



## A CARTESIAN ARGUMENT FOR SUBSTANCE DUALISM

### I

In my recent book *Are we bodies or souls?*<sup>1</sup> I argued in favour of Descartes's version of substance dualism – that humans on earth consist of two parts, body and soul, of which the soul is the one essential part. I understand by a 'soul' a non-physical substance which is capable of having conscious experiences. We exist if and only if our souls exist. In that book I gave arguments of two kinds for this view. The first kind of argument proceeded from the impossibility of giving criteria for personal identity over time (that is, criteria for a person P2 at a time T2 being the same person as a person P1 at a time T1) in terms of the continuing of any physical substance or physical or mental property; and so the only possible necessary criterion for such identity is that P2 is the same person as P1 only if P2 has P1's soul. I shall not discuss this kind of argument in this paper. In this paper I shall summarize my treatment of the second kind of argument, an argument from the mere conceivability by each of us of our existing without a body to the conclusion that each of us has a soul as our one essential part. I shall claim that Descartes's own argument of this kind shows that the existence of the soul of each of us is a sufficient condition for our existence, but not that it is a necessary condition for our existence. I shall go on to claim that a slightly amended version of Descartes's argument will show that the existence of the soul of each of us is a necessary condition for our existence. I shall then defend this view against the normal fashionable objection to any version of any argument of this kind.

Here is Descartes's argument, as presented in *Discours de la Méthode*, which will be very familiar to you all:

Examining attentively that which I was, I saw that I could conceive that I had no body, and that there was no world nor place where I might be; but yet that I could not for all that conceive that I was not. On the contrary, I saw from the very fact that I thought of doubting the truth of other things, it very evidently and certainly followed that I was; on the other hand if I had only ceased from thinking, even if all

---

1 Oxford University Press, 2019, revised edition, 2023.

the rest of what I had ever imagined had really existed, I should have no reason for thinking that I had existed. From that I knew that I was a substance the whole essence or nature of which is to think, and that for its existence there is no need of any place, nor does it depend on any material thing; so that this 'me', that is to say, the soul by which I am what I am, is entirely distinct from body, and is even more easy to know than is the latter; and even if body were not, the soul would not cease to be what it is.<sup>2</sup>

The argument seems to me to have three premises:

**first premise:** I am a substance which is thinking.

**second premise:** it is conceivable that 'I am thinking and I have no body'.

**third premise:** it is not conceivable that 'I am thinking and I do not exist'.

from which he argues to the conclusion

**conclusion** I am a soul, a substance, the essence of which is to think.

Descartes assumes that he is a 'substance'. Although Descartes gives somewhat different definitions of 'substance' in different places in his writings,<sup>3</sup> all that we need to assume that he means here by a 'substance' is 'a component of the world' – one of the things that exist at a particular moment of time and has properties. **Descartes's first premise** is the contingent premise that (at the time when he was considering this argument) 'I am thinking'. Descartes uses 'thinking' in a wide sense. He wrote elsewhere:

By the word thought I understand all that of which we are conscious as operating in us. And that is why not only understanding, willing, imagining but also feeling, are here the same thing as thought.<sup>4</sup>

**Descartes's second premise** is that, while he is thinking, he can, compatibly with his logically contingent premise 'I am thinking', 'conceive that [I] have no body'. For maybe Descartes just dreams that he has a body. I shall understand a proposition being 'conceivable' as it being logically possible, in the sense of not entailing a contradiction. The premise that this is conceivable certainly

2 *The philosophical works of Descartes*, translated by E.S.Haldane and G.R.T. Ross, Cambridge University Press, vol.1, 1968, p.101.

3 In some places he defines a 'substance' as a thing in which properties inhere – see, for example, his *Arguments demonstrating the existence of God and the distinction between soul and body* Definition 5, in Haldane and Ross, vol.2, p. 53. In other places he defines a 'substance' as a thing 'which can exist by itself, without the aid of any other substance' (*Reply to objections* IV in Haldane and Ross, vol.2, p.101.) Elsewhere he qualifies this second definition by acknowledging that 'creative substances... are things which need only the occurrence of God in order to exist' (*Principles of Philosophy*, principle 52, in Haldane and Ross, vol.1, p.240)

4 *Principles of Philosophy*, principle 9, in Haldane and Ross, vol.1, p.222.

seems to those who first hear it, immensely plausible. For what is it to have a body? It is to have a physical substance, a chunk of matter through which one can make a difference to the physical world and through which one learns about the physical world. Some reports of 'near death' experiences of patients undergoing an operation report that patients claim to have experiences of floating above the operating table at the same time as the surgeons certify that those patients are 'brain-dead'. And while we may suspect that really the patients did not have those experiences at exactly the same time as they were brain-dead, or that surgeons may sometimes judge a patient to be 'brain-dead' while there is still some activity in the patient's brain, we can certainly understand what the reports claim, and fairly evidently they do not entail any contradiction. And what the reports claim is that at the time of the 'near-death' experiences, the patient could not control any body or learn about the world through any body; and so their conceivable (though possibly false) claims were claims that they were having experiences at that time when they did not have a body. So— I suggest— Descartes was right in claiming that there is no contradiction entailed by 'I am thinking and I have no body', and so it is conceivable (= is logically possible) that I am thinking and I have no body'. So his second premise is true.

**Descartes's third premise** is that he cannot conceive that 'I am thinking and I do not exist'. And it surely is obvious that that proposition entails a contradiction and so is inconceivable. 'I am thinking' obviously entails 'I exist'; and so 'I am thinking and I do not exist' entails 'I exist and I do not exist', which is a contradiction.

So it follows from the three premises that 'I am a substance which, it is conceivable, can exist without a body'; in his own words 'for my existence there is no need of any place nor does it depend on any material thing'. Yet, if he exists, there must be some part of him which exists; and so if he can exist without a physical (= material) part, he would need a non-physical part in order to exist. And so having a soul would be sufficient for his existence. But it does not follow that he needs a soul now when he has a body, in order to exist now; and so it does not show that having a soul is a necessary condition for Descartes's existence.

However, that will follow if we substitute for what I have construed as Descartes's second premise this **amended second premise, that it is conceivable that 'While I am thinking, my body is suddenly destroyed'**, (that is, in the middle of a period while I am now thinking my body is suddenly destroyed). I suggest that this stronger principle is also correct. If it is conceivable that while I am now thinking, I have no body, it is surely conceivable that while I am now thinking, I suddenly cease to have any control over my body or to be influenced by anything that happens in it; and in that case I would have ceased to 'have' a body; the body would be un-owned by me. Many of the reports of 'near-death' experiences of patients were of experiences of leaving their bodies; and while again we may reasonably doubt

whether the patients really left their bodies, we can understand what their claims to have left their bodies mean –that they were at one time embodied and then left their bodies and observed them from a distance. Fairly evidently these claims do not entail contradictions. So if it is conceivable that I ‘observe’ while I cease to have a body, it is conceivable that I remain conscious when my body has been destroyed. So ‘While I am thinking, my body is suddenly destroyed’ is conceivable. Then there follows from the amended second premise and the third premise this lemma that ‘I am a substance which, it is conceivable, can continue to exist while my body is suddenly destroyed’.

I now add what I will call a **fourth premise, that is inconceivable is that any substance can lose all its parts simultaneously and yet continue to exist**. A table may continue to exist if it loses a leg, but not if it loses all its legs and the table top at the same time. And organisms can continue to exist if over time they lose all their parts, so long as those parts are gradually replaced, each part being replaced over a period of time while other parts continue to exist. A tree can continue to exist if each cell is replaced by a similar cell at different times. But what is inconceivable is that every single part of the tree should be suddenly destroyed and yet that tree should continue to exist. It therefore follows from that, that if it is conceivable that now while Descartes has a body, that body is suddenly destroyed and yet he continues to exist, then he must actually have now also another part which is not destroyed and which is doing the thinking, and which he and I are calling his ‘soul’. For if he didn’t already have that other part when his body is destroyed, he could not continue to exist. And since at every time while he is thinking and has a body, it is conceivable that he should lose his body and yet continue to exist, it follows that at every time while he is thinking, he has a soul. For if at some time he did not have a soul, it would be logically impossible that at that time he should continue to exist when his body is suddenly destroyed. Hence, Descartes, knowing the truth of his first and contingent premise that he is thinking, is entitled to conclude that having a soul is not merely sufficient, but necessary for his existence. And since each of us humans who are conscious can formulate the same argument with respect to ourselves, we are entitled to conclude that each of us has a soul which is necessary and sufficient for our existence. I believe that this amended argument is valid, and – for each of us for whom its premise is true – knowably sound.

However, that puts me in a very small minority of contemporary philosophers. Contemporary philosophers might be prepared to admit that Descartes’s argument is a valid argument (that is, its conclusion is indeed entailed by its premises), and they might well be prepared to admit that my own amended version is also a valid argument. But they claim that Descartes has no justification for asserting his second premise, and likewise they would claim that I have no justification for asserting the amended version of that premise. This is because, in the objectors’ view, we simply have no idea of

what we are referring to by 'I'. Maybe 'I' refers to my body, and in that case of course the second premise is false. Or maybe it refers to some hidden essence of me, about the nature of which neither I nor they have the slightest idea; and in that case, although the second premise might be true, we are not in a position to know that it is true. Shoemaker has claimed more generally that that one could never reach a conclusion about the actual world from mere considerations of what is conceivable: '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.'<sup>5</sup> It was in order to deal with this objection that I invented some new terminology.

## II

I understand by a 'designator' a word (or longer expression) which refers to some object – substance, property, event, or time. I define an 'informative designator' as a designator which is such that if we know what the designator means, we know to which object it is referring; and an 'uninformative designator' as a designator which is such that knowing what the designator means does not ensure that we know to which object it is referring.

So, first, what is it to know what some designator means? Often we know this if we know some formal definition of it, of a kind which one finds in a dictionary. A 'taxidermist' is 'a person who stuffs the skins of dead animals, in order to make them look like living animals.' A 'tort' is 'an infringement of a legal right which leads to a legal liability'. And most scientific terms have precise dictionary definitions. An 'electron' is a subatomic particle with a negative electric charge of  $1.602 \times 10^{-19}$  coulomb and a mass of  $9.109 \times 10^{-31}$  kg. But definitions are of no use unless we know what the words in the definition mean. Those words may be defined by other words – a 'subatomic particle' is 'one of the many kinds of particles of which atoms are composed'. And so on. But in the end if we are to understand any definition its constituent designators have to be understood in some other way than by definition and that is by having an ability to recognise straight off whether they apply to an object or not. This ability is normally acquired by 'acquaintance', that is by seeing (hearing, or perceiving in some other way) objects to which those words apply and having these objects contrasted with other objects to which those words do not apply, and being told that (or coming to understand as a result of listening to conversations that) the word applies to that particular object or – alternatively – to objects like it in a certain respect. We learn what an 'animal' is, by seeing many animals, and being told that 'animals'

5 'Sydney Shoemaker's Reply', in S. Shoemaker & R. Swinburne, *Personal Identity*, Basil Blackwell, 1984, p.144.

are things like this, to be contrasted with plants which are not is like this; we learn what 'mass' means by being told that it is roughly the same as 'weight' and by being shown objects which have different weights and among them objects of 1 kg weight. We learn what '10-3<sup>1</sup>' means by being shown what it means to reduce a quantity to one 10<sup>th</sup> of its value, and what it means to perform an operation 31 times. It is in this way that we learn words (or longer expressions) designating simple observable properties or kinds of substance like 'line', 'angle', 'heavier', 'longer', 'door', 'road', 'straight', 'shirt', 'walks'; and words important in human interaction such as 'face', 'mouth', 'talks', and 'kiss'. All of us learn the meanings of a vast number of words (or longer expressions) by 'acquaintance'. For some of these words, some people learn their meaning by acquaintance, while other people learn their meaning by definition; but that can only happen if the latter learn the meaning of the words in the definition (or words by which those words are defined) by acquaintance, these latter words being words which others might have learnt by definition. In the case of words which are names of particular substances, some people learn what they mean by 'acquaintance'. 'London' is the big city in which we live; 'Stonehenge', is that arrangement of big stones in front of us; 'Woodstock road' is the road on which we live; 'the Mona Lisa' is this painting we are looking at. Other people learn that these words mean what those acquainted with the substance mean by it. This picture of the meaning of proper names is of course similar to Kripke's picture of words getting their meaning by an initial baptism, and other later speakers intending to mean by the word what the earlier speakers meant by them.<sup>6</sup> Clearly language is more complicated than I have been describing it, the meanings of some words depend both on acquaintance and definition, and many words are used in slightly different ways by different speakers. But with amplifications and qualifications it must be basically right. Unless the meanings of many words were determined by the objects to which most speakers would paradigmatically apply them, there could be no language.

Being able to recognise instances of the correct application of a designator straight-off, as I am understanding this, involves being able to recognize whether or not it applies to an object—under ideal conditions. Conditions are ideal when one's faculties are working properly, one is in the best possible position (that is, best possible location relative to the object) for recognizing the property (or whatever) referred to, and one is not subject to an illusion. Thus if someone had normal sight and then became totally blind, their inability now to recognize a face doesn't show that they do not know what the word 'face' means. For now their faculties are not working properly. If someone is too far away from two rods, they may not be able to recognize whether one rod is 'longer than' another rod, but that doesn't show that they do not know what the expression 'longer than' means. For then they are not

6 S. Kripke, *Naming and Necessity*, Blackwell Publishing, 1980, pp. 96-7.

in the best possible position for recognising which rod is the longer. Further, the circumstances must not be illusory, that is such as to make a property (or whatever) look (feel, sound, or whatever) differently from the way it would look in paradigm circumstances (that is, the normal circumstances in which the meaning of its designator is explained to new speakers); or such as to make some other property look differently from the way it looks in paradigm circumstances, so that it looks like the property in question. It does not show that we do not know what the word 'cat' means if we cannot recognise a cat when it is disguised to look like a dog, or we misidentify a robot as a 'cat' if it is made to look and behave like a cat. Nor does it show that we do not know what 'Stonehenge' means if we cannot recognise it when all the standing stones are covered with domes, or we misidentify a perfect copy of Stonehenge made of cardboard as 'Stonehenge'. In these cases observers would be subject to an illusion.

In the case of words whose meaning we know straight-off and so are able to recognize under ideal conditions whether or not they apply, we know—simply in virtue of knowing the meaning of the word—what it is for the object to which they apply to be that object; we know logically necessary and sufficient conditions for something to be that object. For an object to be 'a door' just is for it to look, feel like, and behave like (e.g. open when pushed) paradigm instances of doors. For a person to be 'walking' just is for that person to be doing what we recognize as paradigm instances of persons 'walking' as doing when we observe them under ideal conditions (standing fairly close to that person in daylight, with eyes working properly, and not subject to some illusion). To be London just is to be the big city which we (that is, those of us who learn what 'London' means straight-off) recognize as 'London' under ideal conditions (walking around a big city, with eyes working properly, and not subject to an illusion by being in another city which looks exactly like London.) Hence these words whose meaning we know straight-off are all informative designators. Many words which denote properties and so kinds of substance (such as 'proton') (as opposed to words which denote individual substances which have those properties), and which can be defined by other words whose meaning we know straight-off, are also often used as informative designators; and so we can know to which object (that is, to which property) they refer, in the sense of knowing the logically necessary and sufficient conditions for being that object, merely in virtue of knowing their meaning. Which property a designator denotes is a matter of whether the property satisfies the definition which gives logically necessary and sufficient conditions for the application of the designator. To be a 'taxidermist' just is to be a person who stuffs the skins of dead animals, so as to make them look like living animals.

By contrast, an 'uninformative designator' is a word (or longer expression) which is such that if we know what the word means (that is, the meaning

which is common to its use in different contexts), that is not by itself enough to know to what it refers on a particular occasion of its use. Many 'definite descriptions'—that is, descriptions of an object which pick out that object by some property of that object, such as 'the tallest building in London'—are uninformative designators of that object. We may know the meaning of 'the tallest building in London' and so what property it designates, but knowing this is not enough to show us which building is the tallest building. To know this, we need to compare the heights of different buildings in London and discover the location, size, shape, and composition of the tallest one; and to discover this we need to do much empirical investigation. Likewise, most 'indexicals' are uninformative designators. An indexical is a word like 'he', or 'you', 'that river', or 'now', the referent of which (that is, to what they are referring) depends on the context in which it is uttered; that is, who says it, when, and where. Someone knows what 'that river' means iff they know that it refers to a river to which the speaker has just pointed or alluded, but unless they know the location of the river to which the speaker has pointed and where it is flowing from and to they will not know what that river is, in the sense of what makes the river that particular river. As I shall use the word 'essence', to know an 'essence' of the object is to know a set of logically necessary and sufficient conditions for an object to be that object.

There are other uninformative designators, the referent of which depends on some underlying essence which may be totally unknown. Obvious examples include those used by Kripke, Putnam and innumerable subsequent philosophers to illustrate the use of 'rigid designators' to designate kinds of substance, the logically necessary and sufficient conditions for being of that kind being having some possibly unknown chemical constitution. Thus, the word 'water' was used in the early nineteenth century as a designator of the actual transparent drinkable liquid prevalent in our rivers and seas, and of whatever has the same chemical essence as that liquid. But in ignorance of what that chemical essence was, people in the early nineteenth century did not know what it was to be water. They were able to use the word 'water' as a normally sufficient rigid designator because they knew a collection of properties which were normally sufficient for something to be water – being a transparent drinkable liquid prevalent in our rivers and seas. But they allowed that sometimes some transparent drinkable liquid to be found in some river or sea might not be water; and in ignorance of the necessary conditions for something to be water, they could not be sure whether something (liquid or solid) which was not transparent or drinkable or in our rivers or seas was or was not water. That collection of observable properties by which in practice they picked out something as 'water' is what philosophers have come to call the 'stereotype' of water. But, not knowing a set of logically necessary and sufficient conditions for something to be water (and so not knowing in a crucial sense, the 'essence' of water), they did not fully understand what it is to be water. Subsequent



scientists discovered that to be water is to consist of molecules, each of which consists of two atoms of hydrogen and one atom of oxygen.

I have been assuming that all competent speakers of a language understand a word in the same way. But many of the technical terms of a science which function as informative designators for experts of some science function as uninformative designators for the wider public. The wider public understands words such as 'gold' in what has been called a 'deferential' sense, that is they mean the word to refer to an object which they pick out by a stereotype and which also has whatever properties the experts have discovered to be essential to that object. So they understand by 'gold' a kind of chemical substance which is normally heavy, yellow, and malleable, and has whatever essence scientists have discovered the kind of substance which normally has these properties to have. I shall call an uninformative designator an '[uninformative] deferential designator' of an object iff one of the properties which determine its meaning and so help to determine whether or not it applies to an object is that an expert of some kind judges that it applies. It is involved in the meaning of an uninformative deferential designator that speakers 'defer' to some expert to tell them to what it refers. So while the wider public means something different by 'gold' from what scientists mean, both groups refer to the same chemical substance by the designator "gold" and know that they do. Likewise proper names of individual substances such as 'London' are uninformative deferential designators for those who have never visited London; they mean by 'London' that city which its inhabitants call 'London'; the latter are the experts to whom others defer for the meaning of 'London'.

Since most of the words by which we refer to properties and many of the words by which we refer to substances are informative designators, it is the case of most sentences that if we know what their constituent words mean we know what are the objects to which they refer, and so if such a sentence is logically possible, it is logically possible that the objects referred to could have the properties referred to. This is because if we can refer to an object by an informative designator, we know a set of logically necessary and sufficient conditions for something being that object, and so are in a position to work out what is logically possible and what is logically impossible for that object. Thus, not merely does the sentence 'Stonehenge is 100 million years old' not entail a contradiction and so is logically possible, but it is logically possible that the object referred to by 'Stonehenge' is 13 million years old; and this is because '100 million years old' and 'Stonehenge' (for those who know its meaning by acquaintance) are informative designators, and so in knowing what these expressions mean we know to which substance 'Stonehenge' is referring and to which property '100 million years old' is referring. So we can work out whether among the properties which it is logically possible for that substance to have is the property designated by '100 million years old'. But if a sentence contains an uninformative designator referring to, for example,

some substance such as 'water' (as used in the early nineteenth century), then the sentence may entail no contradiction but what it asserts about water may not be logically possible for the actual substance picked out by 'water'.

The distinctions which I have made in this section will allow us to analyse more precisely the nature of the objection to Descartes's argument and to my amended version of it, that we do not know to what the 'I' in what I have analysed as his second premise, refers. So in terms of my terminology the objection is that 'I' is an uninformative designator; and if that were so, it would be like most other indexicals in this respect. What I shall claim, in response to this objection, is that 'I' is an informative designator for each speaker who uses that expression, but an uninformative deferential designator for anyone who hears someone else use that expression; and that likewise one's own proper name, such as 'Richard Swinburne' is an informative designator for Richard Swinburne, but an uninformative deferential designator for everyone else. The person to whose authority other speakers must defer in order to know to what 'I' refers is the speaker is using the word: and the person to whose authority they must defer in order to know to whom the name of the person is referring is that person.

### III

I shall approach the issue of whether 'I' is an informative designator of a substance by the connected issue of whether the designators of conscious experiences are informative or uninformative. By a person's 'conscious experiences' I mean those events about which a person knows better than anyone else can whether or not they are occurring in them. These events include sensations, occurrent thoughts, and intentions. I shall illustrate this point only with respect to sensations. Each of us has privileged access to our sensations in the sense that whatever way anyone else has of learning whether or not we now have a headache or are hearing a noise, or feeling sick, we could always use the same way as other people use to learn about whether or not we are having these experiences, yet we have a unique way which is not available to others – we know which kind of sensation we are having because we are experiencing it. The purpose of this approach is to argue that the words referring to each person's conscious experiences are informative for that person, but not for others; and so it is to be expected that the word referring to that person themselves, 'I', is informative for that person, but not for others.

We may mean by a word which we use to describe a current sensation, either simply a sensation of the kind which we are currently experiencing, or a sensation of a kind which we have experienced in the past and can recognise its reappearance. Because we have this unique way of access to our sensations, the person having a sensation is the only person who is ever in

the best possible position to know the content of that sensation, that is what kind of sensation it is in the sense of what the sensation feels like. The only faculty involved in our awareness of our sensations is our faculty of awareness, that is being conscious; and so inevitably our faculties are working properly when we have sensations. Hence – unless we are subject to an illusion (in falsely supposing the sensation to be similar to certain past sensations) we can always recognise whether some word whose meaning we know applies to that sensation or not. Hence the condition for a word whose meaning we know being an informative designator is always satisfied with respect to all the words which we use to describe our own sensations.

We normally use words to describe different kinds of sensations by the physical events which normally cause sensations of that kind and/or by the physical events (or the desires to do certain actions) which they normally cause. We learn to describe some food as ‘tasting of coffee’ iff it tastes like the taste which coffee causes. We learn to describe a smell as a ‘smell of burnt almonds’ iff it smells like the smell which burnt almonds cause. We learn to describe an after-image as a ‘red’ after-image iff it has the same visual appearance as do the visual appearances of (and so caused by) certain paradigm public ‘red’ objects (British post boxes, ripe tomatoes, strawberries etc.). We learn to describe a sensation as an itch iff it is the kind of sensation which causes us to scratch (or to desire to scratch) the place which seems to cause the sensation. For some kinds of sensations both their normal causes and their normal effects are important for enabling us to pick out the sensation to which we are referring. We learn to describe a sensation as an ‘acute pain’ iff it is the sort of sensation which is caused in us by certain kinds of bodily events (such as being cut or burnt), and which causes aversive behaviour (causes us to try to stop the pain if we know how to do so, for example by withdrawing a hand from the cause of the cut or running away from the fire which is burning us, or strongly to desire to do so). But having learnt to refer to a particular kind of sensation by events which normally cause it or are caused by it, we are then in a position to refer to a sensation of that particular kind when it does not have these normal causes or effects. What each of us means by a ‘red sensation’, ‘taste of coffee’, ‘smell of burnt almonds’ is what we experience, not the normal cause of that experience.

The fact that most other people learn to use words denoting sensations in the same way as we learn them, and seem to make the same distinctions between different sensations as we do (for example distinguishing red sensations from green sensations, and the taste of coffee from the taste of chocolate) makes it probable that they mean by the words they use to describe their sensations what we mean by the same words. But it can sometimes be reasonable for each of us to doubt whether other people do mean by the words they use to describe their sensations the same as we do. Some people cannot distinguish between red and green objects; red and green objects both

look the same to them. So either red objects do not look to them the way they look to most of us and/or green objects do not look to them the way they look to most of us; and so either when they say ‘it looks red’ or ‘it looks green;’, they must mean that it has a sensory appearance different from the sensory appearance which we describe by these sentences. (Note that whether some public surface is ‘red’ is a public matter, a matter of whether it looks to most observers to be of the same colour as (to produce the same colour sensations as) certain paradigm objects. But whether a surface ‘looks red’ to a certain person – as I am using that expression – is a matter of the particular sensory quality which it has in that person’s experience of it.) And in the case of many sensations and especially tastes, the different reactions which people often have to the same input from their sense organs supports the hypothesis that the sensations caused thereby are sometimes different in different people. Insofar as we have reason to suppose that experiencing certain physical events causes different sensations in others from what it does in ourselves, we have reason to suppose that they mean something different by the words they use to describe their sensations from what we mean by those words.

It follows that expressions used by us to describe our own sensations function as informative designators for us of our own sensations, but when used by others to describe their sensations serve only as uninformative designators for us. Suppose that both John and Mary learn to use ‘acute pain’ as a designator of the kind of sensation which is normally caused in them by certain kinds of bodily events (such as being cut or burnt); and which normally causes aversive behaviour (for example, causes them to withdraw their hand from the cause of the cut or run away from the fire which is burning them, or to desire to do so). Then for John ‘acute pain’ just means the sort of sensation which is caused *in him* by those bodily events and which causes *him* to show aversive behaviour. Whether or not on a particular occasion it has such causes and effects, he uses ‘acute pain’ to refer to the intrinsic character of any sensation which feels like that. But he will understand Mary’s use of ‘acute pain’ to mean the sort of sensation to which Mary alone has privileged access, which is normally caused *in her* by those bodily events and which normally causes *her* to show aversive behaviour, whether or not on a particular occasion it has such causes and effects – although he will also believe, in the absence of contrary evidence, that probably what she means is the same as what he means.

John’s understands Mary’s use of ‘acute pain’ as a ‘deferential’ uninformative designator because John understands Mary’s use of that expression to mean whatever Mary means by whatever sensation of hers is normally picked out by the stereotype of certain bodily manifestations. The situation of John in understanding Mary’s use of ‘acute pain’ to mean whatever she means by a sensation related to certain observable manifestations, is thus like that of the unscientific public who understand the scientists’ use of the word ‘gold’ to mean what the scientists mean by the essence of an object which normally

has certain observable properties. But there is this crucial difference from the scientific case, that while the unscientific public can – if they so choose – learn to use ‘gold’ as an informative designator, John could never learn to understand Mary’s use of ‘acute pain’ as an informative designator. This is because Mary alone can ever be in the best possible position to know whether that expression as used by her applies to some sensation of hers. And of course conversely – Mary can never learn to understand John’s use of ‘acute pain’ as an informative designator.

#### IV

Conscious events are the events they are, not merely in virtue of the properties, such as ‘having a pain’ or ‘smelling a smell of burnt almonds’ involved in them, but in virtue of the persons who have these properties. So finally, are the words by which we refer to persons informative or uninformative designators? I suggest that the understanding of the criteria for some word being an informative designator developed in this paper enables us now at last to answer this crucial question. As when we analyse the meanings of words used to refer to conscious events, we need to distinguish the meanings of words used to refer to our own conscious events from the meanings of words used by others to refer to their conscious events, so too in analysing the meanings of words used to refer to persons, we need to distinguish the meanings of words used to refer to ourselves from the meanings of words used to refer to others. We refer to ourself by the word ‘I’. I suggest that, as used by each person, ‘I’ is an informative designator of themself. For we are always able, when (1) our faculties are working properly, (2) we are in the best possible position for recognizing ourself, and (3) we are not subject to an illusion, to recognize when some person is ‘I’ and when it is not. When each of us refers to themself as the subject of a current conscious event, the faculty which they need for this purpose is clearly in working order. Each of us is in the best possible position for recognising themself when they pick out themself as the subject of a current conscious event, as the person who is now having this pain or that thought. Under those circumstances none of us can possibly be subject to illusion. For an illusion would consist in the circumstances being such that I refer to someone as ‘I’ who is not myself, or I fail to recognize myself as ‘I’. But I couldn’t possibly have a pain and suppose that really it was someone else who was having the pain; nor could someone else be having a pain and I suppose mistakenly that the pain is really mine. Of course, I could suppose that some pain which I am having was of just the same kind as a pain which someone else was having, but what I cannot be mistaken about is that I am having the former pain – because if I thought that I was not having a pain, I wouldn’t be feeling anything. When each of us is referring to themself as the

subject of a current conscious event, we are in Shoemaker's phrase, 'immune to error through misidentification.'<sup>7</sup>

Since (unlike most other indexicals) 'I' as used by me is an informative designator – for me, so too is 'Richard Swinburne' – for me; and so is 'I' and their own proper name – for each other person. So the situation is similar to the situation with respect to description of our sensations. Just as each of us is always in better position than anyone else could be with respect to describing their sensations, so too each of us is always in a better position than anyone else ever could be for recognizing ourselves – and that is when we refer to ourselves as the subject of sensations (or other current conscious experiences). Others can only pick us out as that human who has that particular body and/or brain and/or makes certain memory claims (and perhaps has a certain character) and refers to himself or herself as 'I' or by their own proper name. But since those others do not have the (not merely privileged but infallible) access to who the person is to whom they are referring which I have, their use of 'Richard Swinburne' or some indexical expression to refer to me involves using it as an uninformative designator. Others mean by 'Richard Swinburne' the person whose body is such and such a body, or whose brain is such-and-such a brain, and/or who makes certain memory claims, and refers to himself as 'Richard Swinburne' or 'I'. Their use of the uninformative designator 'Richard Swinburne' is therefore also deferential; they regard me, picked out by physical properties, one of which is that I (picked out by the other physical properties) refer to myself as 'I', as the expert on who I am. But I mean by 'I' the person who is aware of himself as experiencing a certain particular conscious event, and not any person who is not experiencing that conscious event.

Our infallible knowledge of ourself is an infallible knowledge of ourself as existing at the moment at which we are aware of this. We do not have infallible awareness of what we experienced at some past time or even whether we existed at some past time, nor any infallible awareness of what we will experience in future. Nevertheless, when I believe that I experienced such-and-such at a certain past time, or will experience such-and-such at a certain future time, I know infallibly who it is to whom I believe that such and such experiences occurred or will occur. Others who believe that I experienced such-and-such at a certain past time or will experience such-and-such at a certain future time cannot know as well as I do what it would be like for their beliefs to be true because they do not have the infallible access which I have to the identity of the person about whom they have these beliefs.

It follows that Descartes did know to what he was referring by 'I' when he claimed that it was 'conceivable' (= logically possible) that 'while I am thinking, I have no body'. So, knowing a set of logically necessary and

<sup>7</sup> Sydney Shoemaker, 'Introspection and The Self' in (ed.) Q. Cassam, *Self-Knowledge*, Oxford University Press, 1994, p.82.

sufficient conditions for a person to be 'I' (consisting in being the person who is having certain particular experiences of which he is aware), he knew what it would be like for that proposition to be true, given to what 'I' refers; he was in a position to judge whether or not that proposition is conceivable. And the same goes for the similar proposition used in my amended version of Descartes's argument, 'While I am thinking, my body is suddenly destroyed', which I claimed to be conceivable (= logically possible), given what 'I' refers to – a claim which I hope that I made plausible by spelling out one way in which it could be true. So this (now) traditional objection to Descartes's argument that no one knows to what they are referring by 'I', which – if cogent – would apply also to my revised version of it, fails. And since each of us can use Descartes's argument to show the same thing about him or herself, all human beings are substances having one and only one essential part, their soul. Each of us also has a body, but our body is a non-essential part of us. We are who we are independently of the body to which we are linked; and – if it were naturally possible for the stream of our consciousness to continue when our body ceases to function – it would be our soul and so we who continue to exist.