universe there would consequently be a person thinking qualitatively identical thoughts to those that you are currently thinking. That person would, nevertheless, not be you; it would be your duplicate. So, the substance dualist should say that it is not fundamentally in virtue of their different properties that different souls are different. Rather, he or she should admit that souls might in principle differ *solo numero*. (They have what is sometimes called 'thisness'.) Need he or she be embarrassed that he or she can say no more than this? I do not think so.

Presumably the person who believes in units of physical substance will wish to maintain that at least with regards to some of these there is nothing in their nature that prevents their differing *solo numero* too. The classic thought experiment on this topic involves imagining a universe composed simply of two chemically pure iron spheres, each of the same diameter, hanging in otherwise unoccupied space a certain distance away from one another; these spheres would be qualitatively identical to one another, yet they would be numerically distinct. Can the physicalist substance monist say more about how these two spheres manage to retain ontological individuality than that they do, that they differ *solo numero* or have thisness? No. So the substance dualist need not feel embarrassed about being able to say no more than this about how two souls might retain their ontological distinctness even were they to have qualitative identity.

As this discussion might have already indicated, this type of issue – and in fact the one we are about to go on to discuss – is an artefact of believing in substance as such, (i.e. of believing in things to which the principle of the identity of indiscernibles does not apply of necessity). As such, this type of issue and the one we are about to go on to discuss cannot be a reason to prefer any theory that claims that substances exist over any other that claims that they do. Thus it cannot be a reason to prefer substance dualism over physicalism.

Belief in substance raises certain problems at the epistemic level. Of substance dualism, it is sometimes said, souls might be swapping bodies every few minutes but each inheriting the psychological properties of the soul that had just vacated the body into which the new one was now moving. Were this to be the case, no one would be able to detect these changes, yet people (Descartes) or significant parts of people (Aquinas) would constantly be swapping bodies. Furthermore, we seem to face on substance dualism a peculiarly intractable variant of the problem of other minds: how do you know, as you encounter another person through the medium of the physical world, that he or she is a person *at all*, that he or she has a soul in the right sort of causal connection with the body which you observe directly?

Again we may observe that the first problem affects those who believe in substance per se and thus in substance of the physical sort; thus, whatever it

is a reason to believe, it cannot be a reason to believe in physicalism over substance dualism. How do you know that the physical stuff underlying the properties of the desk in front of you has not been swapped out by some malign demon in the last few moments, leaving all the properties 'behind' in the sense of their being inherited by the physical substance which this demon instantaneously moved in to replace that which he was removing? So this worry generalizes to physical substance. But, having said that, it's not too great a worry. The physicalist substance monist's response to this sort of worry seems to me entirely adequate. It is indeed metaphysically possible for the substance of the desk to be being changed in the imperceptible way suggested, or, if this is not metaphysically possible, then that is for reasons exterior to the nature of physical substance per se (e.g. that there can be no spirits of the right sort). But unless we have positive reasons for supposing that such swaps are happening, as it would be simplest to suppose that they are not, so we should suppose that they are not. The same move, then, that both the physicalist substance monist and the substance dualist make with respect to physical substance, the substance dualist makes with respect to soul substance. If it works in one area, what reason is there to suppose it will not work in the other? None.

The problem of other minds is often thought to particularly affect – and thus speak against the truth of – substance dualism; were substance dualism true, it is suggested, there would be peculiar difficulties in our knowing that other people exist. I shall deploy a two-pronged approach to meeting this charge: first, I aim to show, similarly to previous objections, that, if this is a problem, it is a problem that is faced, at least to a greater extent than is often appreciated, by physicalism too. Secondly, as it must be admitted that, *pace* point one, it is faced to a *greater* extent by substance dualism, so I shall aim to show how this 'extra' problem of other minds is not, in fact, one it is implausible to suggest we face. Were substance dualism true, there would indeed be an extra difficulty in knowing that others have minds, but that is not a reason to suppose that substance dualism is false; indeed it is a reason to suppose it true for there is much plausibility in suggesting that we do face this extra epistemic hurdle in coming to know that others have minds.

First, then, though it would take too much time to argue it here, the most plausible physicalism will identify the having of a mind with physical processes that are *recondite* in the extreme. For example, a crude behaviourism, whereby being angry is simply behaving in a certain fashion, which may be specified entirely adequately in terms of movements of the body, movements that are sufficiently macroscopic for us to be able to identify them without any great difficulty, using our unassisted five senses, will not prove adequate to the task. Rather, some neurological happenings of a certain type will need to be called upon in the analysis of anger, but as soon as the physicalist pushes

the happenings which are mind-happenings, interior to the skull, then, unless we meet people who are themselves in fMRI scanners of sufficient sophistication to reveal to us these happenings, we never ourselves see the happenings that are, on the physicalist account, being angry, or what have you. On physicalism no less than substance dualism, we never observe the having of minds other than our own. How then do we know, if physicalism is true, that others have minds?

To cut a long story short, the answer to this question is that they tell us that they do, and we ordinarily have no reason to doubt them. Someone says that he or she is suffering, let us say, from anger. If physicalism is true, they will be speaking truly if a certain happening is occurring in their brain; but we do not see this happening and indeed at the current stage of science might not know that it was their feeling of anger, even if we did see it. But, unless we have reason to doubt them (e.g. they are performing in a play or some such), we are surely rational, whatever the theory of mind to which we subscribe, in believing that they are angry simply on the basis of their saying that they are. Without taking this sort of epistemic route into knowledge of others' minds, it would be impossible for the physicalist substance monist to construct the theories by which he identifies to his satisfaction the having of anger with the brain happening that he could then, in principle, find to be universally correlated with the tendency to report it. (This is sometimes called the 'privilege' that must be given to first-person reports of the mental.) But if that is so, then this same route is open to the substance dualist.

It is true that on the substance dualist view, the actual feeling of anger is something happening in a substance even more recondite than the inner parts of the brain. It is happening in a soul and thus in something that could never be revealed by investigation into the physical world however advanced fMRI scanners became. But the same route which the physicalist substance monist takes in everyday life, before hand-held fMRI scanners and the like become commonplace (and which he or she will have to hold as epistemically authoritative even were they to do so, to accommodate the issue of privilege), is open to the substance dualist. This is how the problem of other minds is to be overcome *whatever one's theory of mind*: by taking claims to have minds as a *prima facie* reason to believe minds are had.

However, moving on to the second point, it seems as if the physicalist substance monist may argue that whatever problems he or she faces in coming to knowledge of others' minds, and however these are to be overcome, the substance dualist *must* face an additional problem unless he or she posits some direct and very reliable telepathic contact between minds as an alternative source of knowledge, which positing would itself be most implausible. This is a true point. But does it speak against or really in favour of substance dualism?

Were physicalism true, then, after science has been completed and presuming it has allowed for hand-held fMRI scanners or some such of sufficient accuracy – let us call them ‘brainoscopes’ – one could perhaps confidently bypass first-person reports as a source to knowledge of others’ minds; one could, instead of speaking to a person, directly apply one’s brainoscope to someone’s skull and, on the basis of its findings, confidently report things like, ‘No need to speak; I see from my brainoscope that you are angry at my having applied it to your head without first asking your permission’. These reports could be unfailingly accurate. (Note: not all physicalists believe that this will prove possible, but we are considering the views of one who does in order to point out the contrast with substance dualism and the ‘extra’ problem of other minds that it faces.) Let us consider a physicalist substance monist who contends that, after science has been completed, one will be in a position to know that a certain brain state or some such may be identified with anger being felt at having had a brainoscope applied without having been asked for permission and, with the technology of the brainoscope properly applied, one will know that this brain state is being had, so, one will know that the person is angry in this way. For such a physicalist substance monist, there will then be no ‘gap’ into which a sceptical doubt may creep. It might appear that nothing similar could happen on substance dualism. But, in fact, the substance dualist may hold that it could. If substance dualism is right, then in a completed science this technology might well be possible. The substance dualist of course would not make the extra step of identifying the brain state or what have you that is revealed by the brainoscope with the mental state, but he or she can acknowledge that there might well turn out to be a perfect correlation of the sort the physicalist we are considering anticipates our finding, and thus the substance dualist might admit that the sort of brainoscope that is capable of bypassing first-person reports in the manner described could well turn out to be possible. But there is, nevertheless, it must be conceded, a gap for the substance dualist here relative to his physicalist substance monist counterpart, a gap generating an ‘extra’ problem of other minds.

The extra problem for the substance dualist is generated because it will always remain possible that the brainoscope is in error, even once the science is completed and the brainoscope working (for all we know) properly, for, according to the substance dualist, the brain state or what have you that the completed science finds universally to be conjoined with a thought of a certain kind (and we are supposing that this is what it will find) and that the brainoscope correctly reports to be present in this case is not to be *identified* with the thought of that kind. According to the substance dualist, one could know everything about the physical world, yet not know *without the possibility of error* what mental state a person was in (or indeed even if they were a person at all) for there is – according to substance dualism – an ontological gap between the

physical world and the mental, a gap which may be 'bridged' by causation, but – causation not being a conceptual relationship – any particular bridge across which may or may not hold and thus any particular judgements using which may be in error. But now this extra problem of other minds for the substance dualist looks more like an asset than a liability, for, as we shall see when looking at 'Mary-type' arguments for substance dualism, it is apparently possible that someone might know everything about the physical world yet not know something about the mental, which appearance has to be ruled out as deceptive by the physicalist substance monist we are considering.

Problems of interaction

The version of substance dualism on which we are focusing suggests that there is two-way causal exchange between physical substance and soul substance. This is often held to generate problems for the view. First, it is suggested that it runs contrary to a finding of physics. In particular, it looks as if the principle that matter/energy is conserved across a closed system such as the physical universe must be violated if substance dualism of the interactionist sort is true. Second, it is suggested that there is something problematic in general in any case – regardless of whatever physics might be telling us – about non-physical substances causing changes in physical ones and vice versa. We know, a priori, that such is an impossibility.⁴ I do not find either of these two lines of thought tempting.

Let us suppose for a moment, what we shall later see is in any case false, that the interactionist substance dualist is committed to laws of physics being violated. It does not seem that an objection arising from this commitment would be any more than a restatement of the objection from the relative complexity of substance dualism over physicalism. Obviously it would be simpler were the universe closed and the laws of physics not violated, and that is indeed, we have already conceded, a reason to suppose that it is so. We should not 'double count' this objection to substance dualism. In fact though, the interactionist substance dualist is not committed to his or her souls' violating natural laws.

With the advance of physics beyond determinism, another possibility arises. The substance dualist may maintain that happenings in the brain which are caused directly by the soul are caused in ways compatible with the preceding brain state and the laws of nature, but – these two not being such as to necessitate what state emerges from them – they are caused to be the particular way that they are by the soul. That the brain be in state q , rather than state r , after it has previously been in state p is something which was always allowed for by the preceding physical states (given indeterminacy), but, in fact, the substance dualist may maintain, that it ended up in state q was caused by the relevant

person's soul. It is no bar to this theory to point out, if such a fact can be pointed out (and it is doubtful that it can be), that, of any individual sub-microscopic event where such quantum indeterminacy plausibly reigns, it seems incapable of producing cascade effects up to the macroscopic level which results in arms being moved and so forth. For presumably some brain state leads to macroscopic happenings such as arms being moved, and this is made up at the sub-microscopic level of many such quantum happenings. So, the substance dualist may maintain that the soul's influence on the brain, in causing it (e.g. to raise one's arm, occurs in a number of disparate tiny locations, any one of which is perhaps not sufficient, or perhaps even necessary, for the event to occur, but which then jointly cause one's arm to rise). Those quantum happenings in the brain which are similar in the properties they reveal to the natural sciences as those happening in an 'inanimate' object where they are indeed uncaused are in fact, when they happen in the animate object that is the brain, caused by the soul of the relevant person. The universe is not indeed causally closed, but no laws of nature need be violated.

So, in short, even were fundamental physics to return to a deterministic mode, the interactionist substance dualist could maintain that souls are able to influence physical stuff (and vice versa) although by doing so he or she would be positing that the laws of physics are violated – little bits of energy come into and go out of existence. However, within the current indeterministic paradigm, no such violations are required as a part of the substance dualist's account of this interaction. The substance dualist may maintain that the soul operates in the causal 'gaps', otherwise filled by randomness, that indeterminism opens up. And of course even were the dominant paradigm of interpretation of the laws of nature within the community of physicists to revert to determinism, it would still be just a paradigm of *interpretation*; there would be no necessity that the substance dualist follow it.

Of course, such suggestions on the part of the substance dualist presuppose that in general a spiritual substance may cause a change in a physical substance and vice-versa, and someone might hold as a matter of principle that the only possible *relata* of causation are physical events, so such a suggestion may be ruled out in advance. But why adopt such a principle? It may be rejected by the substance dualist as mere prejudice if argued for a priori (although of course if argued for validly a priori, the substance dualist will need to find one or more premises to which to object) and the substance dualist will insist that such a principle cannot be discovered a posteriori, for the actual universe is one which has souls operative in it, so does not follow it. Descartes himself said all that, it strikes me, needs to be said on this issue in a letter to one of his objectors.

These questions presuppose among other things an explanation of the union between the soul and the body, which I have not yet dealt with at all.

But I will say, for your benefit at least, that the whole problem contained in such questions arises simply from a supposition that is false and cannot in any way be proved, namely that, if the soul and the body are two substances whose nature is different, this prevents them from being able to act on each other. (Descartes, in Cottingham, vol II, 1994, p. 275).⁵

So, in summary: the reasons for supposing interactionist substance dualism false and physicalism true reduce to the simplicity of the latter over the former. Simplicity is a reason to prefer one theory over the other, but so is explanatory adequacy, and it is far from clear that physicalism will prove adequate, as we shall now see.

Reasons to Believe Substance Dualism True

Various arguments in favour of substance dualism have been put forward over the last two and a half thousand years, and it would be impossible to provide an adequate treatment of all of them in anything smaller than a substantive book. That being so, in the space that remains for me, I wish to focus on just three areas where, it strikes me, the substance dualist can plausibly contend that substance dualism does better than physicalism in accommodating various 'commonsense intuitions' we have about ourselves. Of course commonsense intuitions are hardly the basis for conclusive arguments in favour of substance dualism. After all, if our commonsense intuitions about such issues were not sometimes wrong, there would hardly be any point in the discipline of metaphysics. I conclude then by discussing what weight we may in general give to this type of argument relative to the weight we may give to the virtue of simplicity which, it has been conceded, physicalism has over substance dualism. The three areas are personal identity, freedom, and consciousness. I shall consider them in order.

Personal identity

What is it that makes a person at a later time, $t+1$, the same person as existed at an earlier time, t ? Substance dualism has a simple answer: it is fundamentally the continuity of the same soul (or, for Aquinas perhaps, the same soul and the same body), and souls themselves do not continue in virtue of anything more basic continuing (bodies presumably do). For the physicalist substance monist, the issue is more complicated: there are three options. The person may be identified with a certain set of properties (usually psychological properties are chosen); with a part of the physical substance which makes up his or her

body (usually the brain is chosen); or with a combination of these (e.g. psychological properties p going on in brain b). However, none of these options seems to offer a satisfactory theory of personal identity. There are problems peculiar to each, but a general defect may be observed in play in their dealing with almost all the thought experiments that are used, it is supposed, to illuminate this issue.

So, for example, one is asked to imagine a brain bisection, after which the two resultant hemispheres are transplanted into separate clones of the original body where they take up more or less functional residence. To add weight to the situation, perhaps one of the resultant people is then tortured to death over the next five minutes while the other is given a gin sling to enjoy. Which of these two resultant people, if either, is the person who originally underwent the brain bisection? one is asked. Then the details of the experiment are altered; perhaps one of the two resultant people gets more psychological continuity and the other more of the physical substance of the original brain. What then do we say? For some proportions of psychological continuity and continuity of physical substance, the physicalist must say that it either becomes ontologically indeterminate whether a resultant person is the same as the original, or it remains determinate, yet he or she does not know whether he or she is the same or instead a new person inheriting some of the original's psychology and/or brain matter. But our commonsense intuitions about personal identity do not allow for indeterminacy, as shown most markedly when one thinks of these possibilities from the first-person perspective of someone about to undergo the relevant experiment: 'Either I will survive or I won't; it cannot be ontologically indeterminate in a few minutes time whether I'm there or not'. But nor is there anything unknown left for the physicalist to hang a determinate fact of personal identity from, something which again we might perhaps see most sharply by imagining the first-person perspective: 'If I can know where all the properties are going and know where all the physical substance is going, yet still not know where I am going, then I cannot be identified with any combination of properties or the physical substance; I must be something else, and the only something else left (once we've swept properties and physical substance off the table) is soul substance'. This is not conclusive of course, for one could be – unbeknownst to oneself – identical to some indivisible property or indivisible bit of physical stuff and thus even if one knew in advance of this property/bit of stuff where it was going to go, one would not know that in knowing this one was knowing where one was oneself going.

However, each of these claims – that one is to be identified with an indivisible property or an indivisible bit of physical stuff – would itself be most implausible. Properties and sets of properties (whether properties of physical substance or soul substance) are capable of multiple instantiation, and the sorts of sets with which people might most plausibly be identified (that go into

psychological continuity accounts of personal identity) are themselves capable of degrees of survival. But people are not the sorts of things which seem to commonsense capable of multiple instantiation, or 'division' as the most discussed variant of multiple instantiation is sometimes called. One would not think, 'Maybe, in five minutes time, I'll be two people, one being quickly tortured to death and one enjoying a pleasant drink, so that in ten minutes time I'll be both alive and dead'. And, as already observed, people are not the sorts of things which seem capable of survival by degrees. So it is implausible to suggest that persons are to be identified with any set of properties or any one indivisible property. Physical substance is not capable of multiple instantiation, but the set of bits of physical substance with which people might most plausibly be identified – brains – are capable of division and, as already observed, people are not capable of division. Of course there might be some genuinely indivisible bit of physical substance within the brain – an 'atom' in the original Greek sense – which enables one to side-step this issue if one identifies oneself with that, but to posit such a thing and to identify oneself with it would be most implausible. Substance dualism, then, gives the best theory of personal identity by reference to our commonsense intuitions about persons as not being capable of multiple instantiation/division and as not being capable of survival to a degree.⁶

Freedom

In everyday life, we often suppose ourselves to have been able to do something different from whatever it is that we ended up doing, even had everything else in the physical universe up to the moment of our choice remained exactly the same. Of course, if everything else had really remained *exactly* the same, then we might wonder why we would ever have behaved differently from however it is we ended up behaving, but some of our choices are, we believe, whimsical in the following way. I am offered the choice between tea and coffee; I have no preference between the two but a strong preference to have one rather than remain thirsty and, not wishing to be like Buridan's ass, I thus say, on a whim, 'Tea, please'. In reflecting back moments later, I believe of myself that I could have said 'Coffee, please' instead. In a situation such as this, although perhaps most vividly in situations where things of great moral moment turn on what we end up doing, we suppose that the fact that we end up doing whatever it is we do end up doing is – to some extent at least – down to us. I say 'to some extent at least' as it must be conceded that we always operate within a finite range of options and sometimes this finitude exculpates us from at least some, possibly all, responsibility ('I agree that what I did was bad, but look at the alternatives I faced; each was worse'), but we ordinarily suppose that this finite

range is greater than one – we do genuinely have options – and, when we have options and end up realizing one rather than another as a result of the right sort of conscious choice on our part, we suppose that in that way the causal and moral buck stops with us. We are in this way free agents, responsible to a greater or lesser extent for the choices we make and thus for the shape of our lives and the lives of those we affect.

Substance dualism – of the interactionist sort – gives a straightforward and simple account of how all of this gets to be so. (That is the long-promised reason why interactionist substance dualism, rather than, for example, psychophysical parallelism or epiphenomenalist substance dualism is the most plausible.) According to interactionist substance dualism, the soul, while of course being affected by things going on in the physical world (e.g. in coming to the beliefs that it has about that world), is not always necessitated to do what it does by those effects; sometimes it initiates causal chains, which then impinge upon the physical world when it could yet have initiated different causal chains and thus impinged differently, had it chosen to do so. When my soul does so, that is me (Descartes) or a part of me (Aquinas) making a choice. The commonsense view of ourselves as articulated in the previous paragraph finds its metaphysical grounding.⁷

Physicalism cannot ground this commonsense view. On physicalism, either what I ended up doing was entirely causally necessitated by preceding states extending back through time to the big bang or there was a certain amount of randomness (uncaused-ness) involved in the causal chain that ended up with my doing whatever it was I did. In neither case would the causal – and, one might hence think, moral – buck stop with me; either the happening was caused by factors beyond my control (for they go back to the big bang, which is certainly beyond my control); or it was random; or it was some mixture. Various accounts of how the moral buck might stop earlier than the causal buck and in the right spot – me – have of course been advanced by physicalist substance monists keen to accommodate moral responsibility to their worldview. So, for example, one might say that if my body does what I want it to do as a result of me wanting it to do that thing, then that's me being morally responsible for the doing of that thing, and the fact that my wanting it to do that thing rather than something else was itself caused by factors beyond my control does not detract from that. This account is open to easy counter-arguments, but there are of course much more sophisticated accounts. However, they all suffer from the common feature that whatever psychological states are posited as sufficient to lead to the agent being morally responsible, it seems possible to imagine a skilled enough hypnotist inducing those states in a person and yet we not hold such a victim of such hypnosis to any extent responsible for the actions that then flowed from these states. In cases where we can identify causal responsibility, moral responsibility, we think, falls straight through to it; we are

strongly committed to the causal and moral buck stopping in the same place. Substance dualism of the interactionist sort is the view that accommodates this strong commitment in offering a 'third way' to causal necessitation of the physical sort stretching back beyond our births and randomness: my actions are caused by me (Descartes) or my mental part (Aquinas). Substance dualism, then, gives the best theory of freedom by reference to our commonsense intuitions about ourselves as being the initiators of and thus morally responsible for our actions.⁸

Consciousness

The classic thought experiment here concerns someone called Mary, who, we are asked to imagine, has been brought up in an entirely black and white room. In this room she has access to black and white science textbooks and science is now completed. She thus learns everything there is to know about the physical properties of colour and indeed, let us say, about the physical properties of brains too. She then leaves the room and goes into the outside world. For the first time, she herself sees a red apple. Is it not plausible to suppose of her that she thereby learns something new: what red looks like? We may call this new fact a fact about red qualia: what it is like to see red. From the fact that Mary – *ex hypothesi* – knew everything about the physical qualities of the colour red and the brain prior to leaving her room yet did not know about this 'qualitative', as we may call it, property, so we can conclude that this qualitative property is not a physical property of red or the brain; of what is it a property then? The substance dualist has a ready answer: of red as it is experienced by the soul.

There have been various physicalist responses to Mary-type thought experiments; they tend to deny the fact that Mary comes to know about a qualitative property; rather, they tend to assert, she comes to have an ability which she did not previously have, the ability to recognize red objects in a new way.⁹ This however seems wrong-headed to me, for Mary plausibly will not gain the ability to recognize red objects simply by getting out of the room and seeing a red apple for the first time. She will only gain that ability once someone provides her with information in the following manner: 'That apple you're looking at, Mary, it's red'. In hearing someone say *that*, she will plausibly gain a new ability to recognize red objects thereafter, but she had *already* come to know what red objects looked like prior to hearing someone say that, just by looking at the red apple. She wouldn't say back to the person who'd just said this to her, 'Now, for the first time, I know what red is like'; she'd say something like this: 'Now you've told me that that apple is red, I realize that I already knew – just by looking at it – what it was that red was like, rather than what it

was that blue was like, and so on. But although I didn't know that it was red, the qualitative nature of which I knew about by looking at the apple prior to your telling me, it was red that I had discovered something new about simply by looking at the apple'.¹⁰

Of course the analysis presented here has been, perforce, terribly brief (cannot *property* dualism deal with the issue of consciousness to which we have recently adverted?), but, even so, it appears that the 'facts' of personal identity, freedom of choice and consciousness, as they present themselves to commonsense are, when taken together, easily accommodated by substance dualism and fail to be accommodated by physicalism. Either these facts are not facts at all – commonsense is wrong – or physicalism is wrong.

Conclusion

If we suppose a substance/property structure to our metaphysics of the mind and we suppose that there is physical stuff (two suppositions I have not called into question in anything but the most oblique way here), then reasons of ontological economy alone would suggest that we should believe that we do not have souls, that our mental life could in principle be explained in terms of our physical. Such a world would be simpler – to the tune of one whole class of substance – than the world posited by substance dualists. physicalist substance monism has simplicity on its side when compared with substance dualism, but, as we have seen, it does not seem to have anything else; there are no other reasons for thinking substance dualism false. While simplicity is a virtue, so is explanatory adequacy, and there are things that we have reason to suppose a physicalism cannot explain. That current natural science cannot explain something is of course in itself very slight reason to suppose that future science will not be able to explain it, but there are at least three areas where, it has been argued, we are able to detect difficulties in principle. First, the 'facts' of personal identity as they are presented to commonsense seem to suggest that we as persons are (Descartes) or are constituted in part by (Aquinas) units of substance which are indivisible over time, and souls are the best candidate for such. Secondly, the 'facts' of freedom of choice as they are presented to commonsense – roughly that the causal and moral responsibility bucks stop in the same place – can be accommodated by substance dualism, but not by physicalism. And finally, what we have called the 'qualitative' facts of consciousness, what it is like to see things like red, are not reducible to facts about the physical properties of colours (or indeed colours and brains), something which again can be accommodated by substance dualism but not physicalism. Our discussion of all these points has perforce been very brief, but I hope sufficient to suggest reasons for this analysis. If so, one might sum up

our findings thus: there's one argument against substance dualism (it's more complex) and three in favour (it better explains personal identity, freedom of choice, and consciousness). If that is so, neither substance dualism nor physicalist substance monism will give us everything we want, and we shall naturally turn to considering how we should weigh simplicity against these other considerations when deciding what we have, on balance, most reason to believe.

Moore has taught us that we may take any valid argument in either of two directions, as articulating a reason to suppose its conclusion true or a reason to suppose one or more of its premises false, and that the direction in which it is most reasonable to take a given argument will depend on whether the premises are jointly more obviously true than the conclusion is obviously false. So we may give the considerations presented here some direction by finally asking ourselves this question: Knowing now that you can only believe one, which of the following seems more obviously right to you?

- We are persons in more or less the same way that commonsense suggests; we have freedom of the sort supposed in everyday life; and colours – and indeed mental happenings in general – have qualitative properties.
- The world is as simple a place as physicalism suggests.¹¹