RESPONSE TO REVIEWERS

RICHARD SWINBURNE

University of Oxford

Many thanks to the Editors of *The European Journal* for organising this symposium, and to all the reviewers for the work involved in grappling with my arguments.

I. INFORMATIVE DESIGNATORS AND PROPERTY DUALISM

I claimed that the history (in an objective sense) of the world is just the succession of events (which – briefly – are the instantiations of properties in substances at times), and that we can tell it in many different ways by cutting the world up into substances and properties in different ways; but that to do so we need to pick out substances and properties by informative [rigid] designators. I defined an 'informative designator' as follows:

For a rigid designator of a thing to be an informative designator it must be the case that anyone who knows what the word means (that is, has the linguistic knowledge of how to use it) knows a certain set of conditions necessary and sufficient (in any possible world) for a thing to be that thing (whether or not he can state those conditions in words.) Two informative designators are logically equivalent if and only if they are associated with logically equivalent sets of necessary and sufficient conditions. To 'know' these conditions for the application of a designator – as I shall understand this expression – just is to be able (when favourably positioned, with faculties in working order, and not subject to illusion) to recognize where the informative designator (or, if it is defined in words, the words by which it is defined) applies and where it does not and to be able to make simple inferences to and from its application. (*Mind*, *Brain*, *and Free Will*, p. 12)

I illustrated the difference between an informative designator such as 'red' and an uninformative one such as 'water' (as understood in the eighteenth century). We know what it is for a thing to be red because we can always recognise red things under the stated conditions; but in the eighteenth century when to be 'water' meant to have the same essence as the stuff in our rivers and seas, we could only recognize water when it was in our rivers and seas. With aid of this distinction I argue that we would leave out an important part of the history of the world unless we included in that history both pure mental events (sensations, beliefs, etc.) to which subjects have privileged access and physical events which are publicly accessible events (or events which entail the occurrence of events of both kinds), and both pure mental substances and physical substances (or substances which entail the occurrence of both kinds).

I have quoted my definition of 'informative designator' at length, because it seems to me that neither Lynne Rudder Baker nor William Jaworski have understood it fully or grasped its significance. Baker's first objection (p. 8) to the utility of this concept is that since on my view that 'I' is an informative designator which designates a mental substance, everyone who uses 'I' competently should believe that they are a mental substance, and of course they don't. But Baker has not appreciated the sentence of my definition which states that to 'know' the conditions for the application of the designator 'as I shall understand this expression' just is to be able to recognize when it applies and when it does not; knowing the necessary and sufficient conditions for a thing to be that thing is 'knowing the nature of what I am talking about'. I use the point about 'I', together with a claim about the logical possibility of certain thought experiments, to argue over several pages for the truth of the philosophical thesis that I am a mental substance. To do this I use arguments which, even though in my view they are sound, not everyone will believe to be sound, and so not everyone will believe the philosophical thesis. Yet clearly someone can know something without believing all its logical consequences. Jaworski has a similar general worry - he writes (p. 24) that 'even if I am necessarily right about what "I" refers to, I can still be wrong about what essential properties I have. But on my definition of knowing a set of conditions necessary and sufficient for the application of a designator, 'being necessarily right' about what it refers to is knowing one set of essential properties for being me. There are many different logically equivalent sets of necessary and sufficient conditions

for anything to be what it is; I only need to know one particular set in order to use the word 'I' correctly. But knowing one set of conditions for applying 'I' is quite sufficient to rule out the possession of any physical properties as metaphysically necessary – in virtue of my arguments to show that physical and pure mental properties and so the existence of physical and pure mental substances never entail each other. Of course there may be physical substances with physical properties which are *causally* necessary and sufficient for my existence – but that is irrelevant to my argument.

Then Baker claims (p. 8) that she knows how to use the word 'arthritis' but doesn't know necessary and sufficient conditions for its application. But the meaning of a technical term is determined by a group of experts' (my p. 10), and on the assumption that the experts do know the necessary and sufficient conditions it will be an informative designator; otherwise it is an 'uninformative designator'. And then Baker claims (p. 8) that she does not know the necessary and sufficient conditions for something to be 'red' because of doubt about whether reddish-orange things and such like are red. But a competent language speaker will know that a reddish-orange thing is a borderline case of being red and so (given the falsity of the epistemic theory of vagueness) will know that the necessary and sufficient conditions for being red are not satisfied, and will also know that the necessary and sufficient conditions for being not-red are not satisfied. Baker is mistaken in supposing (p. 8) that on my view whether a word is an 'informative designator' normally varies with the speaker. My view is that words have a 'correct use' (my p. 17) in a language; and normally whom or what is designated by a word is the same whoever uses it. The correct use today of 'water' by everyone is to designate H₂O. However what is designated by an indexical clearly varies with the speaker - what 'you' or 'here' designate varies with who use these words, when, and where. So what 'I' designates, while being always an informative designator, varies with the speaker. And, I argued, the correct use of 'I' or of my own name by myself is that of an informative designator; but that the correct use of my name by someone else is that of an uninformative designator. This is the sole case where whether a word is an 'informative designator' varies with the speaker. 'Richard Swinburne' as used by me is an informative designator, whereas as used by others it 'is an uninformative designator' (my p. 165). This is because 'Richard Swinburne' is used to refer to the actual person who has a certain body. But others do not know the necessary and sufficient conditions for being that person – since that person could now exist without a body or in a different body; and others would not necessarily be able to recognize that person under those conditions. They apply the word, as people applied 'water' in the eighteenth century, on the basis of conditions which are not necessary and sufficient for being that person. I also use 'Richard Swinburne' to refer to the actual person who has a certain body. But I, being aware of myself, am in a position to know who that person is, and so to know those necessary and sufficient conditions for being Richard Swinburne. I alone necessarily would be able to recognize myself now if I had a different body, or if I did not have a body at all. I conclude that none of Baker's criticisms of the utility of the concept of an informative designator have any force.

Jaworski has a similar package of objections to the way I apply the concept of informative designator. He claims (pp. 21-22), that showing that 'red' and 'reflects light of such-and-such wavelengths' are not logically equivalent and so pick out different properties, does not show that one is mental and the other is physical. He is correct in this claim. 'Red' is in my sense a physical property, because the substance in which it is instantiated – a surface – does not have privileged access to its instantiation in it; all observers (when favourably positioned, with faculties in working order, and not subject to illusion) have equal access to whether it is instantiated; and of course 'reflects light of such-and-such wavelengths' is also a physical property. I introduced this example in chapter 1 to illustrate how we need to distinguish between properties - by their informative designators not being logically equivalent; and I deliberately chose an example which was not an example where one property was mental and the other property was physical. But having introduced the concept of an informative designator, I proceeded to argue that there are innumerable properties individuated by informative designators which are such that the substance in whom they are instantiated has privileged access to their instantiation, and which are therefore not the same properties as ones which are such that the substance in which they are instantiated does not have privileged access to their instantiation. Contrary to Jaworski, I do not merely 'tacitly endorse' (p. 22) 'modes of presentation'; I explicitly claim that they are themselves properties – 'a mode of presentation ... is just as much a real characteristic of any object as any property' (my p. 26). My objection was to introducing into our ontology a category of 'mode of presentation' separate from the category of property. Certainly some properties might be distinguished from

each other by different modes of presentation (that is, properties distinct from the properties which they were picking out), but to suppose that all properties could be so picked out would lead to an infinite regress; some properties (those having informative designators) must be recognizable as what they are without intermediaries. However all substances, unlike all properties, may be picked out by different modes of presentation (i.e. different properties); and so I am not denying what Jaworski has to say (p. 22) about 'Cicero' and 'Tully' with the caveat that (except for their use by Cicero himself) they serve as uninformative designators.

Jaworski claims (p. 20) - correctly - that we may find ourselves in unusual conditions and call something 'red' which we would not call 'red' under ideal conditions and come later to recognize our mistake. Then he goes on to claim that if 'red' is like this, then it is possible that redness may be identical to a surface property whose essence we do not know even though we can competently use 'red' to refer to objects having it. But I cannot see why that is supposed to follow. To repeat – words mean what competent language users mean by them. And in my view 'red' is used to designate a certain aspect of the way things look (to most people); and the way things look is the sole determinant of whether they are red. If (implausibly) I am mistaken and 'red' is not used in that way, then we would need another word which is used in this way to designate that aspect (or words which entail that aspect) if we are to give a full description of the world. Jaworski suggests that while all speakers understand 'red' in the same sense, some competent speakers might pick out an object as 'red' by the way it looks, and other speakers might pick it out by the wavelengths of the light it reflects. 'Red' would then have two criteria for its application, each sufficient and only disjunctively necessary; red objects could be picked out by either of two separate properties. But again if that's how 'red' works then - to fully describe the world - we need an informative designator for each of the two properties. My definition that two rigidly designating predicates which are not logically equivalent designate different properties is not (p. 22) a 'tendentious assumption about properties'; it is a definition which I explicitly provide since there are alternative criteria which one could choose to use for the identity of properties. I chose this criterion because with the aid of it we can tell the whole history of the world by listing only a set of events (instantiations of properties in substances at times) which entail all events which ever occur. I am not denying, as Jaworski seems to suppose (pp. 23-24), that properties and substances

may have essential properties of which we are unaware; but I do claim that – if so – those essential properties must be entailed by the ones of which we are aware when we pick them out by informative designators. To deny that is to use different criteria for 'same property' from the one which I have good reason to propose.

II. COMPOSITE SUBSTANCES

Baker criticises my principle of the identity of composites. One way to express the principle, she claims (pp. 9-10), is 'there is no more to any substance than its parts (e.g. fundamental particles) and the way those parts are arranged'. But – crucially my principle, as I stated it (my p. 35) is 'there cannot (logically) be two things which have all the same parts having all the same properties, arranged in the same way. 'Having all the same properties', is crucial (and indeed makes 'arranged in the same way' redundant). Among the properties of a thing are (my p. 5) its relations to other things including (see e.g. my p. 33) properties of spatio-temporal continuity with earlier things. Hence Baker's drinkalator - carburettor example (pp. 9-10) does not count against my principle. All the parts of a drinkalator have the property of being part of a device used for making soft drinks; and the parts of a carburettor do not have this property. All that is ruled out by my principle is that the world could be different if instead of one such machine there was another such machine made of the same parts with the same properties (including the same past-related properties) arranged in the same way. Then I claimed that we can tell the story of the ship of Theseus either by using the 'same planks' criterion or by using the 'gradual replacement' criterion for the identity of the ship 'without anything being omitted'. Baker objects (p. 9) that which later ship is the original ship makes a great difference to the history of the world, because it affects who owns which ship. But on both accounts it may be part of the story that the courts determined which of these ships was the ship of Theseus. But (barring the - to my mind - implausible epistemic theory of vagueness) the courts did not discover the answer to a deep metaphysical issue; they merely determined which criterion should be used for determining who owned which of the two subsequent ships. Once we know what happened to all the planks and what the courts decided, we can still tell the story in two mutually entailing ways. I doubt if the decision of a court (e.g., that 'same planks' determines same ship)

is enough to settle how the expression 'same ship' should normally be used; but even if it does settle it, we can still tell the whole story by using a different criterion (e.g., gradual replacement) from that used by the courts, adding that the courts called the ship formed of the same planks the original ship and so determined that the owner of the original ship now owned, not the original ship but the ship formed of the same planks.

III. MENTAL SUBSTANCES

A mental substance is – on my definition – one for the possession of which some mental property is essential. I argue that we humans are mental substances because we co-experience at a time and over time different mental properties. Baker (pp. 10-11) outlines correctly the structure of my argument from co-experience at one time (e.g. of sensations of different kinds), which I call the argument from the synchronic unity of the human person. She doubts its validity because she suspects that I am using 'delimit' as a causal notion in (2), and that (2) will only support the ontological claim of (4) if it uses 'delimit' in a stronger sense. 'Delimit' is indeed being used in a causal sense in (2), but my claim is that the causal facts determine the ontological facts. What makes a brain my brain (ontological fact) is that my mental events are caused by, or cause, events in that brain (causal fact). It then follows that if I have any physical properties (as Baker and almost all of us hold), then necessarily a mental property of mine determines that certain of my brain properties are (contingently, because of the causal connection) among those physical properties. I claimed at this place in the discussion as elsewhere in the book, that we can cut the world up into substances and trace their continuities in different ways without leaving anything out of the history of the world. In particular we could regard the physical part of me as only a brain – after all, my brain in a vat is still me, or even as only part of a brain, in interaction with another substance, constituted by the rest of my body; and the history of the world which assumed this would entail and be entailed by its history described with our more normal categories. My point at this stage of the discussion, was that if we treat me as having any physical properties at all, certain brain properties must be among them in virtue of their causal relations to a mental property. A mental property determines the minimal set of physical properties which, if I have any physical properties, are essential if a substance is to be me at all. And so, whether or not any physical properties are necessary for

my existence, a mental property is necessary. Jaworski (p. 25) is right to claim that I have to hold that 'I have no physical properties other than those which are determined by my mental properties'. But – and I should have made this point explicit in the text – all my other physical properties (on our normal understanding of my physical boundaries) cause or are caused indirectly by (i.e. via a causal route) my brain properties which cause or are caused 'immediately' (my p. 143) by my mental properties – given that the causation is close enough to regard them as belonging to a single organism. What makes the other physical properties mine is determined (immediately or indirectly) by my mental properties. The rest of my brain is mine because events in it cause the events which are the immediate causes of my sensations, and so on. Even my hair is mine, because it goes where my brain events cause the rest of my body, to which it is attached, to go. So if I had no mental properties, neither would I have any physical properties. If I no longer had any conscious events or continuing mental states (e.g. beliefs and desires which I can have while not being conscious, and even when in a coma), I would not exist - on our normal understanding of what it is for 'me' to exist, which I was trying to analyse. So I reject Jaworski's suggestion (p. 25) that 'the continued existence of some ... mentally irrelevant parts [might] be sufficient for my existence. My corpse is not an existing me.

Although my argument from the diachronic unity of the human person is supported by my argument from its synchronic unity, it does not in fact need it - contrary to Baker's claim. In fact it consists of several connected arguments, none of which need the argument from synchronic unity. I opened the section on diachronic unity with the claim that (among pure mental events, to which – see my pp. 71-72 – my discussion thereafter of 'mental events' was almost entirely confined) all conscious events, such as having a pain or a thought last for a period of time. (As Baker points out, I wrote carelessly that 'all events take time' (my p. 148); I should have written 'all conscious events take time'.) I then argue that for a person to have a conscious event such as a pain lasting for a period (e.g. one second) is for that person to have a pain lasting for the first half of that period and also a pain lasting for the second half of that period. In being aware of one's one-second pain, one is aware of oneself as experiencing pain for the first half of that period and then experiencing pain for the second half of the period, and so of oneself continuing over time. Yet (given – as I argue later -that 'I' is an informative designator) that I have a pain (or any other conscious event) does not entail and is

not entailed by any physical event at all. So the substance involved in that event (myself) could - it is metaphysically possible - exist even if there were no physical events. That entails that I am a pure mental substance. What applies to me applies to other humans, and since the argument is simply an argument from a being being conscious to it having a soul as its one essential part, it applies to any conscious being. But all that that argument shows is that I have a soul as my one essential part for the 'specious present' of one conscious event. I then extend the argument to show that I have a soul as my one essential part for as long as I have a series of overlapping conscious events. I then appeal to various thought experiments to show the logical possibility and then the metaphysical possibility of any conscious being continuing to exist over intervals of not being conscious without that consciousness being dependent on any particular brain and then on any brain at all. Hence being conscious entails being such as not - metaphysically - to need any physical properties in order to exist. Since consciousness entails being a pure mental substance, not being a pure mental substance entails not being conscious. James Dew (p. 31) calls it an 'unstated assumption' that 'it is impossible for purely physical organisms to have conscious experiences'. I think that the argument for that is sufficiently explicit in the arguments from synchronic and diachronic unity summarized above to be found in my chapter 6.

Dew misunderstands my thought experiment in which a tenth of my brain is replaced each year until after ten years none of the original brain remains. He asks (p. 32) 'why couldn't it ... happen' that the same 'stream of consciousness - a first person perspective - is maintained by an everchanging physical organism'? If he means by the 'same stream', a ten-year stream of overlapping conscious experiences, that seems unlikely in view of the human need for sleep, some of which seems to be non-conscious. But if it did happen, the resulting person would indeed be the same person as the original person. But given a few intervals of non-consciousness in such an experiment, as I wrote, it still seems logically possible that the resulting person would be the same as the original person. But, I go on to say, it seems 'also logically possible' that the resulting person would not be the original person. I then go on to argue that what is logically possible is in this case also metaphysically possible. I then conclude that because each scenario is compatible with all the data about the physical parts, and the physical and mental properties of the original and resulting persons, by the principle of the identity of composites the final resulting person would need to have a same non-physical part (the same soul) if he is to be the same person as the original person.

Baker summarises her own view - a view of which Dew writes approvingly - that (to quote Dew, p. 33) as well as some sort of bodily continuity, 'all that is necessary for the persistence of the persons is continuity of first-person perspective' which is 'the activity one has to think of herself as herself'. The continuity of 'first-person perspective' is explicated as the continuity, not of a conscious thought, but of an 'ability to think of herself as herself'. But what is it for such an ability possessed by a person P, on waking up in the morning to be continuous with the ability possessed by the person P₁ who went to sleep the previous evening, and whose body was continuous with the body of P₂? If bodily continuity and/or continuity of memory and character is necessary and sufficient for the resulting ability to be continuous, we run into all the problems raised for the complex theory of personal identity by brain transplant thought experiments. For example, 'continuity of first person perspective' with the previous person could be possessed by more than one subsequent person; and so it cannot provide an answer to which (if either) subsequent person is the original person, and - unless personal identity is a matter of degree (a view which I gave arguments to reject on pp. 154-5), there must be an answer to this question. And the answer which I provide which is immune to such problems, is that 'continuity of first-person perspective' must be construed as the continuing existence of an indivisible part of the original person, her soul; and then, I now add against Baker, no one needs to 'think of herself as herself' in order to have some primitive conscious events, e.g. sensations, and so to continue to exist. Peter van Inwagen's theory, to which Dew is also sympathetic, that personal identity over time is constituted by the continuing life of a physical organism, also runs into all the problems of any complex theory of personal identity. To take a slightly different example, it runs into the problem that any answer to the inevitable question of how many bodily parts can be replaced how gradually for the person still to be the same person will seem highly arbitrary. The obvious non-arbitrary solution to this question is that the truth about when a person continues to exist is a truth over and above any truth about how many bodily parts have been replaced how gradually, but that the fewer parts are replaced and the more gradual the replacement, the more probable it is that the same person continues to exist.

Both Dew and Baker are troubled by consequences of my theory. Dew (p. 33) claims that my theory has the 'troubling and unfortunate consequence' that we may be completely unable 'to identify ourselves across time. But I argued that under normal circumstances (i.e. when brains are not split, and memory and character are continuous) it is 'enormously probable' that I am the same person as any previous person who had the same brain. To ask for more than that is unreasonable. It is only under extremely abnormal circumstances that there will be a serious doubt about which previous person was me. And, as Dew admits, that there will be a serious doubt under such circumstances is no argument against the theory. Baker (p. 13) regards it as a disadvantage of my theory that 'there is not (or rather I cannot think of) any naturalistic way that an immaterial mind could have come into existence. I too cannot think of such a way, but I - unlike Baker - am very happy to endorse this consequence of my theory – given what I regard as strong arguments in favour of that theory.

IV. FREE WILL AND MORAL RESPONSIBILITY

David Hunt has confined his comments to my views on what it is to have 'free will' of the kind that makes us morally responsible. He has read my text very carefully – more carefully than I read some of the texts dealing with PAP, 'The Principle of Alternate Possibilities', when – as Hunt points out – I represented (on my pp. 203-4) various authors as making a claim about free will (in my sense), whereas in fact their claim concerns moral responsibility. Fortunately, as Hunt kindly remarks (p. 45), this error (for which I apologize to my readers) makes no difference to my main arguments in chapter 8 that free will in my sense – an agent acting 'intentionally without their intentions being fully determined by prior causes' (my p. 202) – is necessary for moral responsibility.

Generally Hunt is very sympathetic to my views on these topics, and so I am left to deal with some fairly minor issues. I stand by my argument that belief (that is, belief that some proposition is true) is immune to the will. My claim is that we cannot change our beliefs at will, that is immediately by a decision. However I wrote (p. 77) that 'I can try to brainwash myself, so as to come to hold later a certain belief specified in advance; but I will only succeed if I get myself to be caused to hold the belief at the later time by some cause, e.g some brain event, which

I am not at that time intentionally causing. Pascal had a suggestion of a procedure for doing this - 'taking holy water, having masses said, and so on. That will make you believe quite naturally.1 This point should be enough to deal with Hunt's claim that people actually do voluntarily change their beliefs; to do so requires a period of brainwashing and is not guaranteed success. Even if some, maybe most, people do not accept my philosophical thesis that belief that p is belief that p is more probable than not -p, that doesn't show that the thesis is false! It may show only that some people do not fully understand the logical consequences of 'S believes that p'. Hunt asks for independent evidence that when we do an action intentionally, we always believe that it is in some way a good action to do. He claims (p. 43) that Augustine's action of stealing pears when young, 'didn't fit this paradigm'. It seems to me that Augustine's description of his motivation in stealing the pears does exactly fit the paradigm, even though Augustine himself seems to doubt it. For he writes of his 'pleasure' in stealing, and that he 'loved' the act, and 'loved' the evil in him. You cannot get 'pleasure' from something, or 'love' something, without thinking it in some respect a good thing. But however that may be, the passage of mine which Hunt cites concerns moral beliefs, not any value beliefs; 'moral beliefs' on my definition are (roughly) value beliefs about the overall goodness of actions, of a kind overlapping with the views of many other actual humans; and I spell this out more carefully by illustrating what 'overlap' amounts to. Those are the kind of value beliefs which – I argue – are such that we are morally responsible for acting on them or not acting on them. My principal claim is that we could not have a moral belief without having some inclination to act upon it. Contrary to Hunt's claim (p. 48) that I hold that we are culpable 'only for acting contrary to our value beliefs, my view is that we are culpable only for acting contrary to our moral beliefs; everyone has value beliefs, but not everyone has moral beliefs. Psychopaths - such as perhaps was Hitler are not, in my view, culpable for their actions, unless they have moral beliefs. But of course they may be culpable for acting so often contrary to their moral beliefs that they ceased to have any.

Hunt raises two further issues about the kinds of actions for which – given that we have free will (in my sense) – we are morally responsible. The first is this: I claimed that normally we are praiseworthy only for

¹ Blaise Pascal, *Pensées*, translated by A.J. Krailsheimer (London: Penguin Books, 1966), No. 418.

doing actions which we believe to be good but not morally obligatory; but that we are praiseworthy for doing actions believed obligatory only if we do so by overcoming strong contrary temptations (my p. 212-13). Analogously I added later (my p. 220), that we are not praiseworthy for doing an action believed good but not obligatory, if we were not subject to any contrary desire. Hunt rightly claims that I could have given a simpler account of this total view. The complexity arose, I suspect, because of my own doubts about the (p. 220) claim at the stage of the argument when I put forward the (p. 213) claim. However I stand by this view, which results from my reflection on the principles lying behind what I suggest are our intuitive judgements about many particular cases; and I do not find Hunt's judgement about the woman who rushes into the burning building at all persuasive. She instinctively does the right thing, and so deserves admiration for her character from which her action inevitably and unthinkingly flows, but she can't deserve any extra admiration for acting on that character when circumstances require it. So she doesn't deserve the kind of admiration which belongs to someone who fights contrary temptation in order to do the right thing. And although we should be enormously grateful to God for our creation, I do not think that we should regard God as praiseworthy in the same sense as humans are praiseworthy, except in respect of his actions when incarnate and when in my view he was tempted to do and could have done less than the best, even though he could not have failed to fulfil his obligations.² The second issue concerns whether we are more blameworthy for successful wrongdoing than for unsuccessful wrongdoing. I stick by my view that on a libertarian account of free will it is irrational to hold a bomber more blameworthy when his bomb works than when it doesn't. We owe reparation for the harm we cause, and the successful bomber will owe a lot more reparation than the unsuccessful one; but we must not confuse his resulting indebtedness with his being more blameworthy. And I mentioned (my pp. 211-12) additional reasons why the criminal law is right to punish successful bombers more severely than unsuccessful ones.

² On the point that if God became incarnate as Jesus Christ, he could not have done wrong, but he could have done less than the best, and is praiseworthy for not doing less than the best, see my *Was Jesus God?* (Oxford: Oxford University Press, 2008), pp. 44-47.