

DUALISM INTACT

Richard Swinburne

I have argued in many places that a carefully articulated version of Descartes' argument to show that he is essentially an immaterial soul is sound. It is conceivable that I who am currently conscious continue to exist without my body, and that can only be if there is currently a non-bodily part of me which alone is essential for me. Recent counter-arguments of Alston and Smythe, Moser and van der Nat, Zimmerman, and Shoemaker are rejected.

An argument which I gave in several places¹ in the 1980's in favour of substance-dualism has generated a number of purported refutations in the 1990's. It is time for me to defend myself.

The argument is a modal argument, similar to that given by Descartes² and intended as an improved version thereof. I call the part of the human person composed of ordinary matter their body; and any part (if there is one) which is immaterial their soul. The argument claims that it is logically possible that my body could suddenly be destroyed at an instant and yet (whatever else compatible with this might now be the case), I could continue to live a conscious life. (There is not supposed to be anything peculiar in this to me—the same applies to any other human person now conscious.) Yet if I am to continue to exist, some part of me must continue to exist; and so if it is to be possible for me to continue to exist without my body, I must already have another part whose continued existence constitutes my continued existence—viz. my soul.

I argued for the claim that if I continue to exist, some part of me must continue to exist, from a more general principle which I called the quasi-Aristotelian assumption. This is that for the continuity of a thing, a substance, of which properties are predicated, the substance has to continue to be made at least in part of some of the same stuff as it was made previously; and so some part of the original substance made of that stuff must continue to form part of it. I call this latter claim the 'quasi-Aristotelian assumption' because, unlike a strict Aristotelian assumption it does not explicitly claim that the stuff of which any substance is made and some of which has to be preserved is ordinary matter—it does not rule out the possibility of immaterial stuff. We clearly accept this assumption for inanimate material objects, where the only stuff conceiv-



ably involved is ordinary matter. If a desk after 2 p.m. is to be the same desk as one up to and at 2 p.m., some part of the subsequent desk must be the same. I am happy, for the purposes of the present argument, to allow that the identity of a substance can be preserved with gradual replacement of parts over time, until there has been a total replacement. The full grown oak tree is the same tree as the sapling if the former is obtained from the latter by gradual replacement of parts (i.e. atoms), even if none of the original parts remain. But since my thought experiment concerns the situation immediately before and immediately after an instantaneous destruction of a body, the possibility of identity being preserved by a gradual replacement of parts does not arise. The kind of 'immaterial part' which, given the quasi-Aristotelian assumption, my Cartesian argument shows to exist is what we may call a Bonaventurian soul, because St. Bonaventure thought that the human soul consisted of 'spiritual' stuff (*materia*) informed by a form.³

To avoid misunderstanding and suspicion that I had committed some modal fallacy, I stated my argument in logical symbols, as well as in words. The argument was originally designed to prove that I have a soul in 1984, and I leave it in that form. Updating is always possible for any year in which Premiss 1 is manifestly true. Likewise any name or other referring expression can be substituted for 'I', so long as Premiss 1 remains manifestly true. I use the usual logical symbols—'&' as 'and', '-' as 'not', '◇' as 'it is logically possible that'. I define:

- p = 'I am a conscious person and I exist in 1984'.
- q = 'My body is destroyed in the last instant of 1984'.
- r = 'I have a soul in 1984'.
- s = 'I exist in 1985'
- x ranges over all consistent propositions compatible with (p and q) and describing 1984 states of affairs.
- '(x)' is to be read in the normal way as 'for all propositions x'.⁴

The argument is then as follows:

pPremiss 1
(x) ◇ (p & q & x & s)Premiss 2
- ◇ (p & q & -r & s)Premiss 3

Premiss 2 says that it is possible that I survive into 1985, given that I am conscious in 1984, even if my body is totally destroyed and whatever else might be the case in 1984, compatible with these last two suppositions. Premiss 3 says that it is not possible that I who am conscious in 1984 survive into 1985 if my body is totally destroyed, unless there is a non-bodily part of me in 1984, viz. a soul. It follows from Premiss 2 and Premiss 3 that -r is not within the range of x. But since -r describes a 1984 state of affairs, it follows that it is not compatible with (p & q). Hence (p & q) entails r. But the addition to p of q, which describes what happens to my body at the end of 1984 can hardly affect whether or not p entails r. So I conclude that p by itself entails r. Hence, from Premiss 1, r.

All my opponents acknowledge the validity of the argument, but dispute either Premiss 2 or Premiss 3, normally the former. I begin by considering objections to Premiss 2. Most materialists these days are happy to allow that it is logically possible that a person be non-embodied or even disembodied. But the most common criticism of Premiss 2 is that I have slid carelessly from this *de dicto* claim to the *de re* claim that it is logically possible that I become disembodied, and that there is no reason for believing the latter. I think that a fair reading of the relevant texts will show that I explicitly affirmed the *de re* claim and that all my arguments were designed to substantiate it. However, I acknowledge that there were some rather loose initial softening-up sentences which could, outside this context, be interpreted differently.

Among those who accuse me of sliding from the *de dicto* claim to the *de re* claim are Alston and Smythe.⁵ They accuse me of confusing my premiss (2) with their (2A):

Given p, q, and any consistent proposition about 1984 that is consistent with p and q, it is logically possible that s.

In my symbols, this reads:

(x) { (p & q & x) -> \diamond s }

In other words, p & q and any 1984 state of affairs, compatible with p and q, entails the logical possibility of s (but not of course the actuality of s)—just as my being unmarried at t entails the logical possibility of my being married at t, though not of my actually being married at t.

Their argument now gets muddled because (2A) does not say what their verbal exposition of it claims that it does, viz. “that, given p and q, no matter what is the case in 1984 (provided it is consistent with p and q), it is still logically possible that I continue to exist (in a disembodied state, naturally) in 1985.” Put symbolically, however, this reads:

(x) { p & q & x -> \diamond (p & q & s) }

Lets call this (2B). (2B) captures the ‘continue’ to exist, i.e. exist in both 1984 and 1985, and the ‘disembodied’ state of 1985, of their non-symbolic verbal account which 2A does not capture. They state immediately after giving what they suppose to be its non-symbolic expression that “2A seems to be clearly true”. So I take it that they are referring not to the obviously trivial 2A, but rather to 2B. However (2B) is still not enough. I need (2), and I gave arguments for it in *The Evolution of the Soul* (pp.151f).

The only arguments which can be given to show some supposition to be logically possible are arguments which spell it out, which tell in detail a story of what it would be like for it to be true and do not seem to involve any contradictions,⁶ i.e. arguments from apparent conceivability. Apparent conceivability is evidence (though not of course conclusive evidence) of logical possibility. The arguments have the form: take any

actual human person, currently conscious in 1984, yet having his body destroyed at the end of 1984, conceive of anything you like (compatible therewith) being the case with his body during 1984, or happening to the bodies of other people, to the proportion of nitrogen in the surface of Jupiter or whatever else you choose in 1984; it is still conceivable that he goes on existing in 1985. If anyone does not see that at first, a story can be told in a lot more detail of what it would be like for it to be true, which would help the reader to see it. But of course whatever suppositions x we make about 1984 must be compatible with $(p \ \& \ q)$.

Like all worthwhile arguments, mine purported to start from premises which many an opponent might grant—viz. (1), (2), and (3) as they stand—to establish a conclusion which he did not previously recognise. I suggested that most people not already having a firm philosophical position on the mind/body issue will grant my premises. But someone already having a firm philosophical position contrary to mine can challenge my premiss (2) by inserting an x which he claims to be compatible with p and q and which he claims will show the premiss to be false, where x states a philosophical thesis about the very issue in dispute, contrary to the one which I am seeking to prove. Examples include “I am purely material in 1984” of Alston and Smythe, or “I am identical with my body or some part of it” of Zimmerman.⁷ Now of course I claim that no such x is compatible with $(p \ \& \ q)$. Since I put forward premisses (2) and (3) as purported necessary truths, my argument was designed to show that (given p) r is a necessary truth. The claim therefore that any x of the above type is compatible with $(p \ \& \ q)$ amounts to the denial of my conclusion. Now it is true that my argument will not convince anyone who claims to be more certain that the conclusion is false than that the premises are true. But then that does not discredit my argument—for no argument about anything will convince someone in that position. My argument was designed for those prepared to set aside philosophical dogma concerned explicitly with the mind/body issue, and rely only on philosophical theses and intuitions about logical possibility relating to other or wider issues.

In the earlier part of their paper Alston and Smythe make what is in effect a claim that no mere assertions about logical possibility could have any tendency to show what I in fact am (e.g. that I am not only a body). But given the quasi-Aristotelian assumption (which they only factor into this part of their discussion briefly at the end), their claim must be false. For any substance, there must be something which makes it that substance rather than any other one. A substance has an essence. The quasi-Aristotelian assumption says that the stuff of which a substance is made belongs to the essence of the substance; it is involved in what makes the substance that substance. So if I am only a body, the stuff that makes me me must be ordinary matter. That the stuff which made me me is ordinary matter would then be no mere contingent truth about me; it would be necessarily true. Hence a crucial necessary truth would follow from “I am only a body” that I could not exist without my body (whereas no similar crucial necessary truth follows is the analogy on p.129 of Alston and Smythe from “I am a bachelor”.) Since (given my Premiss 2) the purport-

ed necessary truth is false, so too must be the claim that I am only a body. My argument thus has the same pattern as a well-known argument of Kripke where he argues that the actual constitution of a substance determines the logical possibilities for its constitution:

Supposing this lectern is in fact made of wood, could this very lectern have been made from the very beginning of its existence from ice, say frozen from water in the Thames? One has a considerable feeling that it could *not*.⁸

A bolder and more general principle was invoked by Shoemaker, to show that no thought experiments could establish Premiss 2:

In general, it is quite hopeless to suppose that a claim of *de re* possibility, a claim to the effect that some actually existing thing could undergo such-and-such changes, can be grounded on mere thought-experiments, or on considerations of what can be supposed or imagined without logical or conceptual incoherence.⁹

Shoemaker is correct in supposing that no claim of *de re* possibility can be grounded *merely* on thought-experiments. But in supposing that this point counts against my argument, he is quite mistaken. For my claim is not so grounded—it has a crucial contingent first premiss. And if Shoemaker is claiming that thought-experiments cannot play a crucial role in determining *de re* possibility, he is also mistaken—as the Kripke example illustrates. What happens is that the thought-experiment shows that a certain kind of thing can or cannot undergo certain changes. Then some contingent premiss states or entails that a particular object *a* is (either necessarily or contingently) a certain sort of thing. And then it follows that this or that can or cannot happen to *a*. A contingent premiss tells us that Hesperus is a planet. Thought-experiments show that for any planet to continue to exist it must retain (much of) the same matter. We then conclude that Hesperus cannot lose all its matter. But if the contingent premiss told us that Hesperus is a ghost, that conclusion would not follow.

What is, however, more dubious is my Premiss 3, or rather—as I acknowledge in *The Evolution of the Soul*—the quasi-Aristotelian assumption which I use to argue in support of it. This assumption, as I claimed earlier, clearly states a necessary condition for the identity over time of inanimate material objects—where the only kind of stuff at stake is ordinary matter. But given my premiss (2), I wrote in *The Evolution of the Soul* (p.153) we have a choice: “Either we can say simply that persons are different—in their case continuing matter is not necessary for the continued existence of the substance, or we can try to make sense of this fact by liberalising Aristotle’s account”- (that continuing ordinary matter is necessary for the identity of substances) and allow a non-material stuff. My adoption of the latter alternative was in a way a concession to the materialist: it was trying to preserve as much as possible of his “integrated system of thought” (my p. 153), i.e. a hylemorphic theory,

which says that the identity of all substances depends on the stuff of which they are made, and the form imposed upon it—that is, the properties essential to a substance of that kind. A substance is the same substance as an earlier one if it is made of the same stuff (approximately, or stuff obtained by gradual replacement thereof) and if it retains the essential properties of a substance of the kind in question. (If the original substance is a tree, the later substance can only be the same substance if it is also a tree.)¹⁰ My quasi-Aristotelian assumption kept this basic idea but allowed the possibilities of non-material stuff (and so non-material substances made solely thereof). It thus allowed a relatively unified account of the identity of substances, inanimate and animate.

Now it is indeed true that there are many philosophers who wish to give a very different account of the identity over time of persons (and we may suppose, more generally—of animate beings) from the account they give of inanimate substances. They maintain that whether a later person is the same as some earlier person depends solely on whether the later person has the right relational properties, i.e. exhibits the right relations to the earlier person (when being composed of such and such stuff or parts is not a property). A Humean Theory of personal identity for example makes the identity of a later person with an earlier person depend on such features as the later person having the same apparent memories as, and apparent memories caused by, those of the earlier person. Apparent memories are properties possessed by a substance; they are abilities apparently to recall. The main burden of the criticism of me by Moser and Van der Nat¹¹ seems to be that in putting forward my quasi-Aristotelian assumption, I have not taken seriously the possibility of a relational theory of personal identity.

The reason which I gave implicitly in *The Evolution of the Soul* for not exploring such possibilities is that we need an integrated theory of what makes for the identity of substances over time, and that a relational theory of personal identity made the identity of persons depend on very different factors from the identity of inanimate substances. However, I should perhaps have spelled out there the well-known objections to any relational theory of personal identity, which I did spell out in *Personal Identity* with respect to the best known kind of relational theory—a memory and character theory.

These objections boil down to the following. Any relational theory of personal identity ever proposed either has the consequence that indefinitely many later persons can satisfy the criteria for being the same person as a certain earlier person (the duplication objection), or has the consequence that which later person is the same person as the earlier person depends on what happens to a third person (the arbitrariness objection). The duplication objection will apply if more than one later person can have the relevant relational properties (e.g. 'psychological continuity' with the earlier person). And of course any such account must be false—for the well known reason, that if two later persons are the same as the earlier person they would be the same as each other—and, by hypothesis, they are not. The duplication objection is normally met by insisting that the relational properties include one which can only be sat-

isfied (in each possible world) by at most one person. The trouble is that the crucial property offered is always one which involves the later person exemplifying some relation to the earlier person in some way better than other later persons—e.g. by exemplifying a relation to a greater degree or earlier in time than others. Thus a writer may claim that a later P2 is the same as an earlier P1 only if P2 has the property of being the first person causally connected with P1 to have similar apparent memories; or is in other ways the “closest continuer”¹² of P1. But then whether P2 is the same person as P1 will depend on what happens or does not happen to persons other than P2 e.g. whether P1 causally generated some other similar person before generating P2, or whether some person who would otherwise have been the “closest continuer” dies prematurely. And is it not absurd to suppose that who I (P2) am depends on what happens to someone else? Whether I am elected to some office may depend on how many votes are given to someone else, but that who I am can depend on such extrinsic factors seems absurd. This is the arbitrariness objection—my identity becomes an arbitrary matter.

There are philosophers who are prepared to accept that (in the stated sense) personal identity is an arbitrary matter. Parfit would claim that it is an arbitrary matter but that does not matter, because identity does not matter. What matters is survival. Survival, unlike identity, can be duplicated. A person can have many surviving selves, in which he survives to different degrees.¹³ But the issue is—can any intelligible account be given of what it is for me to “survive” except in terms of a later person being the same person as the earlier one? Many writers have pointed out that for inanimate things, clouds and countries and armies, identity clearly is an arbitrary matter. Whether a later country is the same as an earlier country will often depend on what happens to bits of territory not included in the later country, etc. But the point is just that these are inanimate things. If I undergo some operation, or my brain state is copied in some way before my body is destroyed, there is clearly a truth here about whether or not I have survived this process. Before the process I seek my continued existence. Afterward we may not know whether my endeavours have been fulfilled. But only someone already in the grip of a strong philosophical dogma, could deny that there is a truth about whether or not I have survived an operation or brain-copying process; which does not depend on what happens to someone else. He who hopes to survive his death will not be satisfied by the knowledge that someone very like him will live again.

It was these implausible consequences of every proposed relational property theory of personal identity, which—among other reasons—led philosophers to seek a criterion of identity which did not permit duplication and was not open to the arbitrariness objection. Such a criterion will make who someone is depend solely on non-duplicable factors intrinsic to him. Some sort of Aristotelian or quasi—Aristotelian principle will have the desired result. If you insist that the later person has to have (most of) the matter, or (most of) the brain matter of the earlier person, there will be no duplicates; and who is who is not (in the above sense) arbitrary, for the criterion of having (most of) the (brain) matter of

P1 is a criterion whose satisfaction depends on what is the case with P2, quite independently of what has happened to any other person. And if ordinary matter will not—given my Cartesian argument—provide the requisite satisfactory criterion then immaterial stuff (which, because it is not space-occupying and so not necessarily extended, we may suppose to come in indivisible units, souls) will provide the requisite criterion. All of this provides the reason for holding the quasi—Aristotelian assumption, as well as the fact that it provides criteria of identity for animate beings as similar as can be to those for inanimate beings.

There is, however, one kind of relational theory of personal identity which is immune to the duplication and arbitrariness objections. But it is one which I suspect the average modern analytic philosopher would be even less happy to adopt than he would be to adopt the Bonaventurian theory given above. Any relational theory of personal identity is bound to fail if the only properties considered are universals, i.e. properties which could (maybe only in a different possible world) be instantiated in different individuals. For just because it is contingent in which person the property is instantiated, the property cannot suffice to make the person who he is. We normally think of properties as universals, and may indeed make it a matter of definition that to be a property something has to be a universal¹⁴; but if we do thus make it a matter of definition, there can be no viable relational theory of personal identity. Clearly any relational theory open to the duplication objection has made personal identity depend on universals. And if a theory is open to the arbitrariness objection, this will be because a different person could have had the suggested identifying property. If relational properties are to provide the criterion of identity over time, they must be non-universals.¹⁵ The theory, that is, must state that to be P1 a later individual must have a certain relation (R) to P1 which can be possessed by no other individual than P1 in any other possible world. The relation must hold solely in virtue of factors intrinsic to P1, if the criterion is to avoid the arbitrariness objection. Once we allow non-universal properties, one such could be the property of being identical with P1, for in no possible world could any individual other than a certain individual (viz. P1) have that property. Such a property we may call an individual essence. All other non-universal properties are properties whose possession entails the possession of a certain individual essence—for example the property of remembering doing what P1 did in 1984 (to be distinguished from the property of apparently remembering what P1 did in 1984). Any one who has the former property will have P1's individual essence, but not conversely—P1 may now suffer from amnesia. So the crucial relational "property" is P1's individual essence. P2's having that "property" is both necessary and sufficient for his being P1. If we are prepared to allow individual essences, personal identity can certainly be analysed in terms of possession of relational properties. A later person is the same person as an earlier person if and only if he has the same individual essence. Who an embodied person is depends on which individual essence is coinstantiated with his other properties.

Given the logical possibility for which I agreed earlier in the paper,

that any embodied person can become disembodied, there are two possible theories of how this could come about. One is that there is again immaterial stuff, instantiation of his individual essence in which constitutes the continuing existence of a disembodied person. But on that theory we would have two kinds of new thing—immaterial stuff and individual essences. It is simpler to suppose that the individual essence can exist on its own without being instantiated in any stuff (exist that is in the full-blooded way in which substances exist, not in the pale way in which properties such as squareness “exist” when there are no square objects). It would be a substance on its own, whether or not instantiated in ordinary matter. Individual essences are very strange properties indeed.

Duns Scotus seems explicitly committed to individual essences; and—though he had not thought through his views consistently—Aquinas was also, I believe, implicitly committed to them. Aquinas affirms that the human soul (which is a form, and so a property) is what makes an individual human person the person he is. When the soul of Socrates is instantiated in a body, we have Socrates. But Aquinas also held that the soul of Socrates could exist without the body and live some sort of a mental life, though it would not on its own be a human person. The soul of a human was thus a “subsistent form” and in *Summa Contra Gentiles* though not in *Summa Theologiae* he calls it a substance.¹⁶ (If we understand by a substance something which can exist on its own, that is what he must say that the soul is. But he does not always thus understand it.) So we can do without Bonaventurian souls (parts of persons made of immaterial stuff), if we allow Thomist souls—non-universal properties which are also substances. I think that there is a lot to be said for Thomist souls,¹⁷ but—suspecting that a modern philosopher would be marginally more tolerant of Bonaventurian souls—I commended the quasi-Aristotelian principle. But if you drop it and allow individual essences which can exist by themselves, then, when they are joined to matter to form embodied humans, they are substantial enough (not merely universals which cannot exist un-instantiated) to be parts of those humans (though not parts made of stuff), and so Premiss (3) is true for different reasons. So again, the argument goes through. It remains quite undefeated by any of the counter—arguments which any objector has raised.

University of Oxford

NOTES

1. Principally in my *The Evolution of the Soul* (Oxford: Clarendon Press, 1986) Ch. 8 and Additional Note 2; but also in S. Shoemaker and R. Swinburne, *Personal Identity* (Oxford: Basil Blackwell, 1984) ‘The Dualist Theory,’ Ch. 2; and in ‘The Structure of the Soul’ in (ed.) A. Peacocke and G. Gillett, *Persons and Personality* (Oxford: Basil Blackwell, 1987).

2. *Meditation* 6.

3. St. Bonaventure, II *Sentences*, 17.1.2, responsio

4. Unfortunately there was a misprint in the original version of this sentence in the text, where “states” was printed instead of “propositions”. Alston and Smythe (see below) drew attention to the need for this correction.

5. William P. Alston and Thomas W. Smythe, “Swinburne’s Argument for Dualism”, *Faith and Philosophy* (1994), pp. 127-133.

6. As I argue (with a qualification) at much greater length in *The Coherence of Theism* revised ed. (Oxford: Clarendon Press, 1993) Ch. 3.

7. Dean W. Zimmerman, “Two Cartesian Arguments for the Simplicity of the Soul”. *American Philosophical Quarterly* 28 (1991), pp. 217-26.

8. Saul Kripke, “Identity and Necessity” in (ed.) Milton K. Munitz, *Identity and Individuation* (New York: New York University Press, 1971) p.152.

9. Shoemaker and Swinburne op. cit. p.144.

10. There is of course a crucial issue (which we need not resolve) about the minimum essential kind which has to be preserved—e.g. if *a* is an oak tree, and *b* is a later tree made of the same matter as *a*, does *b* have to be an oak tree in order to be the same tree as *a* or is being a tree sufficient? But, whatever are the criteria for minimum essential kinds, that does not affect the issue at stake here, which is whether the hylemorphic theory, however interpreted, should be accepted.

11. Paul Moser and Arnold van der Nat, “Surviving Souls”, *Canadian Journal of Philosophy* 1993 (23), pp. 101-106.

12. See R. Nozick, *Philosophical Explanation* (Oxford: Clarendon Press, 1981) Ch. 1.

13. Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984) Ch.12.

14. If we do this, it *might* of course follow that not all conjunctions of properties are properties; but it will remain the case that all disjunctions of properties are properties.

15. Or to be totally precise—they could be universal disjunctive properties having a non-universal property as one of the disjuncts. Thus “(being identical with P1 in a world where P1 exists) or (being square in a world where P1 does not exist)” is a universal property, because in different possible worlds it can be possessed by different individuals. But to acknowledge such a disjunctive property as a property involves acknowledging each of the disjuncts, including the first one, as a property; and the first disjunct is a non-universal. So, either way, we are saddled with non-universal properties.

16. For Aquinas’ views see *Summa Theologiae* Ia. 75 and 76 and *Summa Contra Gentiles* 2, 46—end. One tension in Aquinas’ account arises from his claim that a soul is individuated by the body which is ‘fitted’ to occupy (see SCG 2.81). Unlike normal forms, Aquinas held, it could fit only one body. Why that should be is however quite mysterious, if a soul is capable of existing without a body. Why should it not be joined again to a new but perhaps qualitatively similar body? And it is difficult to see how a soul could be fitted to occupy a certain body unless there was something intrinsic to it which made it different from other souls, and which would therefore be sufficient to individuate it. Scotus made this latter criticism in *Ordinatio* II d 3 p 1 q7 nn 230-1. For full analysis of the views of this issue of Duns Scotus, and of the internal tensions within Aquinas’ account of the soul, see my *The Christian God* (Oxford: Clarendon Press, 1994) pp. 47-50. (In line 2 and line 4 of Additional Note 3, read ‘subsistent forms’ instead of ‘substantial forms’.)

17. And indeed have defended it and put my account of personal identity in terms of it in *The Christian God*.