

Rene Descartes**MEDITATIONS ON FIRST PHILOSOPHY**

in which are demonstrated the existence
of God and the distinction between
the human soul and the body

FIRST MEDITATION

What can be called into doubt

[1] Some years ago I was struck by the large number of falsehoods that I had accepted as true in my childhood, and by the highly doubtful nature of the whole edifice that I had subsequently based on them. I realized that it was necessary, once in the course of my life, to demolish everything completely and start again right from the foundations if I wanted to establish anything at all in the sciences that was stable and likely to last. But the task looked an enormous one, and I began to wait until I should reach a mature enough age to ensure that no subsequent time of life would be more suitable for tackling such inquiries. This led me to put the project off for so long that I would now be to blame if by pondering over it any further I wasted the time still left for carrying it out. So today I have expressly rid my mind of all worries and arranged for myself a clear stretch of free time. I am here quite alone, and at last I will devote myself sincerely and without reservation to the general demolition of my opinions.

[2] But to accomplish this, it will not be necessary for me to show that all my opinions are false, which is something I could perhaps never manage. Reason now leads me to think that I should hold back my assent from opinions which are not completely certain and indubitable just as carefully as I do from those which are patently false. So, for the purpose of rejecting all my opinions, it will be enough if I find in each of them at least some reason for doubt. And to do this I will not need to run through them all individually, which would be an endless task. Once the foundations of a building are undermined, anything built on them collapses of its own accord; so I will go straight for the basic principles on which all my former beliefs rested.

[3] Whatever I have up till now accepted as most true I have acquired either from the senses or through the senses. But from time to time I have found that the senses deceive, and it is prudent never to trust completely those who have deceived us even once.

[4] Yet although the senses occasionally deceive us with respect to objects which are very small or in the distance, there are many other beliefs about which doubt is quite impossible, even though they are derived from the senses -- for example, that I am here, sitting by the fire, wearing a winter dressing-gown, holding this piece of paper in my hands, and so on. Again, how could it be denied that these hands or this whole body are mine? Unless perhaps I were to liken myself to madmen, whose brains are so damaged by the persistent vapours of melancholia that they firmly maintain they are kings when they are paupers, or say they are dressed in purple when they are naked, or that their heads are made of earthenware, or that they are pumpkins, or made of glass. But such people are insane, and I would be thought equally mad if I took anything from them as a model for myself.

[5] A brilliant piece of reasoning! As if I were not a man who sleeps at night, and regularly has all the same experiences -- while asleep as madmen do when awake -- indeed sometimes even more improbable ones. How often, asleep at night, am I convinced of just

such familiar events -- that I am here in my dressing-gown, sitting by the fire -- when in fact I am lying undressed in bed! Yet at the moment my eyes are certainly wide awake when I look at this piece of paper; I shake my head and it is not asleep; as I stretch out and feel my hand I do so deliberately, and I know what I am doing. All this would not happen with such distinctness to someone asleep. Indeed! As if I did not remember other occasions when I have been tricked by exactly similar thoughts while asleep! As I think about this more carefully, I see plainly that there are never any sure signs by means of which being awake can be distinguished from being asleep. The result is that I begin to feel dazed, and this very feeling only reinforces the notion that I may be asleep.

[6] Suppose then that I am dreaming, and that these particulars -- that my eyes are open, that I am moving my head and stretching out my hands -- are not true. Perhaps, indeed, I do not even have such hands or such a body at all. Nonetheless, it must surely be admitted that the visions which come in sleep are like paintings, which must have been fashioned in the likeness of things that are real, and hence that at least these general kinds of things -- eyes, head, hands and the body as a whole -- are things which are not imaginary but are real and exist. For even when painters try to create sirens and satyrs with the most extraordinary bodies, they cannot give them natures which are new in all respects; they simply jumble up the limbs of different animals. Or if perhaps they manage to think up something so new that nothing remotely similar has ever been seen before -- something which is therefore completely fictitious and unreal -- at least the colours used in the composition must be real. By similar reasoning, although these general kinds of things -- eyes, head, hands and so on -- could be imaginary, it must at least be admitted that certain other even simpler and more universal things are real. These are as it were the real colours from which we form all the images of things, whether true or false, that occur in our thought.

[7] This class appears to include corporeal nature in general, and its extension; the shape of extended things; the quantity, or size and number of these things; the place in which they may exist, the time through which they may endure, and so on.

[8] So a reasonable conclusion from this might be that physics, astronomy, medicine, and all other disciplines which depend on the study of composite things, are doubtful; while arithmetic, geometry and other subjects of this kind, which deal only with the simplest and most general things, regardless of whether they really exist in nature or not, contain something certain and indubitable. For whether I am awake or asleep, two and three added together are five, and a square has no more than four sides. It seems impossible that such transparent truths should incur any suspicion of being false.

[9] And yet firmly rooted in my mind is the long-standing opinion that there is an omnipotent God who made me the kind of creature that I am. How do I know that he has not brought it about that there is no earth, no sky, no extended thing, no shape, no size, no place, while at the same time ensuring that all these things appear to me to exist just as they do now? What is more, since I sometimes believe that others go astray in cases where they think they have the most perfect knowledge, may I not similarly go wrong every time I add two and three or count the sides of a square, or in some even simpler matter, if that is imaginable? But perhaps God would not have allowed me to be deceived in this way, since he is said to be supremely good. But if it were inconsistent with his goodness to have created me such that I am deceived all the time, it would seem equally foreign to his goodness to allow me to be deceived even occasionally; yet this last assertion cannot be made. [". . . yet I cannot doubt that he does allow this" (French version).]

[10] Perhaps there may be some who would prefer to deny the existence of so powerful a God rather than believe that everything else is uncertain. Let us not argue with them, but grant them that everything said about God is a fiction. According to their supposition, then, I have arrived at my present state by fate or chance or a continuous chain of events, or by some other means; yet since deception and error seem to be imperfections, the less powerful they make my original cause, the more likely it is that I am so imperfect as to be deceived all the time. I have no answer to these arguments, but am finally compelled to admit that there is not one of my former beliefs about which a doubt may not properly be raised; and this is not a flippant or ill-considered conclusion, but is based on powerful and well thought-out reasons. So in future I must withhold my assent from these former beliefs just as carefully as I would from obvious falsehoods, if I want to discover any certainty.

[11] But it is not enough merely to have noticed this; I must make an effort to remember it. My habitual opinions keep coming back, and, despite my wishes, they capture my belief, which is as it were bound over to them as a result of long occupation and the law of custom. I shall never get out of the habit of confidently assenting to these opinions, so long as I suppose them to be what in fact they are, namely highly probable opinions -- opinions which, despite the fact that they are in a sense doubtful, as has just been shown, it is still much more reasonable to believe than to deny. In view of this, I think it will be a good plan to turn my will in completely the opposite direction and deceive myself, by pretending for a time that these former opinions are utterly false and imaginary. I shall do this until the weight of preconceived opinion is counter-balanced and the distorting influence of habit no longer prevents my judgement from perceiving things correctly. In the meantime, I know that no danger or error will result from my plan, and that I cannot possibly go too far in my distrustful attitude. This is because the task now in hand does not involve action but merely the acquisition of knowledge.

[12] I will suppose therefore that not God, who is supremely good and the source of truth, but rather some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgement. I shall consider myself as not having hands or eyes, or flesh, or blood or senses, but as falsely believing that I have all these things. I shall stubbornly and firmly persist in this meditation; and, even if it is not in my power to know any truth, I shall at least do what is in my power that is, resolutely guard against assenting to any falsehoods, so that the deceiver, however powerful and cunning he may be, will be unable to impose on me in the slightest degree. But this is an arduous undertaking, and a kind of laziness brings me back to normal life. I am like a prisoner who is enjoying an imaginary freedom while asleep; as he begins to suspect that he is asleep, he dreads being woken up, and goes along with the pleasant illusion as long as he can. In the same way, I happily slide back into my old opinions and dread being shaken out of them, for fear that my peaceful sleep may be followed by hard labour when I wake, and that I shall have to toil not in the light, but amid the inextricable darkness of the problems I have now raised.

SECOND MEDITATION

The nature of the human mind,
and how it is better known than the body

[1] So serious are the doubts into which I have been thrown as a result of yesterday's meditation that I can neither put them out of my mind nor see any way of resolving them. It feels as if I have fallen unexpectedly into a deep whirlpool which tumbles me around so that I can neither stand on the bottom nor swim up to the top. Nevertheless I will make an effort and once more attempt the same path which I started on yesterday. Anything which admits of the slightest doubt I will set aside just as if I had found it to be wholly false; and I will proceed in this way until I recognize something certain, or, if nothing else, until I at least recognize for certain that there is no certainty. Archimedes used to demand just one firm and immovable point in order to shift the entire earth; so I too can hope for great things if I manage to find just one thing, however slight, that is certain and unshakeable.

[2] I will suppose then, that everything I see is spurious. I will believe that my memory tells me lies, and that none of the things that it reports ever happened. I have no senses. Body, shape, extension, movement and place are chimeras. So what remains true? Perhaps just the one fact that nothing is certain.

[3] Yet apart from everything I have just listed, how do I know that there is not something else which does not allow even the slightest occasion for doubt? Is there not a God, or whatever I may call him, who puts into me the thoughts I am now having? But why do I think this, since I myself may perhaps be the author of these thoughts? In that case am not I, at least, something? But I have just said that I have no senses and no body. This is the sticking point: what follows from this? Am I not so bound up with a body and with senses that I cannot exist without them? But I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it now follow that I too do not exist? No: if I convinced myself of something then I certainly existed. But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me; and let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must finally conclude that this proposition, I am, I exist, is necessarily true whenever it is put forward by me or conceived in my mind.

[4] But I do not yet have a sufficient understanding of what this 'I' is, that now necessarily exists. So I must be on my guard against carelessly taking something else to be this 'I', and so making a mistake in the very item of knowledge that I maintain is the most certain and evident of all. I will therefore go back and meditate on what I originally believed myself to be, before I embarked on this present train of thought. I will then subtract anything capable of being weakened, even minimally, by the arguments now introduced, so that what is left at the end may be exactly and only what is certain and unshakeable.

[5] What then did I formerly think I was? A man. But what is a man? Shall I say "a rational animal"? No; for then I should have to inquire what an animal is, what rationality is, and in this way one question would lead me down the slope to other harder ones, and I do not now have the time to waste on subtleties of this kind. Instead I propose to concentrate on what came into my thoughts spontaneously and quite naturally whenever I used to consider what I was. Well, the first thought to come to mind was that I had a face, hands, arms and the whole mechanical structure of limbs which can be seen in a corpse, and which I called the body. The next thought was that I was nourished, that I moved about, and that I engaged

in sense-perception and thinking; and these actions I attributed to the soul. But as to the nature of this soul, either I did not think about this or else I imagined it to be something tenuous, like a wind or fire or ether, which permeated my more solid parts. As to the body, however, I had no doubts about it, but thought I knew its nature distinctly. If I had tried to describe the mental conception I had of it, I would have expressed it as follows: by a body I understand whatever has a determinable shape and a definable location and can occupy a space in such a way as to exclude any other body; it can be perceived by touch, sight, hearing, taste or smell, and can be moved in various ways, not by itself but by whatever else comes into contact with it. For, according to my judgement, the power of self-movement, like the power of sensation or of thought, was quite foreign to the nature of a body; indeed, it was a source of wonder to me that certain bodies were found to contain faculties of this kind.

[6] But what shall I now say that I am, when I am supposing that there is some supremely powerful and, if it is permissible to say so, malicious deceiver, who is deliberately trying to trick me in every way he can? Can I now assert that I possess even the most insignificant of all the attributes which I have just said belong to the nature of a body? I scrutinize them, think about them, go over them again, but nothing suggests itself; it is tiresome and pointless to go through the list once more. But what about the attributes I assigned to the soul? Nutrition or movement? Since now I do not have a body, these are mere fabrications. Sense-perception? This surely does not occur without a body, and besides, when asleep I have appeared to perceive through the senses many things which I afterwards realized I did not perceive through the senses at all. Thinking? At last I have discovered it -- thought; this alone is inseparable from me. I am, I exist -- that is certain. But for how long? For as long as I am thinking. For it could be that were I totally to cease from thinking, I should totally cease to exist. At present I am not admitting anything except what is necessarily true. I am, then, in the strict sense only a thing that thinks; that is, I am a mind, or intelligence, or intellect, or reason -- words whose meaning I have been ignorant of until now. But for all that I am a thing which is real and which truly exists. But what kind of a thing? As I have just said -- a thinking thing.

[7] What else am I? I will use my imagination to see if I am not something more. I am not that structure of limbs which is called a human body. I am not even some thin vapour which permeates the limbs -- a wind, fire, air, breath, or whatever I depict in my imagination; for these are things which I have supposed to be nothing. Let this supposition stand; for all that I am still something. And yet may it not perhaps be the case that these very things which I am supposing to be nothing, because they are unknown to me, are in reality identical with the 'I' of which I am aware? I do not know, and for the moment I shall not argue the point, since I can make judgements only about things which are known to me. I know that I exist; the question is, what is this 'I' that I know? If the 'I' is understood strictly as we have been taking it, then it is quite certain that knowledge of it does not depend on things of whose existence I am as yet unaware; so it cannot depend on any of the things which I invent in my imagination. And this very word 'invent' shows me my mistake. It would indeed be a case of fictitious invention if I used my imagination to establish that I was something or other; for imagining is simply contemplating the shape or image of a corporeal thing. Yet now I know for certain both that I exist and at the same time that all such images and, in general, everything relating to the nature of body, could be mere dreams. Once this point has been grasped, to say "I will use my imagination to get to know more distinctly what I am" would seem to be as silly as saying "I am now awake, and see some truth; but since my vision is not yet clear enough, I will deliberately fall asleep so that my dreams may provide a truer and clearer representation." I thus realize that none of the things that the imagination enables me to grasp is at all relevant to this knowledge of myself which I

possess, and that the mind must therefore be most carefully diverted from such things if it is to perceive its own nature as distinctly as possible.

[8] But what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions.

[9] This is a considerable list, if everything on it belongs to me. But does it? Is it not one and the same 'I' who is now doubting almost everything, who nonetheless understands some things, who affirms that this one thing is true, denies everything else, desires to know more, is unwilling to be deceived, imagines many things even involuntarily, and is aware of many things which apparently come from the senses? Are not all these things just as true as the fact that I exist, even if I am asleep all the time, and even if he who created me is doing all he can to deceive me? Which of all these activities is distinct from my thinking? Which of them can be said to be separate from myself? The fact that it is I who am doubting and understanding and willing is so evident that I see no way of making it any clearer. But it is also the case that the 'I' who imagines is the same 'I'. For even if, as I have supposed, none of the objects of imagination are real, the power of imagination is something which really exists and is part of my thinking. Lastly, it is also the same 'I' who has sensory perceptions, or is aware of bodily things as it were through the senses. For example, I am now seeing light, hearing a noise, feeling heat. But I am asleep, so all this is false. Yet I certainly seem to see, to hear, and to be warmed. This cannot be false; what is called 'having a sensory perception' is strictly just this, and in this restricted sense of the term it is simply thinking.

[10] From all this I am beginning to have a rather better understanding of what I am. But it still appears -- and I cannot stop thinking this -- that the corporeal things of which images are formed in my thought, and which the senses investigate, are known with much more distinctness than this puzzling 'I' which cannot be pictured in the imagination. And yet it is surely surprising that I should have a more distinct grasp of things which I realize are doubtful, unknown and foreign to me, than I have of that which is true and known -- my own self. But I see what it is: my mind enjoys wandering off and will not yet submit to being restrained within the bounds of truth. Very well then; just this once let us give it a completely free rein, so that after a while, when it is time to tighten the reins, it may more readily submit to being curbed.

[11] Let us consider the things which people commonly think they understand most distinctly of all; that is, the bodies which we touch and see. I do not mean bodies in general -- for general perceptions are apt to be somewhat more confused -- but one particular body. Let us take, for example, this piece of wax. It has just been taken from the honeycomb; it has not yet quite lost the taste of the honey; it retains some of the scent of the flowers from which it was gathered; its colour, shape and size are plain to see; it is hard, cold and can be handled without difficulty; if you rap it with your knuckle it makes a sound. In short, it has everything which appears necessary to enable a body to be known as distinctly as possible. But even as I speak, I put the wax by the fire, and look: the residual taste is eliminated, the smell goes away, the colour changes, the shape is lost, the size increases; it becomes liquid and hot; you can hardly touch it, and if you strike it, it no longer makes a sound. But does the same wax remain? It must be admitted that it does; no one denies it, no one thinks otherwise. So what was it in the wax that I understood with such distinctness? Evidently none of the features which I arrived at by means of the senses; for whatever came under taste, smell, sight, touch or hearing has now altered -- yet the wax remains.

[12] Perhaps the answer lies in the thought which now comes to my mind; namely, the wax was not after all the sweetness of the honey, or the fragrance of the flowers, or the

whiteness, or the shape, or the sound, but was rather a body which presented itself to me in these various forms a little while ago, but which now exhibits different ones. But what exactly is it that I am now imagining? Let us concentrate, take away everything which does not belong to the wax, and see what is left: merely something extended, flexible and changeable. But what is meant here by 'flexible' and 'changeable'? Is it what I picture in my imagination: that this piece of wax is capable of changing from a round shape to a square shape, or from a square shape to a triangular shape? Not at all; for I can grasp that the wax is capable of countless changes of this kind, yet I am unable to run through this immeasurable number of changes in my imagination, from which it follows that it is not the faculty of imagination that gives me my grasp of the wax as flexible and changeable. And what is meant by 'extended'? Is the extension of the wax also unknown? For it increases if the wax melts, increases again if it boils, and is greater still if the heat is increased. I would not be making a correct judgement about the nature of wax unless I believed it capable of being extended in many more different ways than I will ever encompass in my imagination. I must therefore admit that the nature of this piece of wax is in no way revealed by my imagination, but is perceived by the mind alone. (I am speaking of this particular piece of wax; the point is even clearer with regard to wax in general.) But what is this wax which is perceived by the mind alone? It is of course the same wax which I see, which I touch, which I picture in my imagination, in short the same wax which I thought it to be from the start. And yet, and here is the point, the perception I have of it is a case not of vision or touch or imagination -- nor has it ever been, despite previous appearances -- but of purely mental scrutiny; and this can be imperfect and confused, as it was before, or clear and distinct as it is now, depending on how carefully I concentrate on what the wax consists in.

[13] But as I reach this conclusion I am amazed at how weak and prone to error my mind is. For although I am thinking about these matters within myself, silently and without speaking, nonetheless the actual words bring me up short, and I am almost tricked by ordinary ways of talking. We say that we see the wax itself, if it is there before us, not that we judge it to be there from its colour or shape; and this might lead me to conclude without more ado that knowledge of the wax comes from what the eye sees, and not from the scrutiny of the mind alone. But then if I look out of the window and see men crossing the square, as I just happen to have done, I normally say that I see the men themselves, just as I say that I see the wax. Yet do I see any more than hats and coats which could conceal automatons? I judge that they are men. And so something which I thought I was seeing with my eyes is in fact grasped solely by the faculty of judgement which is in my mind.

[14] However, one who wants to achieve knowledge above the ordinary level should feel ashamed at having taken ordinary ways of talking as a basis for doubt. So let us proceed, and consider on which occasion my perception of the nature of the wax was more perfect and evident. Was it when I first looked at it, and believed I knew it by my external senses, or at least by what they call the "common" sense -- that is, the power of imagination? Or is my knowledge more perfect now, after a more careful investigation of the nature of the wax and of the means by which it is known? Any doubt on this issue would clearly be foolish; for what distinctness was there in my earlier perception? Was there anything in it which an animal could not possess? But when I distinguish the wax from its outward forms -- take the clothes off, as it were, and consider it naked -- then although my judgement may still contain errors, at least my perception now requires a human mind.

[15] But what am I to say about this mind, or about myself? (So far, remember, I am not admitting that there is anything else in me except a mind.) What, I ask, is this 'I' which seems to perceive the wax so distinctly? Surely my awareness of my own self is not merely much truer and more certain than my awareness of the wax, but also much more distinct and evident. For if I judge that the wax exists from the fact that I see it, clearly this same

fact entails much more evidently that I myself also exist. It is possible that what I see is not really the wax; it is possible that I do not even have eyes with which to see anything. But when I see, or think I see (I am not here distinguishing the two), it is simply not possible that I who am now thinking am not something. By the same token, if I judge that the wax exists from the fact that I touch it, the same result follows, namely that I exist. If I judge that it exists from the fact that I imagine it, or for any other reason, exactly the same thing follows. And the result that I have grasped in the case of the wax may be applied to everything else located outside me. Moreover, if my perception of the wax seemed more distinct after it was established not just by sight or touch but by many other considerations, it must be admitted that I now know myself even more distinctly. This is because every consideration whatsoever which contributes to my perception of the wax, or of any other body, cannot but establish even more effectively the nature of my own mind. But besides this, there is so much else in the mind itself which can serve to make my knowledge of it more distinct, that is scarcely seems worth going through the contributions made by considering bodily things.

[16] I see that without any effort I have now finally got back to where I wanted. I now know that even bodies are not strictly perceived by the senses or the faculty of imagination but by the intellect alone, and that this perception derives not from their being touched or seen but from their being understood; and in view of this I know plainly that I can achieve an easier and more evident perception of my own mind than of anything else. But since the habit of holding on to old opinions cannot be set aside so quickly, I should like to stop here and meditate for some time on this new knowledge I have gained, so as to fix it more deeply in my memory.

John Lock

Essay Concerning Human Understanding

Chapter XXVII

Of Identity and Diversity

[1] Wherein identity consists. Another occasion the mind often takes of comparing, is the very being of things, when, considering anything as existing at any determined time and place, we compare it with itself existing at another time, and thereon form the ideas of identity and diversity. When we see anything to be in any place in any instant of time, we are sure (be it what it will) that it is that very thing, and not another which at that same time exists in another place, how like and undistinguishable soever it may be in all other respects: and in this consists identity, when the ideas it is attributed to vary not at all from what they were that moment wherein we consider their former existence, and to which we compare the present. For we never finding, nor conceiving it possible, that two things of the same kind should exist in the same place at the same time, we rightly conclude, that, whatever exists anywhere at any time, excludes all of the same kind, and is there itself alone. When therefore we demand whether anything be the same or no, it refers always to something that existed such a time in such a place, which it was certain, at that instant, was the same with itself, and no other. From whence it follows, that one thing cannot have two beginnings of existence, nor two things one beginning; it being impossible for two things of the same kind to be or exist in the same instant, in the very same place; or one and the same thing in different places. That, therefore, that had one beginning, is the same thing; and that which had a different beginning in time and place from that, is not the same, but diverse. That which has made the difficulty about this relation has been the little care and attention used in having precise notions of the things to which it is attributed.

[2] Identity of substances. We have the ideas but of three sorts of substances: 1. God. 2. Finite intelligences. 3. Bodies. First, God is without beginning, eternal, unalterable, and everywhere, and therefore concerning his identity there can be no doubt. Secondly, Finite spirits having had each its determinate time and place of beginning to exist, the relation to that time and place will always determine to each of them its identity, as long as it exists. Thirdly, The same will hold of every particle of matter, to which no addition or subtraction of matter being made, it is the same. For, though these three sorts of substances, as we term them, do not exclude one another out of the same place, yet we cannot conceive but that they must necessarily each of them exclude any of the same kind out of the same place: or else the notions and names of identity and diversity would be in vain, and there could be no such distinctions of substances, or anything else one from another. For example: could two bodies be in the same place at the same time; then those two parcels of matter must be one and the same, take them great or little; nay, all bodies must be one and the same. For, by the same reason that two particles of matter may be in one place, all bodies may be in one place: which, when it can be supposed, takes away the distinction of identity and diversity of one and more, and renders it ridiculous. But it being a contradiction that two or more should be one, identity and diversity are relations and ways of comparing well founded, and of use to the understanding. Identity of modes and relations. All other things being but modes or relations ultimately terminated in substances, the identity and diversity of each particular existence of them too will be by the same way determined: only as to things whose existence is in succession, such as are the actions of finite beings, v.g. motion and thought, both which consist in a continued train of succession, concerning their diversity there can be no question: because each perishing the moment it begins, they cannot exist in different times, or in different places, as permanent beings can at different

times exist in distant places; and therefore no motion or thought, considered as at different times, can be the same, each part thereof having a different beginning of existence.

[3] Principium Individuationis. From what has been said, it is easy to discover what is so much inquired after, the principium individuationis; and that, it is plain, is existence itself; which determines a being of any sort to a particular time and place, incommunicable to two beings of the same kind. This, though it seems easier to conceive in simple substances or modes; yet, when reflected on, is not more difficult in compound ones, if care be taken to what it is applied: v.g. let us suppose an atom, i.e. a continued body under one immutable superficies, existing in a determined time and place; it is evident, that, considered in any instant of its existence, it is in that instant the same with itself. For, being at that instant what it is, and nothing else, it is the same, and so must continue as long as its existence is continued; for so long it will be the same, and no other. In like manner, if two or more atoms be joined together into the same mass, every one of those atoms will be the same, by the foregoing rule: and whilst they exist united together, the mass, consisting of the same atoms, must be the same mass, or the same body, let the parts be ever so differently jumbled. But if one of these atoms be taken away, or one new one added, it is no longer the same mass or the same body. In the state of living creatures, their identity depends not on a mass of the same particles, but on something else. For in them the variation of great parcels of matter alters not the identity: an oak growing from a plant to a great tree, and then lopped, is still the same oak; and a colt grown up to a horse, sometimes fat, sometimes lean, is all the while the same horse: though, in both these cases, there may be a manifest change of the parts; so that truly they are not either of them the same masses of matter, though they be truly one of them the same oak, and the other the same horse. The reason whereof is, that, in these two cases -- a mass of matter and a living body -- identity is not applied to the same thing.

[4] Identity of vegetables. We must therefore consider wherein an oak differs from a mass of matter, and that seems to me to be in this, that the one is only the cohesion of particles of matter any how united, the other such a disposition of them as constitutes the parts of an oak; and such an organization of those parts as is fit to receive and distribute nourishment, so as to continue and frame the wood, bark, and leaves, &c., of an oak, in which consists the vegetable life. That being then one plant which has such an organization of parts in one coherent body, partaking of one common life, it continues to be the same plant as long as it partakes of the same life, though that life be communicated to new particles of matter vitally united to the living plant, in a like continued organization conformable to that sort of plants. For this organization, being at any one instant in any one collection of matter, is in that particular concrete distinguished from all other, and is that individual life, which existing constantly from that moment both forwards and backwards, in the same continuity of insensibly succeeding parts united to the living body of the plant, it has that identity which makes the same plant, and all the parts of it, parts of the same plant, during all the time that they exist united in that continued organization, which is fit to convey that common life to all the parts so united.

[5] Identity of animals. The case is not so much different in brutes but that any one may hence see what makes an animal and continues it the same. Something we have like this in machines, and may serve to illustrate it. For example, what is a watch? It is plain it is nothing but a fit organization or construction of parts to a certain end, which, when a sufficient force is added to it, it is capable to attain. If we would suppose this machine one continued body, all whose organized parts were repaired, increased, or diminished by a constant addition or separation of insensible parts, with one common life, we should have something very much like the body of an animal; with this difference, That, in an animal the fitness of the organization, and the motion wherein life consists, begin together, the motion

coming from within; but in machines the force coming sensibly from without, is often away when the organ is in order, and well fitted to receive it.

[6] The identity of man. This also shows wherein the identity of the same man consists; viz. in nothing but a participation of the same continued life, by constantly fleeting particles of matter, in succession vitally united to the same organized body. He that shall place the identity of man in anything else, but, like that of other animals, in one fitly organized body, taken in any one instant, and from thence continued, under one organization of life, in several successively fleeting particles of matter united to it, will find it hard to make an embryo, one of years, mad and sober, the same man, by any supposition, that will not make it possible for Seth, Ismael, Socrates, Pilate, St. Austin, and Caesar Borgia, to be the same man. For if the identity of soul alone makes the same man; and there be nothing in the nature of matter why the same individual spirit may not be united to different bodies, it will be possible that those men, living in distant ages, and of different tempers, may have been the same man: which way of speaking must be from a very strange use of the word man, applied to an idea out of which body and shape are excluded. And that way of speaking would agree yet worse with the notions of those philosophers who allow of transmigration, and are of opinion that the souls of men may, for their miscarriages, be detrued into the bodies of beasts, as fit habitations, with organs suited to the satisfaction of their brutal inclinations. But yet I think nobody, could he be sure that the soul of Heliogabalus were in one of his hogs, would yet say that hog were a man or Heliogabalus.

[7] Idea of identity suited to the idea it is applied to. It is not therefore unity of substance that comprehends all sorts of identity, or will determine it in every case; but to conceive and judge of it aright, we must consider what idea the word it is applied to stands for: it being one thing to be the same substance, another the same man, and a third the same person, if person, man, and substance, are three names standing for three different ideas;- for such as is the idea belonging to that name, such must be the identity; which, if it had been a little more carefully attended to, would possibly have prevented a great deal of that confusion which often occurs about this matter, with no small seeming difficulties, especially concerning personal identity, which therefore we shall in the next place a little consider.

[8] Same man. An animal is a living organized body; and consequently the same animal, as we have observed, is the same continued life communicated to different particles of matter, as they happen successively to be united to that organized living body. And whatever is talked of other definitions, ingenious observation puts it past doubt, that the idea in our minds, of which the sound man in our mouths is the sign, is nothing else but of an animal of such a certain form. Since I think I may be confident, that, whoever should see a creature of his own shape or make, though it had no more reason all its life than a cat or a parrot, would call him still a man; or whoever should hear a cat or a parrot discourse, reason, and philosophize, would call or think it nothing but a cat or a parrot; and say, the one was a dull irrational man, and the other a very intelligent rational parrot. A relation we have in an author of great note, is sufficient to countenance the supposition of a rational parrot.

...

[9] Personal identity. This being premised, to find wherein personal identity consists, we must consider what person stands for;- which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it: it being impossible for any one to perceive without perceiving that he does perceive. When we see, hear, smell, taste, feel, meditate, or will anything, we know that we do so. Thus it is always as to our present

sensations and perceptions: and by this every one is to himself that which he calls self: - it not being considered, in this case, whether the same self be continued in the same or divers substances. For, since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity, i.e. the sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.

[10] Consciousness makes personal identity. But it is further inquired, whether it be the same identical substance. This few would think they had reason to doubt of, if these perceptions, with their consciousness, always remained present in the mind, whereby the same thinking thing would be always consciously present, and, as would be thought, evidently the same to itself. But that which seems to make the difficulty is this, that this consciousness being interrupted always by forgetfulness, there being no moment of our lives wherein we have the whole train of all our past actions before our eyes in one view, but even the best memories losing the sight of one part whilst they are viewing another; and we sometimes, and that the greatest part of our lives, not reflecting on our past selves, being intent on our present thoughts, and in sound sleep having no thoughts at all, or at least none with that consciousness which remarks our waking thoughts, -- I say, in all these cases, our consciousness being interrupted, and we losing the sight of our past selves, doubts are raised whether we are the same thinking thing, i.e. the same substance or no. Which, however reasonable or unreasonable, concerns not personal identity at all. The question being what makes the same person; and not whether it be the same identical substance, which always thinks in the same person, which, in this case, matters not at all: different substances, by the same consciousness (where they do partake in it) being united into one person, as well as different bodies by the same life are united into one animal, whose identity is preserved in that change of substances by the unity of one continued life. For, it being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed solely to one individual substance, or can be continued in a succession of several substances. For as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self. For it is by the consciousness it has of its present thoughts and actions, that it is self to itself now, and so will be the same self, as far as the same consciousness can extend to actions past or to come. and would be by distance of time, or change of substance, no more two persons, than a man be two men by wearing other clothes to-day than he did yesterday, with a long or a short sleep between: the same consciousness uniting those distant actions into the same person, whatever substances contributed to their production.

[11] Personal identity in change of substance. That this is so, we have some kind of evidence in our very bodies, all whose particles, whilst vitally united to this same thinking conscious self, so that we feel when they are touched, and are affected by, and conscious of good or harm that happens to them, as a part of ourselves; i.e. of our thinking conscious self. Thus, the limbs of his body are to every one a part of Himself; he sympathizes and is concerned for them. Cut off a hand, and thereby separate it from that consciousness he had of its heat, cold, and other affections, and it is then no longer a part of that which is himself, any more than the remotest part of matter. Thus, we see the substance whereof personal self consisted at one time may be varied at another, without the change of personal identity; there being no question about the same person, though the limbs which but now were a part of it, be cut off.

[12] Personality in change of substance. But the question is, Whether if the same substance which thinks be changed, it can be the same person; or, remaining the same, it can be different persons? And to this I answer: First, This can be no question at all to those who place thought in a purely material animal constitution, void of an immaterial substance. For, whether their supposition be true or no, it is plain they conceive personal identity preserved in something else than identity of substance; as animal identity is preserved in identity of life, and not of substance. And therefore those who place thinking in an immaterial substance only, before they can come to deal with these men, must show why personal identity cannot be preserved in the change of immaterial substances, or variety of particular immaterial substances, as well as animal identity is preserved in the change of material substances, or variety of particular bodies: unless they will say, it is one immaterial spirit that makes the same life in brutes, as it is one immaterial spirit that makes the same person in men; which the Cartesians at least will not admit, for fear of making brutes thinking things too.

[13] Whether in change of thinking substances there can be one person. But next, as to the first part of the question, Whether, if the same thinking substance (supposing immaterial substances only to think) be changed, it can be the same person? I answer, that cannot be resolved but by those who know what kind of substances they are that do think; and whether the consciousness of past actions can be transferred from one thinking substance to another. I grant were the same consciousness the same individual action it could not: but it being a present representation of a past action, why it may not be possible, that that may be represented to the mind to have been which really never was, will remain to be shown. And therefore how far the consciousness of past actions is annexed to any individual agent, so that another cannot possibly have it, will be hard for us to determine, till we know what kind of action it is that cannot be done without a reflex act of perception accompanying it, and how performed by thinking substances, who cannot think without being conscious of it. But that which we call the same consciousness, not being the same individual act, why one intellectual substance may not have represented to it, as done by itself, what it never did, and was perhaps done by some other agent -- why, I say, such a representation may not possibly be without reality of matter of fact, as well as several representations in dreams are, which yet whilst dreaming we take for true -- will be difficult to conclude from the nature of things. And that it never is so, will by us, till we have clearer views of the nature of thinking substances, be best resolved into the goodness of God; who, as far as the happiness or misery of any of his sensible creatures is concerned in it, will not, by a fatal error of theirs, transfer from one to another that consciousness which draws reward or punishment with it. How far this may be an argument against those who would place thinking in a system of fleeting animal spirits, I leave to be considered. But yet, to return to the question before us, it must be allowed, that, if the same consciousness (which, as has been shown, is quite a different thing from the same numerical figure or motion in body) can be transferred from one thinking substance to another, it will be possible that two thinking substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved.

[14] Whether, the same immaterial substance remaining, there can be two persons. As to the second part of the question, Whether the same immaterial substance remaining, there may be two distinct persons; which question seems to me to be built on this, -- Whether the same immaterial being, being conscious of the action of its past duration, may be wholly stripped of all the consciousness of its past existence, and lose it beyond the power of ever retrieving it again: and so as it were beginning a new account from a new period, have a consciousness that cannot reach beyond this new state. All those who hold pre-existence are evidently of this mind; since they allow the soul to have no remaining consciousness of what it did in that pre-existent state, either wholly separate from body, or informing any

other body; and if they should not, it is plain experience would be against them. So that personal identity, reaching no further than consciousness reaches, a pre-existent spirit not having continued so many ages in a state of silence, must needs make different persons. Suppose a Christian Platonist or a Pythagorean should, upon God's having ended all his works of creation the seventh day, think his soul hath existed ever since; and should imagine it has revolved in several human bodies; as I once met with one, who was persuaded his had been the soul of Socrates (how reasonably I will not dispute; this I know, that in the post he filled, which was no inconsiderable one, he passed for a very rational man, and the press has shown that he wanted not parts or learning;) -- would any one say, that he, being not conscious of any of Socrates's actions or thoughts, could be the same person with Socrates? Let any one reflect upon himself, and conclude that he has in himself an immaterial spirit, which is that which thinks in him, and, in the constant change of his body keeps him the same: and is that which he calls himself: let him also suppose it to be the same soul that was in Nestor or Thersites, at the siege of Troy, (for souls being, as far as we know anything of them, in their nature indifferent to any parcel of matter, the supposition has no apparent absurdity in it), which it may have been, as well as it is now the soul of any other man: but he now having no consciousness of any of the actions either of Nestor or Thersites, does or can he conceive himself the same person with either of them? Can he be concerned in either of their actions? attribute them to himself, or think them his own, more than the actions of any other men that ever existed? So that this consciousness, not reaching to any of the actions of either of those men, he is no more one self with either of them than if the soul or immaterial spirit that now informs him had been created, and began to exist, when it began to inform his present body; though it were never so true, that the same spirit that informed Nestor's or Thersites' body were numerically the same that now informs his. For this would no more make him the same person with Nestor, than if some of the particles of matter that were once a part of Nestor were now a part of this man; the same immaterial substance, without the same consciousness, no more making the same person, by being united to any body, than the same particle of matter, without consciousness, united to any body, makes the same person. But let him once find himself conscious of any of the actions of Nestor, he then finds himself the same person with Nestor.

[15] The body, as well as the soul, goes to the making of a man. And thus may we be able, without any difficulty, to conceive the same person at the resurrection, though in a body not exactly in make or parts the same which he had here, -- the same consciousness going along with the soul that inhabits it. But yet the soul alone, in the change of bodies, would scarce to any one but to him that makes the soul the man, be enough to make the same man. For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same person with the prince, accountable only for the prince's actions: but who would say it was the same man? The body too goes to the making the man, and would, I guess, to everybody determine the man in this case, wherein the soul, with all its princely thoughts about it, would not make another man: but he would be the same cobbler to every one besides himself. I know that, in the ordinary way of speaking, the same person, and the same man, stand for one and the same thing. And indeed every one will always have a liberty to speak as he pleases, and to apply what articulate sounds to what ideas he thinks fit, and change them as often as he pleases. But yet, when we will inquire what makes the same spirit, man, or person, we must fix the ideas of spirit, man, or person in our minds; and having resolved with ourselves what we mean by them, it will not be hard to determine, in either of them, or the like, when it is the same, and when not.

[16] Consciousness alone unites actions into the same person. But though the same immaterial substance or soul does not alone, wherever it be, and in whatsoever state, make

the same man; yet it is plain, consciousness, as far as ever it can be extended -- should it be to ages past -- unites existences and actions very remote in time into the same person, as well as it does the existences and actions of the immediately preceding moment: so that whatever has the consciousness of present and past actions, is the same person to whom they both belong. Had I the same consciousness that I saw the ark and Noah's flood, as that I saw an overflowing of the Thames last winter, or as that I write now, I could no more doubt that I who write this now, that saw' the Thames overflowed last winter, and that viewed the flood at the general deluge, was the same self, -- place that self in what substance you please -- than that I who write this am the same myself now whilst I write (whether I consist of all the same substance, material or immaterial, or no) that I was yesterday. For as to this point of being the same self, it matters not whether this present self be made up of the same or other substances -- I being as much concerned, and as justly accountable for any action that was done a thousand years since, appropriated to me now by this self-consciousness, as I am for what I did the last moment.

[17] Self depends on consciousness, not on substance. Self is that conscious thinking thing, -- whatever substance made up of, (whether spiritual or material, simple or compounded, it matters not) -- which is sensible or conscious of pleasure and pain, capable of happiness or misery, and so is concerned for itself, as far as that consciousness extends. Thus every one finds that, whilst comprehended under that consciousness, the little finger is as much a part of himself as what is most so. Upon separation of this little finger, should this consciousness go along with the little finger, and leave the rest of the body, it is evident the little finger would be the person, the same person; and self then would have nothing to do with the rest of the body. As in this case it is the consciousness that goes along with the substance, when one part is separate from another, which makes the same person, and constitutes this inseparable self: so it is in reference to substances remote in time. That with which the consciousness of this present thinking thing can join itself, makes the same person, and is one self with it, and with nothing else; and so attributes to itself, and owns all the actions of that thing, as its own, as far as that consciousness reaches, and no further; as every one who reflects will perceive. ...

[18] Persons, not substances, the objects of reward and punishment. In this personal identity is founded all the right and justice of reward and punishment; happiness and misery being that for which every one is concerned for himself, and not mattering what becomes of any substance, not joined to, or affected with that consciousness. For, as it is evident in the instance I gave but now, if the consciousness went along with the little finger when it was cut off, that would be the same self which was concerned for the whole body yesterday, as making part of itself, whose actions then it cannot but admit as its own now. Though, if the same body should still live, and immediately from the separation of the little finger have its own peculiar consciousness, whereof the little finger knew nothing, it would not at all be concerned for it, as a part of itself, or could own any of its actions, or have any of them imputed to him.

[19] Which shows wherein personal identity consists. This may show us wherein personal identity consists: not in the identity of substance, but, as I have said, in the identity of consciousness, wherein if Socrates and the present mayor of Queinborough agree, they are the same person: if the same Socrates waking and sleeping do not partake of the same consciousness, Socrates waking and sleeping is not the same person. And to punish Socrates waking for what sleeping Socrates thought, and waking Socrates was never conscious of, would be no more of right, than to punish one twin for what his brother-twin did, whereof he knew nothing, because their outsides were so like, that they could not be distinguished; for such twins have been seen.

Where Am I?

Dan Dennett

[1] Now that I've won my suit under the Freedom of Information Act, I am at liberty to reveal for the first time a curious episode in my life that may be of interest not only to those engaged in research in the philosophy of mind, artificial intelligence and neuroscience but also to the general public.

[2] Several years ago I was approached by Pentagon officials who asked me to volunteer for a highly dangerous and secret mission. In collaboration with NASA and Howard Hughes, the Department of Defense was spending billions to develop a Supersonic Tunneling Underground Device, or STUD. It was supposed to tunnel through the earth's core at great speed and deliver a specially designed atomic warhead "right up the Red's missile silos," as one of the Pentagon brass put it.

[3] The problem was that in an early test they had succeeded in lodging a warhead about a mile deep under Tulsa, Oklahoma, and they wanted me to retrieve it for them. "Why me?" I asked. Well, the mission involved some pioneering applications of current brain research, and they had heard of my interest in brains and of course my Faustian curiosity and great courage and so forth ... Well, how could I refuse? The difficulty that brought the Pentagon to my door was that the device I'd been asked to recover was fiercely radioactive, in a new way. According to monitoring instruments, something about the nature of the device and its complex interactions with pockets of material deep in the earth had produced radiation that could cause severe abnormalities in certain tissues of the brain. No way had been found to shield the brain from these deadly rays, which were apparently harmless to other tissues and organs of the body. So it had been decided that the person sent to recover the device should leave his brain behind. It would be kept in a safe place where it could execute its normal control functions by elaborate radio links. Would I submit to a surgical procedure that would completely remove my brain, which would then be placed in a life-support system at the Manned Spacecraft Center in Houston? Each input and output pathway, as it was severed, would be restored by a pair of microminiaturized radio transceivers, one attached precisely to the brain, the other to the nerve stumps in the empty cranium. No information would be lost, all the connectivity would be preserved. At first I was a bit reluctant. Would it really work? The Houston brain surgeons encouraged me. "Think of it," they said, "as a mere stretching of the nerves. If your brain were just moved over an inch in your skull, that would not alter or impair your mind. We're simply going to make the nerves indefinitely elastic by splicing radio links into them."

[4] I was shown around the life-support lab in Houston and saw the sparkling new vat in which my brain would be placed, were I to agree. I met the large and brilliant support team of neurologists, hematologists, biophysicists, and electrical engineers, and after several days of discussions and demonstrations, I agreed to give it a try. I was subjected to an enormous array of blood tests, brain scans, experiments, interviews, and the like. They took down my autobiography at great length, recorded tedious lists of my beliefs, hopes, fears, and tastes. They even listed my favorite stereo recordings and gave me a crash session of psychoanalysis.

[5] The day for surgery arrived at last and of course I was anesthetized and remember nothing of the operation itself. When I came out of anesthesia, I opened my eyes, looked around, and asked the inevitable, the traditional, the lamentably hackneyed post-operative question:

[6] "Where am I?" The nurse smiled down at me. "You're in Houston," she said, and I reflected that this still had a good chance of being the truth one way or another. She handed me a mirror. Sure enough, there were the tiny antennae poking up through their titanium ports cemented into my skull.

[7] "I gather the operation was a success," I said, "I want to go see my brain." They led me (I was a bit dizzy and unsteady) down a long corridor and into the life-support lab. A cheer went up from the assembled support team, and I responded with what I hoped was a jaunty salute. Still feeling lightheaded, I was helped over to the life-support vat. I peered through the glass. There, floating in what looked like ginger-ale, was undeniably a human brain, though it was almost covered with printed circuit chips, plastic tubules, electrodes, and other paraphernalia. "Is that mine?" I asked. "Hit the output transmitter switch there on the side of the vat and see for yourself," the project director replied. I moved the switch to OFF, and immediately slumped, groggy and nauseated, into the arms of the technicians, one of whom kindly restored the switch to its ON position. While I recovered my equilibrium and composure, I thought to myself: "Well, here I am, sitting on a folding chair, staring through a piece of plate glass at my own brain. ... But wait," I said to myself, "shouldn't I have thought, 'Here I am, suspended in a bubbling fluid, being stared at by my own eyes?'" I tried to think this latter thought. I tried to project it into the tank, offering it hopefully to my brain, but I failed to carry off the exercise with any conviction. I tried again. "Here am I, Daniel Dennett, suspended in a bubbling fluid, being stared at by my own eyes." No, it just didn't work. Most puzzling and confusing. Being a philosopher of firm physicalist conviction I believed unswervingly that the tokening of my thoughts was occurring somewhere in my brain: yet, when I thought "Here I am," where the thought occurred to me was here, outside the vat, where I, Dennett, was standing staring at my brain.

[8] I tried and tried to think myself into the vat, but to no avail. I tried to build up to the task by doing mental exercises. I thought to myself, "The sun is shining over there," five times in rapid succession, each time mentally ostending a different place: in order, the sunlit corner of the lab, the visible front lawn of the hospital, Houston, Mars, and Jupiter. I found I had little difficulty in getting my "there's" to hop all over the celestial map with their proper references. I could loft a "there" in an instant through the farthest reaches of space, and then aim the next "there" with pinpoint accuracy at the upper left quadrant of a freckle on my arm. Why was I having such trouble with "here"? "Here in Houston" worked well enough, and so did "here in the lab," and even "here in this part of the lab," but "here in the vat" always seemed merely an unmeant mental mouthing. I tried closing my eyes while thinking it. This seemed to help, but still I couldn't manage to pull it off, except perhaps for a fleeting instant. I couldn't be sure. The discovery that I couldn't be sure was also unsettling. How did I know where I meant by "here" when I thought "here"? Could I think I meant one place when in fact I meant another? I didn't see how that could be admitted without untying the few bonds of intimacy between a person and his own mental life that had survived the onslaught of the brain scientists and philosophers, the physicalists and behaviorists. Perhaps I was incorrigible about where I meant when I said "here." But in my present circumstances it seemed that either I was doomed by sheer force of mental habit to thinking systematically false indexical thoughts, or where a person is (and hence where his thoughts are tokened for purposes of semantic analysis) is not necessarily where his brain, the physical seat of his soul, resides. Nagged by confusion, I attempted to orient myself by falling back on a favorite philosopher's ploy. I began naming things.

[9] "Yorick," I said aloud to my brain, "you are my brain. The rest of my body, seated in this chair, I dub 'Hamlet.'" So here we all are: Yorick's my brain, Hamlet's my body, and I am Dennett. Now, where am I? And when I think "where am I?" where's that thought tokened?

Is it tokened in my brain, lounging about in the vat, or right here between my ears where it seems to be tokened? Or nowhere? Its temporal coordinates give me no trouble; must it not have spatial coordinates as well? I began making a list of the alternatives.

[10] (1) Where Hamlet goes, there goes Dennett. This principle was easily refuted by appeal to the familiar brain transplant thought-experiments so enjoyed by philosophers. If Tom and Dick switch brains, Tom is the fellow with Dick's former body -- just ask him; he'll claim to be Tom, and tell you the most intimate details of Tom's autobiography. It was clear enough, then, that my current body and I could part company, but not likely that I could be separated from my brain. The rule of thumb that emerged so plainly from the thought experiments was that in a brain-transplant operation, one wanted to be the donor, not the recipient. Better to call such an operation a body-transplant, in fact. So perhaps the truth was,

[11] (2) Where Yorick goes, there goes Dennett. This was not at all appealing, however. How could I be in the vat and not about to go anywhere, when I was so obviously outside the vat looking in and beginning to make guilty plans to return to my room for a substantial lunch? This begged the question I realized, but it still seemed to be getting at something important. Casting about for some support for my intuition, I hit upon a legalistic sort of argument that might have appealed to Locke.

[12] Suppose, I argued to myself, I were now to fly to California, rob a bank, and be apprehended. In which state would I be tried: In California, where the robbery took place, or in Texas, where the brains of the outfit were located? Would I be a California felon with an out-of-state brain, or a Texas felon remotely controlling an accomplice of sorts in California? It seemed possible that I might beat such a rap just on the undecidability of that jurisdictional question, though perhaps it would be deemed an inter-state, and hence Federal, offense. In any event, suppose I were convicted. Was it likely that California would be satisfied to throw Hamlet into the brig, knowing that Yorick was living the good life and luxuriously taking the waters in Texas? Would Texas incarcerate Yorick, leaving Hamlet free to take the next boat to Rio? This alternative appealed to me. Barring capital punishment or other cruel and unusual punishment, the state would be obliged to maintain the life-support system for Yorick though they might move him from Houston to Leavenworth, and aside from the unpleasantness of the opprobrium, I, for one, would not mind at all and would consider myself a free man under those circumstances. If the state has an interest in forcibly relocating persons in institutions, it would fail to relocate me in any-institution by locating Yorick there. If this were true, it suggested. a third alternative.

[13] (3) Dennett is wherever he thinks he is. Generalized, the claim was as follows: At any given time a person has a point of view, and the location of the point of view (which is determined internally by the content of the point of view) is also the location of the person.

[14] Such a proposition is not without its perplexities, but to me it seemed a step in the right direction. The only trouble was that it seemed to place one in a heads-I-win/tails-you-lose situation of unlikely infallibility as regards location. Hadn't I myself often been wrong about where I was, and at least as often uncertain? Couldn't one get lost? Of course, but getting lost geographically is not the only way one might get lost. If one were lost in the woods one could attempt to reassure oneself with the consolation that at least one knew where one was: one was right here in the familiar surroundings of one's own body. Perhaps in this case one would not have drawn one's attention to much to be thankful for. Still, there were worse plights imaginable, and I wasn't sure I wasn't in such a plight right now.

[15] Point of view clearly had something to do with personal location, but it was itself an unclear notion. It was obvious that the content of one's point of view was not the same as or determined by the content of one's beliefs or thoughts. For example, what should we say about the point of view of the Cinerama viewer who shrieks and twists in his seat as the roller-coaster footage overcomes his psychic distancing? Has he forgotten that he is safely seated in the theater? Here I was inclined to say that the person is experiencing an illusory shift in point of view. In other cases, my inclination to call such shifts illusory was less strong. The workers in laboratories and plants who handle dangerous materials by operating feedback-controlled mechanical arms and hands undergo a shift in point of view that is crisper and more pronounced than anything Cinerama can provoke. They can feel the heft and slipperiness of the containers they manipulate with their metal fingers. They know perfectly well where they are and are not fooled into false beliefs by the experience, yet it is as if they were inside the isolation chamber they are peering into. With mental effort, they can manage to shift their point of view back and forth, rather like making a transparent Necker cube or an Escher drawing change orientation before one's eyes. It does seem extravagant to suppose that in performing this bit of mental gymnastics, they are transporting themselves back and forth.

[16] Still their example gave me hope. If I was in fact in the vat in spite of my intuitions, I might be able to train myself to adopt that point of view even as a matter of habit. I should dwell on images of myself comfortably floating in my vat, beaming volitions to that familiar body out there. I reflected that the ease or difficulty of this task was presumably independent of the truth about the location of one's brain. Had I been practicing before the operation, I might now be finding it second nature. You might now yourself try such a *tromp l'oeil*. Imagine you have written an inflammatory letter which has been published in the Times, the result of which is that the Government has chosen to impound your brain for a probationary period of three years in its Dangerous Brain Clinic in Bethesda, Maryland. Your body of course is allowed freedom to earn a salary and thus to continue its function of laying up income to be taxed. At this moment, however, your body is seated in an auditorium listening to a peculiar account by Daniel Dennett of his own similar experience. Try it. Think yourself to Bethesda, and then hark back longingly to your body, far away, and yet seeming so near. It is only with long-distance restraint (yours? the Government's?) that you can control your impulse to get those hands clapping in polite applause before navigating the old body to the rest room and a well-deserved glass of evening sherry in the lounge. The task of imagination is certainly difficult, but if you achieve your goal the results might be consoling.

[17] Anyway, there I was in Houston, lost in thought as one might say, but not for long. My speculations were soon interrupted by the Houston doctors, who wished to test out my new prosthetic nervous system before sending me off on my hazardous mission. As I mentioned before, I was a bit dizzy at first, and not surprisingly, although I soon habituated myself to my new circumstances (which were, after all, well nigh indistinguishable from my old circumstances). My accommodation was not perfect, however, and to this day I continue to be plagued by minor coordination difficulties. The speed of light is fast, but finite, and as my brain and body move farther and farther apart, the delicate interaction of my feedback systems is thrown into disarray by the time lags. Just as one is rendered close to speechless by a delayed or echoic hearing of one's speaking voice so, for instance, I am virtually unable to track a moving object with my eyes whenever my brain and my body are more than a few miles apart. In most matters my impairment is scarcely detectable, though I can no longer hit a slow curve ball with the authority of yore. There are some compensations of course. Though liquor tastes as good as ever, and warms my gullet while corroding my liver, I can drink it in any quantity I please, without becoming the slightest bit inebriated, a curiosity some of my close friends may have noticed (though I occasionally have feigned

inebriation, so as not to draw attention to my unusual circumstances). For similar reasons, I take aspirin orally for a sprained wrist, but if the pain persists I ask Houston to administer codeine to me in vitro. In times of illness the phone bill can be staggering.

[18] But to return to my adventure. At length, both the doctors and I were satisfied that I was ready to undertake my subterranean mission. And so I left my brain in Houston and headed by helicopter for Tulsa. Well, in any case, that's the way it seemed to me. That's how I would put it, just off the top of my head as it were. On the trip I reflected further about my earlier anxieties and decided that my first post-operative speculations had been tinged with panic. The matter was not nearly as strange or metaphysical as I had been supposing. Where was I? In two places, clearly: both inside the vat and outside it. Just as one can stand with one foot in Connecticut and the other in Rhode Island, I was in two places at once. I had become one of those scattered individuals we used to hear so much about. The more I considered this answer, the more obviously true it appeared. But, strange to say, the more true it appeared, the less important the question to which it could be the true answer seemed. A sad, but not unprecedented, fate for a philosophical question to suffer. This answer did not completely satisfy me, of course. There lingered some question to which I should have liked an answer, which was neither "Where are all my various and sundry parts?" nor "What is my current point of view?" Or at least there seemed to be such a question. For it did seem undeniable that in some sense I and not merely most of me was descending into the earth under Tulsa in search of an atomic warhead.

[19] When I found the warhead, I was certainly glad I had left my brain behind, for the pointer on the specially built Geiger counter I had brought with me was off the dial. I called Houston on my ordinary radio and told the operation control center of my position and my progress. In return, they gave me instructions for dismantling the vehicle, based upon my on-site observations. I had set to work with my cutting torch when all of a sudden a terrible thing happened. I went stone deaf. At first I thought it was only my radio earphones that had broken, but when I tapped on my helmet, I heard nothing. Apparently the auditory transceivers had gone on the fritz. I could no longer hear Houston or my own voice, but I could speak, so I started telling them what had happened. In midsentence, I knew something else had gone wrong. My vocal apparatus had become paralyzed. Then my right hand went limp -- another transceiver had gone. I was truly in deep trouble. But worse was to follow. After a few more minutes, I went blind. I cursed my luck, and then I cursed the scientists who had led me into this grave peril. There I was, deaf, dumb, and blind, in a radioactive hole more than a mile under Tulsa. Then the last of my cerebral radio links broke, and suddenly I was faced with a new and even more shocking problem: whereas an instant before I had been buried alive in Oklahoma, now I was disembodied in Houston. My recognition of my new status was not immediate. It took me several very anxious minutes before it dawned on me that my poor body lay several hundred miles away, with heart pulsing and lungs respirating, but otherwise as dead as the body of any heart transplant donor, its skull packed with useless, broken electronic gear. The shift in perspective I had earlier found well nigh impossible now seemed quite natural. Though I could think myself back into my body in the tunnel under Tulsa, it took some effort to sustain the illusion. For surely it was an illusion to suppose I was still in Oklahoma: I had lost all contact with that body.

[20] It occurred to me then, with one of those rushes of revelation of which we should be suspicious, that I had stumbled upon an impressive demonstration of the immateriality of the soul based upon physicalist principles and premises. For as the last radio signal between Tulsa and Houston died away, had I not changed location from Tulsa to Houston at the speed of light? And had I not accomplished this without any increase in mass? What moved from A to B at such speed was surely myself, or at any rate my soul or mind -- the massless

center of my being and home of my consciousness. My point of view had lagged somewhat behind, but I had already noted the indirect bearing of point of view on personal location. I could not see how a physicalist philosopher could quarrel with this except by taking the dire and counter-intuitive route of banishing all talk of persons. Yet the notion of personhood was so well entrenched in everyone's world view, or so it seemed to me, that any denial would be as curiously unconvincing, as systematically disingenuous, as the Cartesian negation, "non sum."

[21] The joy of philosophic discovery thus tided me over some very bad minutes or perhaps hours as the helplessness and hopelessness of my situation became more apparent to me. Waves of panic and even nausea swept over me, made all the more horrible by the absence of their normal body-dependent phenomenology. No adrenalin rush of tingles in the arms, no pounding heart, no premonitory salivation. I did feel a dread sinking feeling in my bowels at one point, and this tricked me momentarily into the false hope that I was undergoing a reversal of the process that landed me in this fix -- a gradual undisembodiment. But the isolation and uniqueness of that twinge soon convinced me that it was simply the first of a plague of phantom body hallucinations that I, like any other amputee, would be all too likely to suffer.

[22] My mood then was chaotic. On the one hand, I was fired up with elation at my philosophic discovery and was wracking my brain (one of the few familiar things I could still do), trying to figure out how to communicate my discovery to the journals; while on the other was bitter, lonely, and filled with dread and uncertainty. Fortunately, this did not last long, for my technical support team sedated me into a dreamless sleep from which I awoke, hearing with magnificent fidelity the familiar opening strains of my favorite Brahms piano trio. So that was why they had wanted a list of my favorite recordings! It did not take me long to realize that I was hearing the music without ears. The output from the stereo stylus was being fed through some fancy rectification circuitry directly into my auditory nerve. I was mainlining Brahms, an unforgettable experience for any stereo buff. At the end of the record it did not surprise me to hear the reassuring voice of the project director speaking into a microphone that was now my prosthetic ear. He confirmed my analysis of what had gone wrong and assured me that steps were being taken to reembody me. He did not elaborate, and after a few more recordings, I found myself drifting off to sleep. My sleep lasted, I later learned, for the better part of a year, and when I awoke, it was to find myself fully restored to my senses. When I looked into the mirror, though, I was a bit startled to see an unfamiliar face. Bearded and a bit heavier, bearing no doubt a family resemblance to my former face, and with the same look of spritely intelligence and resolute character, but definitely a new face. Further self-explorations of an intimate nature left me no doubt that this was a new body and the project director confirmed my conclusions. He did not volunteer any information on the past history of my new body and I decided (wisely, I think in retrospect) not to pry. As many philosophers unfamiliar with my ordeal have more recently speculated, the acquisition of a new body leaves one's person intact. And after a period of adjustment to a new voice, new muscular strengths and weaknesses, and so forth, one's personality is by and large also preserved. More dramatic changes in personality have been routinely observed in people who have undergone extensive plastic surgery, to say nothing of sex change operations, and I think no one contests the survival of the person in such cases. In any event I soon accommodated to my new body, to the point of being unable to recover any of its novelties to my consciousness or even memory. The view in the mirror soon became utterly familiar. That view, by the way, still revealed antennae, and so I was not surprised to learn that my brain had not been moved from its haven in the life-support lab.

[23] I decided that good old Yorick deserved a visit. I and my new body, whom we might as well call Fortinbras, strode into the familiar lab to another round of applause from the technicians, who were of course congratulating themselves, not me. Once more I stood before the vat and contemplated poor Yorick, and on a whim I once again cavalierly flicked off the output transmitter switch. Imagine my surprise when nothing unusual happened. No fainting spell, no nausea, no noticeable change. A technician hurried to restore the switch to ON, but still I felt nothing. I demanded an explanation, which the project director hastened to provide. It seems that before they had even operated on the first occasion, they had constructed a computer duplicate of my brain, reproducing both the complete information processing structure and the computational speed of my brain in a giant computer program. After the operation, but before they had dared to send me off on my mission to Oklahoma, they had run this computer system and Yorick side by side. The incoming signals from Hamlet were sent simultaneously to Yorick's transceivers and to the computer's array of inputs. And the outputs from Yorick were not only beamed back to Hamlet, my body; they were recorded and checked against the simultaneous output of the computer program, which was called "Hubert" for reasons obscure to me. Over days and even weeks, the outputs were identical and synchronous, which of course did not prove that they had succeeded in copying the brain's functional structure, but the empirical support was greatly encouraging.

[24] Hubert's input, and hence activity, had been kept parallel with Yorick's during my disembodied days. And now, to demonstrate this, they had actually thrown the master switch that put Hubert for the first time in on-line control of my body -- not Hamlet, of course, but Fortinbras. (Hamlet, I learned, had never been recovered from its underground tomb and could be assumed by this time to have largely returned to the dust. At the head of my grave still lay the magnificent bulk of the abandoned device, with the word STUD emblazoned on its side in large letters -- a circumstance which may provide archeologists of the next century with a curious insight into the burial rites of their ancestors.)

[25] The laboratory technicians now showed me the master switch, which had two positions, labeled B, for Brain (they didn't know my brain's name was Yorick) and H, for Hubert. The switch did indeed point to H, and they explained to me that if I wished, I could switch it back to B. With my heart in my mouth (and my brain in its vat), I did this. Nothing happened. A click, that was all. To test their claim, and with the master switch now set at B, I hit Yorick's output transmitter switch on the vat and sure enough, I began to faint. Once the output switch was turned back on and I had recovered my wits, so to speak, I continued to play with the master switch, flipping it back and forth. I found that with the exception of the transitional click, I could detect no trace of a difference. I could switch in mid-utterance, and the sentence I had begun speaking under the control of Yorick was finished without a pause or hitch of any kind under the control of Hubert. I had a spare brain, a prosthetic device which might some day stand me in very good stead, were some mishap to befall Yorick. Or alternatively, I could keep Yorick as a spare and use Hubert. It didn't seem to make any difference which I chose, for the wear and tear and fatigue on my body did not have any debilitating effect on either brain, whether or not it was actually causing the motions of my body, or merely spilling its output into thin air.

[26] The one truly unsettling aspect of this new development was the prospect, which was not long in dawning on me, of someone detaching the spare -- Hubert or Yorick, as the case might be -- from Fortinbras and hitching it to yet another body -- some Johnny-come-lately Rosencrantz or Guildenstern. Then (if not before) there would be two people, that much was clear. One would be me, and the other would be a sort of super-twin brother. If there were two bodies, one under the control of Hubert and the other being controlled by Yorick, then which would the world recognize as the true Dennett? And whatever the rest of the world

decided, which one would be me? Would I be the Yorick-brained one, in virtue of Yorick's causal priority and former intimate relationship with the original Dennett body, Hamlet? That seemed a bit legalistic, a bit too redolent of the arbitrariness of consanguinity and legal possession, to be convincing at the metaphysical level. For, suppose that before the arrival of the second body on the scene, I had been keeping Yorick as the spare for years, and letting Hubert's output drive my body -- that is, Fortinbras -- all that time. The Hubert-Fortinbras couple would seem then by squatter's rights (to combat one legal intuition with another) to be the true Dennett and the lawful inheritor of everything that was Dennett's. This was an interesting question, certainly, but not nearly so pressing as another question that bothered me. My strongest intuition was that in such an eventuality I would survive so long as either brain-body couple remained intact, but I had mixed emotions about whether I should want both to survive.

[27] I discussed my worries with the technicians and the project director. The prospect of two Dennetts was abhorrent to me, I explained, largely for social reasons. I didn't want to be my own rival for the affections of my wife, nor did I like the prospect of the two Dennetts sharing my modest professor's salary. Still more vertiginous and distasteful, though, was the idea of knowing that much about another person, while he had the very same goods on me. How could we ever face each other? My colleagues in the lab argued that I was ignoring the bright side of the matter. Weren't there many things I wanted to do but, being only one person, had been unable to do? Now one Dennett could stay at home and be the professor and family man, while the other could strike out on a life of travel and adventure -- missing the family of course, but happy in the knowledge that the other Dennett was keeping the home fires burning. I could be faithful and adulterous at the same time. I could even cuckold myself -- to say nothing of other more lurid possibilities my colleagues were all too ready to force upon my overtaxed imagination. But my ordeal in Oklahoma (or was it Houston?) had made me less adventurous, and I shrank from this opportunity that was being offered (though of course I was never quite sure it was being offered to me in the first place).

[28] There was another prospect even more disagreeable -- that the spare, Hubert or Yorick as the case might be, would be detached from any input from Fortinbras and just left detached. Then, as in the other case, there would be two Dennetts, or at least two claimants to my name and possessions, one embodied in Fortinbras, and the other sadly, miserably disembodied. Both selfishness and altruism bade me take steps to prevent this from happening. So I asked that measures be taken to ensure that no one could ever tamper with the transceiver connections or the master switch without my (our? no, my) knowledge and consent. Since I had no desire to spend my life guarding the equipment in Houston, it was mutually decided that all the electronic connections in the lab would be carefully locked: both those that controlled the life-support system for Yorick and those that controlled the power supply for Hubert would be guarded with fail-safe devices, and I would take the only master switch, outfitted for radio remote control, with me wherever I went. I carry it strapped around my waist and -- wait a moment -- here it is. Every few months I reconnoiter the situation by switching channels. I do this only in the presence of friends of course, for if the other channel were, heaven forbid, either dead or otherwise occupied, there would have to be somebody who had my interests at heart to switch it back, to bring me back from the void. For while I could feel, see, hear and otherwise sense whatever befell my body, subsequent to such a switch, I'd be unable to control it. By the way, the two positions on the switch are intentionally unmarked, so I never have the faintest idea whether I am switching from Hubert to Yorick or vice versa. (Some of you may think that in this case I really don't know who I am, let alone where I am. But such reflections no longer make much of a dent on my essential Dennett-ness, on my own sense of who I am. If it is

true that in one sense I don't know who I am then that's another one of your philosophical truths of underwhelming significance.)

[29] In any case, every time I've flipped the switch so far, nothing has happened. So let's give it a try....

[30] "THANK GOD! I THOUGHT YOU'D NEVER FLIP THAT SWITCH!

[31] You can't imagine how horrible it's been these last two weeks -- but now you know, it's your turn in purgatory. How I've longed for this moment! You see, about two weeks ago -- excuse me, ladies and gentle-men, but I've got to explain this to my ... um, brother, I guess you could say, but he's just told you the facts, so you'll understand -- about two weeks ago our two brains drifted just a bit out of synch. I don't know whether my brain is now Hubert or Yorick, any more than you do, but in any case, the two brains drifted apart, and of course once the process started, it snowballed, for I was in a slightly different receptive state for the input we both received, a difference that was soon magnified. In no time at all the illusion that I was in control of my body -- our body -- was completely dissipated. There was nothing I could do -- no way to call you. YOU DIDN'T EVEN KNOW I EXISTED! It's been like being carried around in a cage, or better, like being possessed -- hearing my own voice say things I didn't mean to say, watching in frustration as my own hands performed deeds I hadn't intended. You'd scratch our itches, but not the way I would have, and you kept me awake, with your tossing and turning. I've been totally exhausted, on the verge of a nervous breakdown, carried around helplessly by your frantic round of activities, sustained only by the knowledge that some day you'd throw the switch.

[32] "Now it's your turn, but at least you'll have the comfort of knowing I know you're in there. Like an expectant mother, I'm eating -- or at any rate tasting, smelling, seeing-for two now, and I'll try to make it easy for you. Don't worry. Just as soon as this colloquium is over, you and I will fly to Houston, and we'll see what can be done to get one of us another body. You can have a female body -- your body could be any color you like. But let's think it over. I tell you what -- to be fair, if we both want this body, I promise I'll let the project director flip a coin to settle which of us gets to keep it and which then gets to choose a new body. That should guarantee justice, shouldn't it? In any case, I'll take care of you, I promise. These people are my witnesses.

[34] "Ladies and gentlemen, this talk we have just heard is not exactly the talk I would have given, but I assure you that everything he said was perfectly true. And now if you'll excuse me, I think I'd -- we'd -- better sit down."

The Self as a Center of Narrative Gravity

Dan Dennett

[1] What is a self? I will try to answer this question by developing an analogy with something much simpler, something which is nowhere near as puzzling as a self, but has some properties in common with selves. What I have in mind is *the center of gravity* of an object.

[2] This is a well-behaved concept in Newtonian physics. But a center of gravity is not an atom or a subatomic particle or any other physical item in the world. It has no mass; it has no color; it has no physical properties at all, except for spatio-temporal location. It is a fine example of what Hans Reichenbach would call an *abstractum*. It is a purely abstract object. It is, if you like, a theorist's fiction. It is not one of the real things in the universe in addition to the atoms. But it is a fiction that has nicely defined, well delineated and well behaved role within physics. Let me remind you how robust and familiar the idea of a center of gravity is.

[3] Consider a chair. Like all other physical objects, it has a center of gravity. If you start tipping it, you can tell more or less accurately whether it would start to fall over or fall back in place if you let go of it. We're all quite good at making predictions involving centers of gravity and devising explanations about when and why things fall over. Place a book on the chair. It, too, has a center of gravity. If you start to push it over the edge, we know that at some point will fall. It will fall when its center of gravity is no longer directly over a point of its supporting base (the chair seat). Notice that that statement is itself virtually tautological. The key terms in it are all interdefinable. And yet it can also figure in explanations that appear to be causal explanations of some sort. We ask "Why doesn't that lamp tip over?" We reply "Because its center of gravity is so low." Is this a causal explanation? It can compete with explanations that are clearly causal, such as: "Because it's nailed to the table," and "Because it's supported by wires."

[4] We can manipulate centers of gravity. For instance, I change the center of gravity of a water pitcher easily, by pouring some of the water out. So, although a center of gravity is a purely abstract object, it has a spatio-temporal career, which I can affect by my actions. It has a history, but its history can include some rather strange episodes. Although it moves around in space and time, its motion can be discontinuous. For instance, if I were to take a piece of bubble gum and suddenly stick it on the pitcher's handle, that would shift the pitcher's center of gravity from point A to point B. But the center of gravity would not have to move through all the intervening positions. As an abstractum, it is not bound by all the constraints of physical travel.

[5] Consider the center of gravity of a slightly more complicated object. Suppose we wanted to keep track of the career of the center of gravity of some complex machine with lots of turning gears and camshafts and reciprocating rods -- the engine of a steam-powered unicycle, perhaps. And suppose our theory of the machine's operation permitted us to plot the complicated trajectory of the center of gravity precisely. And suppose -- most improbably -- that we discovered that in this particular machine the trajectory of the center of gravity was precisely the same as the trajectory of a particular iron atom in the crankshaft. Even if this were

discovered, we would be wrong even to *entertain* the hypothesis that the machine's center of gravity was (identical with) that iron atom. That would be a category mistake. A center of gravity is *just* an abstractum. It's just a fictional object. But when I say it's a fictional object, I do not mean to disparage it; it's a wonderful fictional object, and it has a perfectly legitimate place within serious, sober, *echt* physical science.

[6] A self is also an abstract object, a theorist's fiction. The theory is not particle physics but what we might call a branch of people-physics; it is more soberly known as a phenomenology or hermeneutics, or soul-science (*Geisteswissenschaft*). The physicist does an *interpretation*, if you like, of the chair and its behavior, and comes up with the theoretical abstraction of a center of gravity, which is then very useful in characterizing the behaviour of the chair in the future, under a wide variety of conditions. The hermeneuticist or phenomenologist -- or anthropologist -- sees some rather more complicated things moving about in the world -- human beings and animals -- and is faced with a similar problem of interpretation. It turns out to be theoretically perspicuous to organize the interpretation around a central abstraction: each person has a *self* (in addition to a center of gravity). In fact we have to posit selves for *ourselves* as well. The theoretical problem of self-interpretation is at least as difficult and important as the problem of other-interpretation.

[7] Now how does a self differ from a center of gravity? It is a much more complicated concept. I will try to elucidate it via an analogy with another sort of fictional object: fictional characters in literature. Pick up *Moby Dick* and open it up to page one. It says, "Call me Ishmael." Call whom Ishmael? Call Melville Ishmael? No. Call Ishmael Ishmael. Melville has created a fictional character named Ishmael. As you read the book you learn about Ishmael, about his life, about his beliefs and desires, his acts and attitudes. You learn a lot more about Ishmael than Melville ever explicitly tells you. Some of it you can read in by implication. Some of it you can read in by extrapolation. But beyond the limits of such extrapolation fictional worlds are simply *indeterminate*. Thus, consider the following question (borrowed from David Lewis's "Truth and Fiction," *American Philosophical Quarterly*, 1978, 15, pp.37-46). Did Sherlock Holmes have three nostrils? The answer of course is no, but not because Conan Doyle ever says that he doesn't, or that he has two, but because we're entitled to make that extrapolation. In the absence of evidence to the contrary, Sherlock Holmes' nose can be supposed to be normal. Another question: Did Sherlock Holmes have a mole on his left shoulder blade? The answer to this question is neither yes nor no. Nothing about the text or about the principles of extrapolation from the text permit an answer to that question. There is simply no fact of the matter. Why? Because Sherlock Holmes is a merely fictional character, created by, or constituted out of, the text and the culture in which that text resides.

[8] This indeterminacy is a fundamental property of fictional objects which strongly distinguishes them from another sort of object scientists talk about: theoretical entities, or what Reichenbach called *illata* -- inferred entities, such as atoms, molecules and neutrinos. A logician might say that the "principle of bivalence" does not hold for fictional objects. That is to say, with regard to any actual man, living or dead, the question of whether or not he has or had a mole on his left shoulder blade has an answer, yes or no. Did Aristotle have such a mole? There is a fact of the matter even if we can never discover it. But with regard to a fictional character, that question may have no answer at all.

[9] We can imagine someone, a benighted literary critic, perhaps, who doesn't understand that fiction is fiction. This critic has a strange theory about how fiction works. He thinks that something literally magical happens when a novelist writes a novel. When a novelist sets down words on paper, this critic says (one often hears claims like this, but not meant to be taken completely literally), the novelist actually *creates a world*. A litmus test for this bizarre view is the principle of bivalence: when our imagined critic speaks of a fictional world he means a strange sort of *real* world, a world in which the principle of bivalence holds. Such a critic might seriously wonder whether Dr Watson was *really* Moriarty's second cousin, or whether the conductor of the train that took Holmes and Watson to Aldershot was also the conductor of the train that brought them back to London. That sort of question can't properly arise if you understand fiction correctly, of course. Whereas analogous questions about historical personages have to have yes or no answers, even if we may never be able to dredge them up.

[10] Centers of gravity, as fictional objects, exhibit the same feature. They have only the properties that the theory that constitutes them endowed them with. If you scratch your head and say, "I wonder if maybe centers of gravity are really neutrinos!" you have misunderstood the theoretical status of a center of gravity.

[11] Now how can I make the claim that a self -- your own real self, for instance -- is rather like a fictional character? Aren't all *fictional* selves dependent for their very creation on the existence of *real* selves? It may seem so, but I will argue that this is an illusion. Let's go back to Ishmael. Ishmael is a fictional character, although we can certainly learn all about him. One might find him in many regards more real than many of one's friends. But, one thinks, Ishmael was created by Melville, and Melville is a real character -- was a real character. A real self. Doesn't this show that it takes a real self to create a fictional self? I think not, but if I am to convince you, I must push you through an exercise of the imagination.

[12] First of all, I want to imagine something some of you may think incredible: a novel-writing machine. We can suppose it is a product of artificial intelligence research, a computer that has been designed or programmed to write novels. But it has not been designed to write any particular novel. We can suppose (if it helps) that it has been given a great stock of whatever information it might need, and some partially random and hence unpredictable ways of starting the seed of a story going, and building upon it. Now imagine that the designers are sitting back, wondering what kind of novel their creation is going to write. They turn the thing on and after a while the high speed printer begins to go clickety-clack and out comes the first sentence. "Call me Gilbert," it says. What follows is the apparent autobiography of some fictional Gilbert. Now Gilbert is a fictional, created self but its creator is no self. Of course there were human designers who designed the machine, but they didn't design Gilbert. Gilbert is a product of a design or invention process in which there aren't any selves at all. That is, I am *stipulating* that this is not a conscious machine, not a "thinker." It is a dumb machine, but it does have the power to write a passable novel. (If you think this is strictly impossible I can only challenge you to show why you think this must be so, and invite you read on; in the end you may not have an interest in defending such a precarious impossibility-claim.)

[13] So we are to imagine that a passable story is emitted from the machine. Notice that we can perform the same sort of literary exegesis with regard to this novel as we can with any other. In fact if you were to pick up a novel at random out of a library, you could not tell with certainty that it wasn't written by something like this

machine. (And if you're a New Critic you shouldn't care.) You've got a text and you can interpret it, and so you can learn the story, the life and adventures of Gilbert. Your expectations and predictions, as you read, and your interpretive reconstruction of what you have already read, will congeal around the central node of the fictional character, Gilbert.

[14] But now I want to twiddle the knobs on this thought experiment. So far we've imagined the novel, *The Life and Times of Gilbert*, clanking out of a computer that is just a box, sitting in the corner of some lab. But now I want to change the story a little bit and suppose that the computer has arms and legs -- or better: wheels. (I don't want to make it too anthropomorphic.) It has a television eye, and it moves around in the world. It also begins its tale with "Call me Gilbert," and tells a novel, but now we notice that if we do the trick that the New Critics say you should never do, and *look outside the text*, we discover that there's a truth-preserving interpretation of that text in the real world. The adventures of Gilbert, the fictional character, now bear a striking and presumably non-coincidental relationship to the adventures of this robot rolling around in the world. If you hit the robot with a baseball bat, very shortly thereafter the story of Gilbert includes his being hit with a baseball bat by somebody who looks like you. Every now and then the robot gets locked in the closet and then says "Help me!" Help whom? Well, help Gilbert, presumably. But who is Gilbert? Is Gilbert the robot, or merely the fictional self created by the robot? If we go and help the robot out of the closet, it sends us a note: "Thank you. Love, Gilbert." At this point we will be unable to ignore the fact that the fictional career of the fictional Gilbert bears an interesting resemblance to the "career" of this mere robot moving through the world. We can still maintain that the robot's *brain*, the robot's computer, really knows nothing about the world; *it's* not a self. It's just a clanky computer. It doesn't know what it's doing. It doesn't even know that it's creating a fictional character. (The same is just as true of your brain; *it* doesn't know what it's doing either.) Nevertheless, the patterns in the behavior that is being controlled by the computer are interpretable, by us, as accreting biography -- telling the narrative of a self. But we are not the only interpreters. The robot novelist is also, of course, an interpreter: a *self*-interpreter, providing its own account of its activities in the world.

[15] I propose that we take this analogy seriously. "Where is the self?" a materialist philosopher or neuroscientist might ask. It is a category mistake to start looking around for the self in the brain. Unlike centers of gravity, whose sole property is their spatio-temporal position, selves have a spatio-temporal position that is only grossly defined. Roughly speaking, in the normal case if there are three human beings sitting on a park bench, there are three selves there, all in a row and roughly equidistant from the fountain they face. Or we might use a rather antique turn of phrase and talk about how many *souls* are located in the park. ("All twenty souls in the starboard lifeboat were saved, but those that remained on deck perished.")

[16] Brain research may permit us to make some more fine-grained localizations, but the capacity to achieve *some* fine-grained localization does not give one grounds for supposing that the process of localization can continue indefinitely and that the day will finally come when we can say, "That cell there, right in the middle of hippocampus (or wherever) -- that's the self!"

[17] There's a big difference, of course, between fictional characters and our own selves. One I would stress is that a fictional character is usually encountered as a *fait accompli*. After the novel has been written and published, you read it. At that point it

is too late for the novelist to render determinate anything indeterminate that strikes your curiosity. Dostoevsky is dead; you can't ask him what *else* Raskolnikov thought while he sat in the police station. But novels don't have to be that way. John Updike has written three novels about Rabbit Angstrom: *Rabbit Run*, *Rabbit Redux*, and *Rabbit is Rich*. Suppose that those of us who particularly liked the first novel were to get together and compose a list of questions for Updike -- things we wished Updike has talked about in that first novel, when Rabbit was a young former basketball star. We could send our questions to Updike and ask him to consider writing another novel in the series, only this time not continuing the chronological sequence. Like Lawrence Durrell's *Alexandria Quarter*, the Rabbit series could include another novel about Rabbit's early days when he was still playing basketball, and this novel could answer our questions.

[18] Notice what we would *not* be doing in such a case. We would not be saying to Updike, "Tell us the answers that you already know, the answers that are already fixed to those questions. Come on, let us know all those secrets you've been keeping from us." Nor would we be asking Updike to do research, as we might ask the author of a multi-volume biography of a real person, We would be asking him to write a new novel, to invent some more novel for us, on demand. And if he acceded, he would enlarge and make more determinate the character of Rabbit Angstrom in the process of writing the new novel. In this way matters which are indeterminate at one time can become determined later by a creative step.

[19] I propose that this imagined exercise with Updike, getting him to write more novels on demand to answer our questions, is actually a familiar exercise. That is the way we treat each other; that is the way we are. We cannot undo those parts of our pasts that are determinate, but our selves are constantly being made more determinate as we go along in response to the way the world impinges on us. Of course it is also possible for a person to engage in auto-hermeneutics, interpretation of one's self, and in particular to go back and think about one's past, and one's memories, and to rethink them and rewrite them. This process does change the "fictional" character, the character that you are, in much the way that Rabbit Angstrom, after Updike writes the second novel about him as a young man, comes to be a rather different fictional character, determinate in ways he was never determinate before. This would be an utterly mysterious and magical prospect (and hence something no one should take seriously) *if the self were anything but an abstractum*.

[20] I want to bring this out by extracting one more feature from the Updike thought experiment. Updike might take up our request but then he might prove to be forgetful. After all, it's been many years since he wrote *Rabbit Run*. He might not want to go back and reread it carefully; and when he wrote the new novel it might end up being inconsistent with the first. He might have Rabbit being in two places at one time, for instance. If we wanted to settle what the *true* story was, we'd be falling into error; there is no true story. In such a circumstance there would be simply be a failure of coherence of all the data that we had about Rabbit. And because Rabbit is a fictional character, we wouldn't smite our foreheads in wonder and declare "Oh my goodness! There's a rift in the universe; we've found a contradiction in nature!" Nothing is easier than contradiction when you're dealing with fiction; a fictional character can have contradictory properties because it's *just* a fictional character. We find such contradictions intolerable, however, when we are trying to interpret something or someone, even a fictional character, so we typically *bifurcate* the character to resolve the conflict.

[21] Something like this seems to happen to real people on rare occasions. Consider the putatively true case histories recorded in *The Three Faces of Eve* and *Sybil*. (Corbett H. Thigpen and Hervey Cleckly, *The Three Faces of Eve*, McGraw Hill, 1957, and Flora Rheta Schreiber, *Sybil*, Warner paperback, 1973.) Eve's three faces were the faces of three distinct personalities, it seems, and the woman portrayed in *Sybil* had *many* different selves, or so it seems. How can we make sense of this? Here is one way -- a solemn, skeptical way favored by some of the psychotherapists with whom I've talked about such cases: when Sybil went in to see her therapist the first time, she wasn't several different people rolled into one body. Sybil was a novel-writing machine that fell in with a very ingenious questioner, a very eager reader. And together they collaborated -- innocently -- to write many, many chapters of a new novel. And, of course, since Sybil was a sort of living novel, she went out and engaged in the world with these new selves, more or less created on demand, under the eager suggestion of a therapist.

[22] I now believe that this is overly skeptical. The population explosion of new characters that typically follows the onset of psychotherapy for sufferers of Multiple Personality Disorder (MPD) is probably to be explained along just these lines, but there is quite compelling evidence in some cases that some multiplicity of selves (two or three or four, let us say) had already begun laying down biography before the therapist came along to do the "reading". And in any event, Sybil is only a strikingly pathological case of something quite normal, a behavior pattern we can find in ourselves. We are all, at times, confabulators, telling and retelling ourselves the story of our own lives, with scant attention to the question of truth. Why, though do we behave this way? Why are we all such inveterate and inventive autobiographical novelists? As Umberto Maturana has (uncontroversially) observed: "Everything said is said by a speaker to another speaker that may be himself." But why should one talk to oneself? Why isn't that an utterly idle activity, as systematically futile as trying to pick oneself up by one's own bootstraps?

[23] A central clue comes from the sort of phenomena uncovered by Michael Gazzaniga's research on those rare individuals -- the "split-brain subjects" -- whose *corpus callosum* has been surgically severed, creating in them two largely independent cortical hemispheres that can, on occasion, be differently informed about the current scene. Does the operation *split* the self in two? After the operation, patients normally exhibit no signs of psychological splitting, appearing to be no less unified than you or I except under particularly contrived circumstances. But on Gazzaniga's view, this does not so much show that the patients have preserved their pre-surgical unity as that the unity of normal life is an illusion.

[24] According to Gazzaniga, the normal mind is *not* beautifully unified, but rather a problematically yoked-together bundle of partly autonomous systems. All parts of the mind are not equally accessible to each other at all times. These modules or systems sometimes have internal communication problems which they solve by various ingenious and devious routes. If this is true (and I think it is), it may provide us with an answer to a most puzzling question about conscious thought: what good is it? Such a question begs for a evolutionary answer, but it will have to be speculative, of course. (It is not critical to my speculative answer, for the moment, where genetic evolution and transmission breaks off and cultural evolution and transmission takes over.)

[25] In the beginning -- according to Julian Jaynes (*The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston: Houghton Mifflin, 1976), whose account I am adapting -- were speakers, our ancestors, who weren't really conscious. They spoke, but they just sort of blurted things out, more or less the way bees do bee dances, or the way computers talk to each other. That is not conscious communication, surely. When these ancestors had problems, sometimes they would "ask" for help (more or less like Gilbert saying "Help me!" when he was locked in the closet), and sometimes there would be somebody around to hear them. So they got into the habit of asking for assistance and, particularly, asking questions. Whenever they couldn't figure out how to solve some problem, they would ask a question, addressed to no one in particular, and sometimes whoever was standing around could answer them. And they also came to be designed to be provoked on many such occasions into answering questions like that -- to the best of their ability -- when asked.

[26] Then one day one of our ancestors asked a question in what was apparently an inappropriate circumstance: there was nobody around to be the audience. Strangely enough, he heard his own question, and this stimulated him, cooperatively, to think of an answer, and sure enough the answer came to him. He had established, without realizing what he had done, a communication link between two parts of his brain, between which there was, for some deep biological reason, an accessibility problem. One component of the mind had confronted a problem that another component could solve; if only the problem could be posed for the latter component! Thanks to his habit of asking questions, our ancestor stumbled upon a route via the ears. What a discovery! Sometimes talking and listening to yourself can have wonderful effects, not otherwise obtainable. All that is needed to make sense of this idea is the hypothesis that the modules of the mind have different capacities and ways of doing things, and are not perfectly interaccessible. Under such circumstances it could be true that the way to get yourself to figure out a problem is to tickle your ear with it, to get that part of your brain which is best stimulated by *hearing* a question to work on the problem. Then sometimes you will find yourself with the answer you seek on the tip of your tongue.

[27] This would be enough to establish the evolutionary endorsement (which might well be only culturally transmitted) of the behavior of *talking to yourself*. But as many writers have observed, conscious thinking seems -- much of it -- to be a variety of particularly efficient and private talking to oneself. The evolutionary transition to thought is then easy to conjure up. All we have to suppose is that the route, the circuit that at first went via mouth and ear, got shorter. People "realized" that the actual vocalization and audition was a rather inefficient part of the loop. Besides, if there were other people around who might overhear it, you might give away more information than you wanted. So what developed was a habit of subvocalization, and this in turn could be streamlined into conscious, verbal thought.

[28] In his posthumous book *On Thinking* (ed. Konstantin Kolenda, Totowa New Jersey, Rowman and Littlefield, 1979), Gilbert Ryle asks: "What is *Le Penseur* doing?" For behaviorists like Ryle this is a real problem. One bit of chin-on-fist-with-knitted-brow looks pretty much like another bit, and yet some of it seems to arrive at good answers and some of it doesn't. What can be going on here? Ironically, Ryle, the arch-behaviorist, came up with some very sly suggestions about what might be going on. Conscious thought, Ryle claimed, should be understood on the model of self-teaching, or better, perhaps: self-schooling or training. Ryle had little to say about how this self-schooling might actually work, but we can get some initial

understanding of it on the supposition that we are *not* the captains of our ships; there is no conscious self that is unproblematically in command of the mind's resources. Rather, we are somewhat disunified. Our component modules have to act in opportunistic but amazingly resourceful ways to produce a modicum of behavioral unity, *which is then enhanced by an illusion of greater unity.*

[29] What Gazzaniga's research reveals, sometimes in vivid detail, is how this must go on. Consider some of his evidence for the extraordinary resourcefulness exhibited by (something in) the right hemisphere when it is faced with a communication problem. In one group of experiments, split-brain subjects must reach into a closed bag with the left hand to feel an object, which they are then to identify verbally. The sensory nerves in the left hand lead to the right hemisphere, whereas the control of speech is normally in the left hemisphere, but for most of us, this poses no problem. In a normal person, the left hand can know what the right hand is doing thanks to the corpus colosum, which keeps both hemispheres mutually informed. But in a split-brain subject, this unifying link has been removed; the right hemisphere gets the information about the touched object from the left hand, but the left, language-controlling, hemisphere must make the identification public. So the "part which can speak" is kept in the dark, while the "part which knows" cannot make public its knowledge.

[30] There is a devious solution to this problem, however, and split-brain patients have been observed to discover it. Whereas ordinary tactile sensations are represented contralaterally -- the signals go to the opposite hemisphere -- pain signals are also represented ipsilaterally. That is, thanks to the way the nervous system is wired up, pain stimuli go to both hemispheres. Suppose the object in the bag is a pencil. The right hemisphere will sometimes hit upon a very clever tactic: hold the pencil in your left hand so its point is pressed hard into your palm; this creates pain, and lets the left hemisphere know there's something sharp in the bag, which is enough of a hint so that it can begin guessing; the right hemisphere will signal "getting warmer" and "got it" by smiling or other controllable sings, and in a very short time "the subject" -- the *apparently* unified "sole inhabitant" of the body - will be able to announce the correct answer.

[31] Now either the split-brain subjects have developed this extraordinarily devious talent as a reaction to the operation that landed them with such radical accessibility problem, or the operation *reveals* -- but does not create -- a virtuoso talent to be found also in normal people. Surely, Gazzaniga claims, the latter hypothesis is the most likely one to investigate. That is, it does seem that we are all virtuoso novelists, who find ourselves engaged in all sorts of behavior, more or less unified, but sometimes disunified, and we always put the best "faces" on it we can. We try to make all of our material cohere into a single good story. And that story is our autobiography.

[32] The chief fictional character at the center of that autobiography is one's *self*. And if you still want to know what the self *really* is, you're making a category mistake. After all, when a human being's behavioral control system becomes seriously impaired, it can turn out that the best hermeneutical story we can tell about that individual says that there is more than one character "inhabiting" that body. This is quite possible on the view of the self that I have been presenting; it does not require any fancy metaphysical miracles. One can discover multiple selves in a person just as unproblematically as one could find Early Young Rabbit and Late Young Rabbit in the imagined Updike novels: all that has to be the case is that the

story doesn't cohere around one self, one imaginary point, but coheres (coheres much better, in any case) around two different imaginary points.

[33] We sometimes encounter psychological disorders, or surgically created disunities, where the only way to interpret or make sense of them is to posit in effect two centers of gravity, two selves. One isn't creating or discovering a little bit of ghost stuff in doing that. One is simply creating another abstraction. It is an abstraction one uses as part of a theoretical apparatus to understand, and predict, and make sense of, the behavior of some very complicated things. The fact that these abstract selves seem so robust and real is not surprising. They are much more complicated theoretical entities than a center of gravity. And remember that even a center of gravity has a fairly robust presence, once we start playing around with it. But no one has ever seen or ever will see a center of gravity. As David Hume noted, no one has ever seen a self, either:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception.... If anyone, upon serious and unprejudiced reflection, thinks he has a different notion of *himself*, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls *himself*; though I am certain there is no such principle in me. (*Treatise on Human Nature*, I, IV, sec. 6.)

THE SELF AND THE FUTURE

Bernard Williams

[1] Suppose that there were some process to which two persons, A and B, could be subjected as a result of which they might be said -- question-beggingly -- to have *exchanged bodies*. That is to say -- less question-beggingly -- there is a certain human body which is such that when previously we were confronted with it, we were confronted with person A, certain utterances coming from it were expressive of memories of the past experiences of A, certain movements of it partly constituted the actions of A and were taken as expressive of the character of A, and so forth; but now, after the process is completed, utterances coming from this body are expressive of what seem to be just those memories which previously we identified as memories of the past experiences of B, its movements partly constitute actions expressive of the character of B, and so forth; and conversely with the other body.

[2] There are certain important philosophical limitations on how such imaginary cases are to be constructed, and how they are to be taken when constructed in various ways. I shall mention two principal limitations, not in order to pursue them further here, but precisely in order to get them out of the way.

[3] There are certain limitations, particularly with regard to character and mannerisms, to our ability to imagine such cases even in the most restricted sense of our being disposed to take the later performances of that body which was previously A's as expressive of B's character; if the previous A and B were extremely unlike one another both physically and psychologically, and if, say, in addition, they were of different sex, there might be grave difficulties in reading B's dispositions in any possible performances of A's body. Let us forget this, and for the present purpose just take A and B as being sufficiently alike (however alike that has to be) for the difficulty not to arise; after the experiment, persons familiar with A and B are just *overwhelmingly struck* by the B-ish character of the doings associated with what was previously A's body, and conversely. Thus the feat of imagining an exchange of bodies is supposed possible in the most restricted sense. But now there is a further limitation which has to be overcome if the feat is to be not merely possible in the most restricted sense but also is to have an outcome which, on serious reflection, we are prepared to describe as A and B having changed bodies -- that is, an outcome where, confronted with what was previously A's body, we are prepared seriously to say that we are now confronted with B.

[4] It would seem a necessary condition of so doing that the utterances coming from that body be taken as genuinely expressive of memories of B's past. But memory is a causal notion; and as we actually use it, it seems a necessary condition on x's present knowledge of x's earlier experiences constituting memory of those experiences that the causal chain linking the experiences and the knowledge should not run outside x's body. Hence if utterances coming from a given body are to be taken as expressive of memories of the experiences of B, there should be some suitable causal link between the appropriate state of that body and the original happening of those experiences to B. One radical way of securing that condition in the imagined exchange case is to suppose, with Shoemaker, that the brains of A and of B are transposed. We may not need so radical a condition. Thus suppose it were possible to extract information from a man's brain and store it in a device while his brain was repaired, or even renewed, the information then being replaced: it would seem exaggerated to insist that the resultant man could not possibly have the memories he had before the operation. With regard to our knowledge of our own past, we draw distinctions

between merely recalling, being reminded, and learning again, and those distinctions correspond (roughly) to distinctions between no new input, partial new input, and total new input with regard to the information in question; and it seems clear that the information-parking case just imagined would not count as new input in the sense necessary and sufficient for "learning again." Hence we can imagine the case we are concerned with in terms of information extracted into such devices from *A*'s and *B*'s brains and replaced in the other brain; this is the sort of model which, I think not unfairly for the present argument, I shall have in mind.

[5] We imagine the following. The process considered above exists; two persons can enter some machine, let us say, and emerge changed in the appropriate ways. If *A* and *B* are the persons who enter, let us call the persons who emerge the *A-body-person* and the *B-body-person*: the *A-body-person* is that person (whoever it is) with whom I am confronted when, after the experiment, I am confronted with that body which previously was *A*'s body -- that is to say, that person who would naturally be taken for *A* by someone who just saw this person, was familiar with *A*'s appearance before the experiment, and did not know about the happening of the experiment. A non-question-begging description of the experiment will leave it open which (if either) of the persons *A* and *B* the *A-body-person* is; the description of the experiment as "persons changing bodies" of course implies that the *A-body-person* is actually *B*.

[6] We take two persons *A* and *B* who are going to have the process carried out on them. (We can suppose, rather hazily, that they are willing for this to happen; to investigate at all closely at this stage why they might be willing or unwilling, what they would fear, and so forth, would anticipate some later issues.) We further announce that one of the two resultant persons, the *A-body-person* and the *B-body-person*, is going after the experiment to be given \$100,00, while the other is going to be tortured. We then ask each *A* and *B* to choose which treatment should be dealt out to which of the persons who will emerge from the experiment, the choice to be made (if it can be) on selfish grounds.

[7] Suppose that *A* chooses that the *B-body-person* should get the pleasant treatment and the *A-body-person* the unpleasant treatment; and *B* chooses conversely (this might indicate that they thought that "changing bodies" was indeed a good description of the outcome). The experimenter cannot act in accordance with both these sets of preferences, those expressed by *A* and those expressed by *B*. Hence there is one clear sense in which *A* and *B* cannot both get what they want: namely, that if the experimenter, before the experiment, announces to *A* and *B* that he intends to carry out the alternative (for example), of treating the *B-body-person* unpleasantly and the *A-body-person* pleasantly -- then *A* can say rightly, "That's not the outcome I chose to happen," and *B* can say rightly, "That's just the outcome I chose to happen." So, evidently, *A* and *B* before the experiment can each come to know either that the outcome he chose will be that which will happen, or that the one he chose will not happen, and in that sense they can get or fail to get what they wanted. But is it also true that when the experimenter proceeds after the experiment to act in accordance with one of the preferences and not the other, then one of *A* and *B* will have got what he wanted, and the other not?

[8] There seems very good ground for saying so. For suppose the experimenter, having elicited *A*'s and *B*'s preference, says nothing to *A* and *B* about what he will do; conducts the experiment; and then, for example, gives the unpleasant treatment to the *B-body-person* and the pleasant treatment to the *A-body-person*. Then the *B-body-person* will not only complain of the unpleasant treatment as such, but will complain (since he has *A*'s memories) that that was not the outcome he chose, since he chose that the *B-body-person* should be well treated; and since *A* made his choice in selfish spirit, he may add that he

precisely chose in that way because he did not want the unpleasant things to happen to him. The A-body-person meanwhile will express satisfaction both at the receipt of the \$100,000, and also at the fact that the experimenter has chosen to act in the way that he, B, so wisely chose. These facts make a strong case for saying that the experimenter has brought it about that B did in the outcome get what he wanted and A did not. It is therefore a strong case for saying that the B-body-person really is A, and the A-body-person really is B; and therefore for saying that the process of the experiment really is that of changing bodies. For the same reasons it would seem that A and B in our example really did choose wisely, and that it was A's bad luck that the choice he correctly made was not carried out, B's good luck that the choice he correctly made was carried out. This seems to show that to care about what happens to me in the future is not necessarily to care about what happens to this body (the one I now have); and this in turn might be taken to show that in some sense of Descartes's obscure phrase, I and my body are "really distinct" (though, of course, nothing in these considerations could support the idea that I could exist without a body at all).

[9] These suggestions seem to be reinforced if we consider the cases where A and B make other choices with regard to the experiment. Suppose that A chooses that the A-body-person should get the money, and the B-body-person get the pain, and B chooses conversely. Here again there can be no outcome which matches the expressed preferences of both of them: they cannot both get what they want. The experimenter announces, before the experiment, that the A-body-person will in fact get the money, and the B-body-person will get the pain. So A at this stage gets what he wants (the announced outcome matches his expressed preference). After the experiment, the distribution is carried out as announced. Both the A-body-person and the B-body-person will have to agree that what is happening is in accordance with the preference that A originally expressed. The B-body-person will naturally express this acknowledgment (since he has A's memories) by saying that this is the distribution he chose; he will recall, among other things, the experimenter announcing this outcome, his approving it as what he chose, and so forth. However, he (the B-body-person) certainly does not like what is now happening to him, and would much prefer to be receiving what the A-body-person is receiving—namely, \$100,000. The A-body-person will on the other hand recall choosing an outcome other than this one, but will reckon it good luck that the experimenter did not do what he recalls choosing. It looks, then, as though the A-body-person had gotten what he wanted, but not what he chose, while the B-body-person has gotten what he chose, but not what he wanted. So once more it looks as though they are, respectively, B and A; and that in this case the original choices of both A and B were unwise.

[10] Suppose, lastly, that in the original choice A takes the line of the first case and B of the second: that is, A chooses that the B-body-person should get the money and the A-body-person the pain, and B chooses exactly the same thing. In this case, the experimenter would seem to be in the happy situation of giving both persons what they want -- or at least, like God, what they have chosen. In this case, the B-body-person likes what he is receiving, recalls choosing it, and congratulates himself on the wisdom of (as he puts it) his choice; while the A-body-person does not like what he is receiving, recalls choosing it, and is forced to acknowledge that (as he puts it) his choice was unwise. So once more we seem to get results to support the suggestions drawn from the first case.

[11] Let us now consider the question, not of A and B choosing certain outcomes to take place after the experiment, but of their willingness to engage in the experiment at all. If they were initially inclined to accept the description of the experiment as "changing bodies" then one thing that would interest them would be the character of the other person's body. In this respect also what would happen after the experiment would seem to suggest that

"changing bodies" was a good description of the experiment. If A and B agreed to the experiment, being each not displeased with the appearance, physique, and so forth of the other person's body; after the experiment the B-body-person might well be found saying such things as: "When I agreed to this experiment, I thought that B's face was quite attractive, but now I look at it in the mirror, I am not so sure"; or the A-body-person might say "When I agreed to this experiment I did not know that A had a wooden leg; but now, after it is over, I find that I have this wooden leg, and I want the experiment reversed." It is possible that he might say further that he finds the leg very uncomfortable, and that the B-body-person should say, for instance, that he recalls that he found it very uncomfortable at first, but one gets used to it: but perhaps one would need to know more than at least I do about the physiology of habituation to artificial limbs to know whether the A-body-person would find the leg uncomfortable: that body, after all, has had the leg on it for some time. But apart from this sort of detail, the general line of the outcome regarded from this point of view seems to confirm our previous conclusions about the experiment.

[12] Now let us suppose that when the experiment is proposed (in non-question-begging terms) A and B think rather of their psychological advantages and disadvantages. A's thoughts turn primarily to certain sorts of anxiety to which he is very prone, while B is concerned with the frightful memories he has of past experiences which still distress him. They each hope that the experiment will in some way result in their being able to get away from these things. They may even have been impressed by philosophical arguments to the effect that bodily continuity is at least a necessary condition of personal identity: A, for example, reasons that, granted the experiment comes off, then the person who is bodily continuous with him will not have this anxiety, while the other person will no doubt have some anxiety -- perhaps in some sense his anxiety -- and at least that person will not be he. The experiment is performed and the experimenter (to whom A and B previously revealed privately their several difficulties and hopes) asks the A-body-person whether he has gotten rid of his anxiety. This person presumably replies that he does not know what the man is talking about; he never had such anxiety, but he did have some very disagreeable memories, and recalls engaging in the experiment to get rid of them, and is disappointed to discover that he still has them. The B-body-person will react in a similar way to questions about his painful memories, pointing out that he still has his anxiety. These results seem to confirm still further the description of the experiment as "changing bodies." And all the results suggest that the only rational thing to do, confronted with such an experiment, would be to identify oneself with one's memories, and so forth, and not with one's body. The philosophical arguments designed to show that bodily continuity was at least a necessary condition of personal identity would seem to be just mistaken.

[13] Let us now consider something apparently different. Someone in whose power I am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to tomorrow in great apprehension. He adds that when the time comes, I shall not remember being told that this was going to happen to me, since shortly before the torture something else will be done to me which will make me forget the announcement. This certainly will not cheer me up, since I know perfectly well that I can forget things, and that there is such a thing as indeed being tortured unexpectedly because I had forgotten or been made to forget a prediction of the torture: that will still be a torture which, so long as I do know about the prediction, I look forward to in fear. He then adds that my forgetting the announcement will be only part of a larger process: when the moment of torture comes, I shall not remember any of the things I am now in a position to remember. This does not cheer me up, either, since I can readily conceive of being involved in an accident, for instance, as a result of which I wake up in a completely amnesiac state and also in great pain; that could certainly happen to me, I should not like it to happen to me, nor to know that it was going to happen to me. He now further adds that at the moment of torture I shall not only not remember the

things I am now in a position to remember, but will have a different set of impressions of my past, quite different from the memories I now have. I do not think that this would cheer me up, either. For I can at least conceive the possibility, if not the concrete reality, of going completely mad, and thinking perhaps that I am George IV or somebody; and being told that something like that was going to happen to me would have no tendency to reduce the terror of being told authoritatively that I was going to be tortured, but would merely compound the horror. Nor do I see why I should be put into any better frame of mind by the person in charge adding lastly that the impressions of my past with which I shall be equipped on the eve of torture will exactly fit the past of another person now living, and that indeed I shall acquire these impressions by (for instance) information now in his brain being copied into mine. Fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen -- torture, which one can indeed expect to happen to oneself, and to be preceded by certain mental derangements as well.

[14] If this is right, the whole question seems now to be totally mysterious. For what we have just been through is of course merely one side, differently represented, of the transaction which we considered before; and it represents it as a perfectly hateful prospect, while the previous considerations represented it as something one should rationally, perhaps even cheerfully, choose out of the options there presented. It is differently presented, of course, and in two notable respects; but when we look at these two differences of presentation, can we really convince ourselves that the second presentation is wrong or misleading, thus leaving the road open to the first version which at the time seemed so convincing? Surely not.

[15] The first difference is that in the second version the torture is throughout represented as going to happen to me: "you," the man in charge persistently says. Thus he is not very neutral. But should he have been neutral? Or, to put it another way, does his use of the second person have a merely emotional and rhetorical effect on me, making me afraid when further reflection would have shown that I had no reason to be? It is certainly not obviously so. The problem just is that through every step of his predictions I seem to be able to follow him successfully. And if I reflect on whether what he has said gives me grounds for fearing that I shall be tortured, I could consider that behind my fears lies some principle such as this: that my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time, with the platitudinous exception of those psychological states which in themselves exclude experiencing pain, notably (if it is a psychological state) unconsciousness. In particular, what impressions I have about the past will not have any effect on whether I undergo the pain or not. This principle seems sound enough.

[16] It is an important fact that not everything I would, as things are, regard as an evil would be something that I should rationally fear as an evil if it were predicted that it would happen to me in the future and also predicted that I should undergo significant psychological changes in the meantime. For the fact that I regard that happening, things being as they are, as an evil can be dependent on factors of belief or character which might themselves be modified by the psychological changes in question. Thus if I am appallingly subject to acrophobia, and am told that I shall find myself on top of a steep mountain in the near future, I shall to that extent be afraid; but if I am told that I shall be psychologically changed in the meantime in such a way as to rid me of my acrophobia (and as with the other prediction, I believe it), then I have no reason to be afraid of the predicted happening, or at least not the same reason. Again, I might look forward to meeting a certain person again with either alarm or excitement because of my memories of our past relations. In some part, these memories operate in connection with my emotion, not only on the present time, but projectively forward: for it is to a meeting itself affected by the presence of those

memories that I look forward. If I am convinced that when the time comes I shall not have those memories, then I shall not have just the same reasons as before for looking forward to that meeting with the one emotion or the other. (Spiritualism, incidentally, appears to involve the belief that I have just the same reasons for a given attitude toward encountering people again after I am dead, as I did before: with the one modification that I can be sure it will all be very nice.)

[17] Physical pain, however, the example which for simplicity (and not for any obsessional reason) I have taken, is absolutely minimally dependent on character or belief. No amount of change in my character or my beliefs would seem to affect substantially the nastiness of tortures applied to me; correspondingly, no degree of predicted change in my character and beliefs can unseat the fear of torture which, together with those changes, is predicted for me.

[18] I am not at all suggesting that the only basis, or indeed the only rational basis, for fear in the face of these various predictions is how things will be relative to my psychological state in the eventual outcome. I am merely pointing out that this is one component; it is not the only one. For certainly one will fear and otherwise reject the changes themselves, or in very many cases one would. Thus one of the old paradoxes of hedonistic utilitarianism; if one had assurances that undergoing certain operations and being attached to a machine would provide one for the rest of one's existence with an unending sequence of delicious and varied experiences, one might very well reject the option, and react with fear if someone proposed to apply it compulsorily; and that fear and horror would seem appropriate reactions in the second case may help to discredit the interpretation (if anyone has the nerve to propose it) that one's reason for rejecting the option voluntarily would be a consciousness of duties to others which one in one's hedonic state would leave undone. The prospect of contented madness or vegetableness is found by many (not perhaps by all) appalling in ways which are obviously not a function of how things would then be for them, for things would then be for them not appalling. In the case we are at present discussing, these sorts of considerations seem merely to make it clearer that the predictions of the man in charge provide a double ground of horror: at the prospect of torture, and at the prospect of the change in character and in impressions of the past that will precede it. And certainly, to repeat what has already been said, the prospect of the second certainly seems to provide no ground for rejecting or not fearing the prospect of the first.

[19] I said that there were two notable differences between the second presentation of our situation and the first. The first difference, which we have just said something about, was that the man predicted the torture for me, a psychologically very changed "me." We have yet to find a reason for saying that he should not have done this, or that I really should be unable to follow him if he does; I seem to be able to follow him only too well. The second difference is that in this presentation he does not mention the other man, except in the somewhat incidental role of being the provenance of the impressions of the past I end up with. He does not mention him at all as someone who will end up with impressions of the past derived from me (and, incidentally, with \$100,000 as well -- a consideration which, in the frame of mind appropriate to this version, will merely make me jealous).

[20] But why *should* he mention this man and what is going to happen to him? My selfish concern is to be told what is going to happen to me, and now I know: torture, preceded by changes of character, brain operations, changes in impressions of the past. The knowledge that one other person, or none, or many will be similarly mistreated may affect me in other ways, of sympathy, greater horror at the power of this tyrant, and so forth; but surely it cannot affect my expectations of torture? But -- someone will say -- this is to leave out exactly the feature which, as the first presentation of the case showed, makes all the

difference: for it is to leave out the person who, as the first presentation showed, will be you. It is to leave out not merely a feature which should fundamentally affect your fears, it is to leave out the very person for whom you are fearful. So of course, the objector will say, this makes all the difference.

[21] But can it? Consider the following series of cases. In each case we are to suppose that after what is described, A is, as before, to be tortured; we are also to suppose the person A is informed beforehand that just these things followed by the torture will happen to him:

(i) A is subjected to an operation which produces total amnesia;

(ii) amnesia is produced in A, and other interference leads to certain changes in his character;

(iii) changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;

(iv) the same as (iii), except that both the character traits and the "memory" impressions are designed to be appropriate to another actual person, B;

(v) the same as (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;

(vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction.

[22] I take it that no one is going to dispute that A has reasons, and fairly straightforward reasons, for fear of pain when the prospect is that of situation (i); there seems no conceivable reason why this should not extend to situation (ii), and the situation (iii) can surely introduce no difference of principle -- it just seems a situation which for more than one reason we should have grounds for fearing, as suggested above. Situation (iv) at least introduces the person B, who was the focus of the objection we are now discussing. But it does not seem to introduce him in any way which makes a material difference; if I can expect pain through a transformation which involves new "memory" -- impressions, it would seem a purely external fact, relative to that, that the "memory"-impressions had a model. Nor, in (iv), do we satisfy a causal condition which I mentioned at the beginning for the "memories" actually being memories; though notice that if the job were done thoroughly, I might well be able to elicit from the A-body-person the kinds of remarks about his previous expectations of the experiment -- remarks appropriate to the original B -- which so impressed us in the first version of the story. I shall have a similar assurance of this being so in situation (v), where, moreover, a plausible application of the causal condition is available.

[23] But two things are to be noticed about this situation. First, if we concentrate on A and the A-body-person, we do not seem to have added anything which from the point of view of his fears makes any material difference; just as, in the move from (iii) to (iv), it made no relevant difference that the new "memory"-impressions which precede the pain had, as it happened, a model, so in the move from (iv) to (v) all we have added is that they have a model which is also their cause: and it is still difficult to see why that, to him looking forward, could possibly make the difference between expecting pain and not expecting pain. To illustrate that point from the case of character: if A is capable of expecting pain, he is capable of expecting pain preceded by a change in his dispositions -- and to that

expectation it can make no difference, whether that change in his dispositions is modeled on, or indeed indirectly caused by, the dispositions of some other person. If his fears can, as it were, reach through the change, it seems a mere trimming how the change is in fact induced. The second point about situation (v) is that if the crucial question for A's fears with regard to what befalls the A-body-person is whether the A-body-person is or is not the person B, then that condition has not yet been satisfied in situation (v): for there we have an undisputed B in addition to the A-body-person, and certainly those two are not the same person.

[24] But in situation (vi), we seemed to think, that is finally what he is. But if A's original fears could reach through the expected changes in (v), as they did in (iv) and (iii), then certainly they can reach through in (vi). Indeed, from the point of view of A's expectations and fears, there is less difference between (vi) and (v) than there is between (v) and (iv) or between (iv) and (iii). In those transitions, there were at least differences -- though we could not see that they were really relevant differences -- in the content and cause of what happened to him; in the present case there is absolutely no difference at all in what happens to him, the only difference being in what happens to someone else. If he can fear pain when (v) is predicted, why should he cease to when (vi) is?

[25] I can see only one way of relevantly laying great weight on the transition from (v) to (vi); and this involves a considerable difficulty. This is to deny that, as I put it, the transition from (v) to (vi) involves merely the addition of something happening to *somebody else*; what rather it does, it will be said, is to involve the reintroduction of A himself, as the B-body-person; since he has reappeared in this form, it is for this person, and not for the unfortunate A-body-person, that A will have his expectations. This is to reassert, in effect, the viewpoint emphasized in our first presentation of the experiment. But this surely has the consequence that A should not have fears for the A-body-person who appeared in situation (v). For by the present argument, the A-body-person in (vi) is not A; the B-body-person is. But the A-body-person in (v) is, in character, history, everything, exactly the same as the A-body-person in (vi); so if the latter is not A, then neither is the former. (It is this point, no doubt, that encourages one to speak of the difference that goes with [vi] as being, on the present view, the *reintroduction* of A.) But no one else in (v) has any better claim to be A. So in (v), it seems, A just does not exist. This would certainly explain why A should have no fears for the state of things in (v) -- though he might well have fears for the path to it. But it rather looked earlier as though he could well have fears for the state of things in (v). Let us grant, however, that that was an illusion, and that A really does not exist in (v); then does he exist in (iv), (iii), (ii), or (i)? It seems very difficult to deny it for (i) and (ii); are we perhaps to draw the line between (iii) and (iv)?

[26] Here someone will say: you must not insist on drawing a line -- borderline cases are borderline cases, and you must not push our concepts beyond their limits. But this well-known piece of advice, sensible as it is in many cases, seems in the present case to involve an extraordinary difficulty. It may intellectually comfort observers of A's situation; but what is A supposed to make of it? To be told that a future situation is a borderline one for its being myself that is hurt, that it is conceptually undecidable whether it will be me or not, is something which, it seems, I can do nothing with; because, in particular, it seems to have no comprehensible representation in my expectations and the emotions that go with them.

[27] If I expect that a certain situation, S, will come about in the future, there is of course a wide range of emotions and concerns, directed on S, which I may experience now in relation to my expectation. Unless I am exceptionally egoistic, it is not a condition on my being concerned in relation to this expectation, that I myself will be involved in S -- where my

being "involved" in S means that I figure in S as someone doing something at that time or having something done to me, or, again, that S will have consequences affecting me at that or some subsequent time. There are some emotions, however, which I will feel only if I will be involved in S, and fear is an obvious example.

[28] Now the description of S under which it figures in my expectations will necessarily be, in various ways, indeterminate; and one way in which it may be indeterminate is that it leave open whether I shall be involved in S or not. Thus I may have good reason to expect that one out of us five is going to get hurt, but no reason to expect it to be me rather than one of the others. My present emotions will be correspondingly affected by this indeterminacy. Thus, sticking to the egoistic concern involved in fear, I shall presumably be somewhat more cheerful than if I knew it was going to be me, somewhat less cheerful than if I had been left out altogether. Fear will be mixed with, and qualified by, apprehension; and so forth. These emotions revolve around the thought of the eventual determination of the indeterminacy; moments of straight fear focus on its really turning out to be me, of hope on its turning out not to be me. All the emotions are related to the coming about of what I expect: and what I expect in such a case just cannot come about save by coming about in one of the ways or another.

[29] There are other ways in which indeterminate expectations can be related to fear. Thus I may expect (perhaps neurotically) that something nasty is going to happen to me, indeed expect that when it happens it will take some determinate form, but have no range, or no closed range, of candidates for the determinate form to rehearse in my present thought. Different from this would be the fear of something radically indeterminate -- the fear (one might say) of a nameless horror. If somebody had such a fear, one could even say that he had, in a sense, a perfectly determinate expectation: if what he expects indeed comes about, there will be nothing more determinate to be said about it after the event than was said in the expectation. Both these cases of course are cases of *fear* because one thing that is fixed amid the indeterminacy is the belief that it is to me to which the things will happen.

[30] Central to the expectation of S is the thought of what it will be like when it happens -- thought which may be indeterminate, range over alternatives, and so forth. When S involves me, there can be the possibility of a special form of such thought: the thought of how it will be for me, the imaginative projection of myself as participant in S.

[31] I do not have to think about S in this way, when it involves me; but I may be able to. (It might be suggested that this possibility was even mirrored in the language, in the distinction between "expecting to be hurt" and "expecting that I shall be hurt"; but I am very doubtful about this point, which is in any case of no importance.)

[32] Suppose now that there is an S with regard to which it is for conceptual reasons undecidable whether it involves me or not, as is proposed for the experimental situation by the line we are discussing. It is important that the expectation of S is not *indeterminate* in any of the ways we have just been considering. It is not like the nameless horror, since the fixed point of that case was that it was going to happen to the subject, and that made his state unequivocally fear. Nor is it like the expectation of the man who expects one of the five to be hurt; his fear was indeed equivocal, but its focus, and that of the expectation, was that when S came about, it would certainly come about in one way or the other. In the present case, fear (of the torture, that is to say, not of the initial experiment) seems neither appropriate, nor inappropriate, nor appropriately equivocal. Relatedly, the subject has an incurable difficulty about how he may think about S. If he engages in projective imaginative thinking (about how it will be for him), he implicitly answers the necessarily unanswerable question; if he thinks that he cannot engage in such thinking, it looks very much as if he

also answers it, though in the opposite direction. Perhaps he must just refrain from such thinking; but is he just refraining from it, if it is incurably undecidable whether he can or cannot engage in it?

[33] It may be said that all that these considerations can show is that fear, at any rate, does not get its proper footing in this case; but that there could be some other, more ambivalent, form of concern which would indeed be appropriate to this particular expectation, the expectation of the conceptually undecidable situation. There are, perhaps, analogous feelings that actually occur in actual situations. Thus material objects do occasionally undergo puzzling transformations which leave a conceptual shadow over their identity. Suppose I were sentimentally attached to an object to which this sort of thing then happened; then it might be that I could neither feel about it quite as I did originally, nor be totally indifferent to it, but would have some other and rather ambivalent feeling toward it. Similarly, it may be said, toward the prospective sufferer of pain, my identity relations with whom are conceptually shadowed, I can feel neither as I would if he were certainly me, nor as I would if he were certainly not, but rather some such ambivalent concern.

[34] But this analogy does little to remove the most baffling aspect of the present case -- an aspect which has already turned up in what was said about the subject's difficulty in thinking either projectively or non-projectively about the situation. For to regard the prospective pain-sufferer just like the transmogrified object of sentiment, and to conceive of my ambivalent distress about his future pain as just like ambivalent distress about some future damage to such an object, is of course to leave him and me clearly distinct from one another, and thus to displace the conceptual shadow from its proper place. I have to get nearer to him than that. But is there any nearer that I can get to him without expecting his pain? If there is, the analogy has not shown us it. We can certainly not get nearer by expecting, as it were, *ambivalent* pain; there is no place at all for that. There seems to be an obstinate bafflement to mirroring in my expectations a situation in which it is conceptually undecidable whether I occur.

[35] The bafflement seems, moreover, to turn to plain absurdity if we move from conceptual undecidability to its close friend and neighbor, conventionalist decision. This comes out if we consider another description, overtly conventionalist, of the series of cases which occasioned the present discussion. This description would reject a point I relied on in an earlier argument -- namely, that if we deny that the A-body-person in (vi) is A (because the B-body-person is), then we must deny that the A-body-person in (v) is A, since they are exactly the same. "No," it may be said, "this is just to assume that we say the same in different sorts of situation. No doubt when we have the very good candidate for being A -- namely, the B-body-person-we call him A; but this does not mean that we should not call the A-body-person A in that other situation when we have no better candidate around. Different situations call for different descriptions." This line of talk is the sort of thing indeed appropriate to lawyers deciding the ownership of some property which has undergone some bewildering set of transformations; they just have to decide, and in each situation, let us suppose, it has got to go to somebody, on as reasonable grounds as the facts and the law admit. But as a line to deal with a person's fears or expectations about his own future, it seems to have no sense at all. If A's fears can extend to what will happen to the A-body-person in (v), I do not see how they can be rationally diverted from the fate of the exactly similar person in (vi) by his being told that someone would have a reason in the latter situation which he would not have in the former for deciding to call another person A.

[36] Thus, to sum up, it looks as though there are two presentations of the imagined experiment and the choice associated with it, each of which carries conviction, and which lead to contrary conclusions. The idea, moreover, that the situation after the experiment is

conceptually undecidable in the relevant respect seems not to assist, but rather to increase, the puzzlement; while the idea (so often appealed to in these matters) that it is conventionally decidable is even worse. Following from all that, I am not in the least clear which option it would be wise to take if one were presented with them before the experiment. I find that rather disturbing.

[37] Whatever the puzzlement, there is one feature of the arguments which have led to it which is worth picking out, since it runs counter to something which is, I think, often rather vaguely supposed. It is often recognized that there are "first-personal" and "third-personal" aspects of questions about persons, and that there are difficulties about the relations between them. It is also recognized that "mentalist" considerations (as we may vaguely call them) and considerations of bodily continuity are involved in questions of personal identity (which is not to say that there are mentalistic and bodily criteria of personal identity). It is tempting to think that the two distinctions run in parallel: roughly, that a first-personal approach concentrates attention on mentalistic considerations, while a third-personal approach emphasizes considerations of bodily continuity. The present discussion is an illustration of exactly the opposite. The first argument, which led to the "mentalist" conclusion that A and B would change bodies and that each person should identify himself with the destination of his memories and character, was an argument entirely conducted in third-personal terms. The second argument, which suggested the bodily continuity identification, concerned itself with the first-personal issue of what A could expect. That this is so seems to me (though I will not discuss it further here) of some significance.

[38] I will end by suggesting one rather shaky way in which one might approach a resolution of the problem, using only the limited materials already available.

[39] The apparently decisive arguments of the first presentation, which suggested that A should identify himself with the B-body-person, turned on the extreme neatness of the situation in satisfying, if any could, the description of "changing bodies." But this neatness is basically artificial; it is the product of the will of the experimenter to produce a situation which would naturally elicit, with minimum hesitation, that description. By the sorts of methods he employed, he could easily have left off earlier or gone on further. He could have stopped at situation (v), leaving B as he was; or he could have gone on and produced two persons each with A-like character and memories, as well as one or two with B-like characteristics. If he had done either of those, we should have been in yet greater difficulty about what to say; he just chose to make it as easy as possible for us to find something to say. Now if we had some model of ghostly persons in bodies, which were in some sense actually moved around by certain procedures, we could regard the neat experiment just as the *effective* experiment: the one method that really did result in the ghostly persons' changing places without being destroyed, dispersed, or whatever. But we cannot seriously use such a model. The experimenter has not in the sense of that model *induced* a change of bodies; he has rather produced the one situation out of a range of equally possible situations which we should be most disposed to call a change of bodies. As against this, the principle that one's fears can extend to future pain whatever psychological changes precede it seems positively straightforward. Perhaps, indeed, it is not; but we need to be shown what is wrong with it. Until we are shown what is wrong with it, we should perhaps decide that if we were the person A then, if we were to decide selfishly, we should pass the pain to the B-body-person. It would be risky: that there is room for the notion of a risk here is itself a major feature of the problem.

PERSONAL IDENTITY

Derek Parfit

[1] We can, I think, describe cases in which, though we know the answer to every other question, we have no idea how to answer a question about personal identity. These cases are not covered by the criteria of personal identity that we actually use. Do they present a problem?

[2] It might be thought that they do not, because they could never occur. I suspect that some of them could. (Some, for instance, might become scientifically possible.) But I shall claim that even if they did they would present no problem.

[3] My targets are two beliefs: one about the nature of personal identity, the other about its importance.

[4] The first is that in these cases the question about identity must have an answer. No one thinks this about, say, nations or machines. Our criteria for the identity of these do not cover certain cases. No one thinks that in these cases the questions "Is it the same nation?" or "Is it the same machine?" must have answers. Some people believe that in this respect they are different. They agree that our criteria of personal identity do not cover certain cases, but they believe that the nature of their own identity through time is, somehow, such as to guarantee that in these cases questions about their identity must have answers. This belief might be expressed as follows: "Whatever happens between now and any future time, either I shall still exist, or I shall not. Any future experience will either be *my* experience, or it will not."

[5] This first belief -- in the special nature of personal identity -- has, I think, certain effects. It makes people assume that the principle of self-interest is more rationally compelling than any moral principle. And it makes them more depressed by the thought of aging and of death. I cannot see how to disprove this first belief. I shall describe a problem case. But this can only make it seem implausible.

[6] Another approach might be this. We might suggest that one cause of the belief is the projection of our emotions. When we imagine ourselves in a problem case, we do feel that the question "Would it be me?" must have an answer. But what we take to be a bafflement about a further fact may be only the bafflement of our concern.

[7] I shall not pursue this suggestion here. But one cause of our concern is the belief which is my second target. This is that unless the question about identity has an answer, we cannot answer certain important questions (questions about such matters as survival, memory, and responsibility). Against this second belief my claim will be this. Certain important questions do presuppose a question about personal identity. But they can be freed of this presupposition. And when they are, the question about identity has no importance.

I

[8] We can start by considering the much-discussed case of the man who, like an amoeba, divides. Wiggins has recently dramatized this case. He first referred to the operation imagined by Shoemaker. We suppose that my brain is transplanted into someone else's

(brainless) body, and that the resulting person has my character and apparent memories of my life. Most of us would agree, after thought, that the resulting person is me. I shall here assume such agreement.

[9] Wiggins then imagined his own operation. My brain is divided, and each half is housed in a new body. Both resulting people have my character and apparent memories of my life. What happens to me? There seem only three possibilities: (1) I do not survive; (2) I survive as one of the two people; (3) I survive as both.

[10] The trouble with (1) is this. We agreed that I could survive if my brain were successfully transplanted. And people have in fact survived with half their brains destroyed. It seems to follow that I could survive if half my brain were successfully transplanted and the other half were destroyed. But if this is so, how could I *not* survive if the other half were also successfully transplanted? How could a double success be a failure?

[11] We can move to the second description. Perhaps one success is the maximum score. Perhaps I shall be one of the resulting people. The trouble here is that in Wiggins' case each half of my brain is exactly similar, and so, to start with, is each resulting person. So how can I survive as only one of the two people? What can make me one of them rather than the other?

[12] It seems clear that both of these descriptions -- that I do not survive, and that I survive as one of the people -- are highly implausible. Those who have accepted them must have assumed that they were the only possible descriptions. What about our third description: that I survive as both people?

[13] It might be said, "If 'survive' implies identity, this description makes no sense -- you cannot be two people. If it does not, the description is irrelevant to a problem about identity." I shall later deny the second of these remarks. But there are ways of denying the first. We might say, "What we have called 'the two resulting people' are not two people. They are one person. I do survive Wiggins' operation. Its effect is to give me two bodies and a divided mind." It would shorten my argument if this were absurd. But I do not think it is. It is worth showing why. We can, I suggest, imagine a divided mind. We can imagine a man having two simultaneous experiences, in having each of which he is unaware of having the other.

[14] We may not even need to imagine this. Certain actual cases, to which Wiggins referred, seem to be best described in these terms. These involve the cutting of the bridge between the hemispheres of the brain. The aim was to cure epilepsy. But the result appears to be, in the surgeon's words, the creation of "two separate spheres of consciousness," each of which controls one half of the patient's body. What is experienced in each is, presumably, experienced by the patient. There are certain complications in these actual cases. So let us imagine a simpler case.

[15] Suppose that the bridge between my hemispheres is brought under my voluntary control. This would enable me to disconnect my hemispheres as easily as if I were blinking. By doing this I would divide my mind. And we can suppose that when my mind is divided I can, in each half, bring about reunion.

[16] This ability would have obvious uses. To give an example: I am near the end of a maths exam, and see two ways of tackling the last problem. I decide to divide my mind, to work, with each half, at one of two calculations, and then to reunite my mind and write a fair copy of the best result. What shall I experience?

[17] When I disconnect my hemispheres, my consciousness divides into two streams. But this division is not something that I experience. Each of my two streams of consciousness seems to have been straightforwardly continuous with my one stream of consciousness up to the moment of division. The only changes in each stream are the disappearance of half my visual field and the loss of sensation in, and control over, half my body.

[18] Consider my experiences in what we can call my "right-handed" stream. I remember that I assigned my right hand to the longer calculation. This I now begin. In working at this calculation I can see, from the movements of my left hand, that I am also working at the other. But I am not aware of working at the other. So I might, in my right-handed stream, wonder how, in my left-handed stream, I am getting on.

[19] My work is now over. I am about to reunite my mind. What should I, in each stream, expect? Simply that I shall suddenly seem to remember just having thought out two calculations, in thinking out each of which I was not aware of thinking out the other. This, I submit, we can imagine. And if my mind was divided, these memories are correct. In describing this episode, I assumed that there were two series of thoughts, and that they were both mine. If my two hands visibly wrote out two calculations, and if I claimed to remember two corresponding series of thoughts, this is surely what we should want to say. If it is, then a person's mental history need not be like a canal, with only one channel. It could be like a river, with islands, and with separate streams.

[20] To apply this to Wiggins' operation: we mentioned the view that it gives me two bodies and a divided mind. We cannot now call this absurd. But it is, I think, unsatisfactory. There were two features of the case of the exam that made us want to say that only one person was involved. The mind was soon reunited, and there was only one body. If a mind was permanently divided and its halves developed in different ways, the point of speaking of one person would start to disappear. Wiggins' case, where there are also two bodies, seems to be over the borderline. After I have had his operation, the two "products" each have all the attributes of a person. They could live at opposite ends of the earth. (If they later met, they might even fail to recognize each other.) It would become intolerable to deny that they were different people.

[21] Suppose we admit that they are different people. Could we still claim that I survived as both, using "survive" to imply identity? We could. For we might suggest that two people could compose a third. We might say, "I do survive Wiggins' operation as two people. They can be different people, and yet be me, in just the way in which the Pope's three crowns are one crown." This is a possible way of giving sense to the claim that I survive as two different people, using "survive" to imply identity. But it keeps the language of identity only by changing the concept of a person. And there are obvious objections to this change. (Suppose the resulting people fight a duel. Are there three people fighting, one on each side, and one on both? And suppose one of the bullets kills. Are there two acts, one murder and one suicide? How many people are left alive? One? Two?)

[22] The alternative, for which I shall argue, is to give up the language of identity. We can suggest that I survive as two different people without implying that I am these people. When I first mentioned this alternative, I mentioned this objection: "If your new way of talking does not imply identity, it cannot solve our problem. For that is about identity. The problem is that all the possible answers to the question about identity are highly implausible." We can now answer this objection.

[23] We can start by reminding ourselves that this is an objection only if we have one or both of the beliefs which I mentioned at the start of this paper. The first was the belief that to any question about personal identity, in any describable case, there must be a true answer. For those with this belief, Wiggins' case is doubly perplexing. If all the possible answers are implausible, it is hard to decide which of them is true, and hard even to keep the belief that one of them must be true. If we give up this belief; as I think we should, these problems disappear. We shall then regard the case as like many others in which, for quite unpuzzling reasons, there *is* no answer to a question about identity. (Consider "Was England the same nation after 1066?")

[24] Wiggins' case makes the first belief implausible. It also makes it trivial. For it undermines the second belief. This was the belief that important questions turn upon the question about identity. (It is worth pointing out that those who have only this second belief do not think that there must *be* an answer to this question, but rather that we must decide upon an answer.)

[25] Against this second belief my claim is this. Certain questions do presuppose a question about personal identity. And because these questions *are* important, Wiggins' case does present a problem. But we cannot solve this problem by answering the question about identity. We can solve this problem only by taking these important questions and prizing them apart from the question about identity. After we have done this, the question about identity (though we might for the sake of neatness decide it) has no further interest. Because there are several questions which presuppose identity, this claim will take some time to fill out.

[26] We can first return to the question of survival. This is a special case, for survival does not so much presuppose the retaining of identity as seem equivalent to it. It is thus the general relation which we need to prize apart from identity. We can then consider particular relations, such as those involved in memory and intention.

[27] "Will I survive?" seems, I said, equivalent to "Will there be some person alive who is the same person as me?" If we treat these questions as equivalent, then the least unsatisfactory description of Wiggins' case is, I think, that I survive with two bodies and a divided mind.

[28] Several writers have chosen to say that I am neither of the resulting people. Given our equivalence, this implies that I do not survive, and hence, presumably, that even if Wiggins' operation is not literally death, I ought, since I will not survive it, to regard it *as* death. But this seemed absurd.

[29] It is worth repeating why. An emotion or attitude can be criticized for resting on a false belief, or for being inconsistent. A man who regarded Wiggins' operation as death must, I suggest, be open to one of these criticisms. He might believe that his relation to each of the resulting people fails to contain some element which is contained in survival. But how can this be true? We agreed that he *would* survive if he stood in this very same relation to only *one* of the resulting people. So it cannot be the nature of this relation which makes it fail, in Wiggins' case, to be survival. It can only be its duplication.

[30] Suppose that our man accepts this, but still regards division as death. His reaction would now seem wildly inconsistent. He would be like a man who, when told of a drug that could double his years of life, regarded the taking of this drug as death. The only difference in the case of division is that the extra years are to run concurrently. This is an interesting difference. But it cannot mean that there are *no* years to run.

[31] I have argued this for those who think that there must, in Wiggins' case, be a true answer to the question about identity. For them, we might add, "Perhaps the original person does lose his identity. But there may be other ways to do this than to die. One other way might be to multiply. To regard these as the same is to confuse nought with two."

[32] For those who think that the question of identity is up for decision, it would be clearly absurd to regard Wiggins' operation as death. These people would have to think, "We could have chosen to say that I should be one of the resulting people. If we had, I should not have regarded it as death. But since we have chosen to say that I am neither person, I do." This is hard even to understand.

[33] My first conclusion, then, is this. The relation of the original person to each of the resulting people contains all that interests us -- all that matters -- in any ordinary case of survival. This is why we need a sense in which one person can survive as two. One of my aims in the rest of this paper will be to suggest such a sense. But we can first make some general remarks.

II

[34] Identity is a one-one relation. Wiggins' case serves to show that what matters in survival need not be one-one. Wiggins' case is of course unlikely to occur. The relations which matter are, in fact, one-one. It is because they are that we can imply the holding of these relations by using the language of identity. This use of language is convenient. But it can lead us astray. We may assume that what matters *is* identity and, hence, has the properties of identity.

[35] In the case of the property of being one-one, this mistake is not serious. For what matters is in fact one-one. But in the case of another property, the mistake *is* serious. Identity is all-or-nothing. Most of the relations which matter in survival are, in fact, relations of degree. If we ignore this, we shall be led into quite ill-grounded attitudes and beliefs.

[36] The claim that I have just made -- that most of what matters are relations of degree -- I have yet to support. Wiggins' case shows only that these relations need not be one-one. The merit of the case is not that it shows this in particular, but that it makes the first break between what matters and identity. The belief that identity *is* what matters is hard to overcome. This is shown in most discussions of the problem cases which actually occur: cases, say, of amnesia or of brain damage. Once Wiggins' case has made one breach in this belief, the rest should be easier to remove.

[37] To turn to a recent debate: most of the relations which matter can be provisionally referred to under the heading "psychological continuity" (which includes causal continuity). My claim is thus that we use the language of personal identity in order to imply such continuity. This is close to the view that psychological continuity provides a criterion of identity. Williams has attacked this view with the following argument. Identity is a one-one relation. So any criterion of identity must appeal to a relation which is logically one-one. Psychological continuity is not logically one-one. So it cannot provide a criterion.

[38] Some writers have replied that it is enough if the relation appealed to is always in fact one-one. I suggest a slightly different reply. Psychological continuity is a ground for speaking of identity when it is one-one. If psychological continuity took a one-many or

branching form, we should need, I have argued, to abandon the language of identity. So this possibility would not count against this view.

[39] We can make a stronger claim. This possibility would count in its favor. The view might be defended as follows. Judgments of personal identity have great importance. What gives them their importance is the fact that they imply psychological continuity. This is why, whenever there is such continuity, we ought, if we can, to imply it by making a judgment of identity.

[40] If psychological continuity took a branching form, no coherent set of judgments of identity could correspond to, and thus be used to imply, the branching form of this relation. But what we ought to do, in such a case, is take the importance which would attach to a judgment of identity and attach this importance directly to each limb of the branching relation. So this case helps to show that judgments of personal identity do derive their importance from the fact that they imply psychological continuity. It helps to show that when we can, usefully, speak of identity, this relation is our ground.

[41] This argument appeals to a principle which Williams put forward. The principle is that an important judgment should be asserted and denied only on importantly different grounds. Williams applied this principle to a case in which one man is psychologically continuous with the dead Guy Fawkes, and a case in which two men are. His argument was this. If we treat psychological continuity as a sufficient ground for speaking of identity, we shall say that the one man is Guy Fawkes. But we could not say that the two men are, although we should have the same ground. This disobeys the principle. The remedy is to deny that the one man is Guy Fawkes, to insist that sameness of the body is necessary for identity.

[42] Williams' principle can yield a different answer. Suppose we regard psychological continuity as more important than sameness of the body. And suppose that the one man really is psychologically (and causally) continuous with Guy Fawkes. If he is, it would disobey the principle to deny that he is Guy Fawkes, for we have the same important ground as in a normal case of identity. In the case of the two men, we again have the same important ground. So we ought to take the importance from the judgment of identity and attach it directly to this ground. We ought to say, as in Wiggins' case, that each limb of the branching relation is as good as survival. This obeys the principle.

[43] To sum up these remarks: even if psychological continuity is neither logically, nor always in fact, one-one, it can provide a criterion of identity. For this can appeal to the relation of *non-branching* psychological continuity, which is logically one-one.

[44] The criterion might be sketched as follows. "X and Y are the same person if they are psychologically continuous and there is no person who is contemporary with either and psychologically continuous with the other." We should need to explain what we mean by "psychologically continuous" and say how much continuity the criterion requires. We should then, I think, have described a sufficient condition for speaking of identity.

[45] We need to say something more. If we admit that psychological continuity might not be one-one, we need to say what we ought to do if it were not one-one. Otherwise our account would be open to the objections that it is incomplete and arbitrary. I have suggested that if psychological continuity took a branching form, we ought to speak in a new way, regarding what we describe as having the same significance as identity. This answers these objections.

[46] We can now return to our discussion. We have three remaining aims. One is to suggest a sense of "survive" which does not imply identity. Another is to show that most of what matters in survival are relations of degree. A third is to show that none of these relations needs to be described in a way that presupposes identity. We can take these aims in the reverse order.

III

[47] The most important particular relation is that involved in memory. This is because it is so easy to believe that its description must refer to identity. This belief about memory is an important cause of the view that personal identity has a special nature. But it has been well discussed by Shoemaker and by Wiggins. So we can be brief.

[48] It may be a logical truth that we can only remember our own experiences. But we can frame a new concept for which this is not a logical truth. Let us call this "q-memory." To sketch a definition: I am q-remembering an experience if (1) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way (whatever that is) in which a memory of an experience is dependent upon it.

[49] According to (1) q-memories seem like memories. So I q-remember *having* experiences. This may seem to make q-memory presuppose identity. One might say, "My apparent memory of *having* an experience is an apparent memory of *my* having an experience. So how could I q-remember my having other people's experiences?"

[50] This objection rests on a mistake. When I seem to remember an experience, I do indeed seem to remember *having* it. But it cannot be a part of what I seem to remember about this experience that I, the person who now seems to remember it, am the person who had this experience. That I am is something that I automatically assume. (My apparent memories sometimes come to me simply as the belief that I had a certain experience.) But it is something that I am justified in assuming only because I do not in fact have q-memories of other people's experiences.

[51] Suppose that I did start to have such q-memories. If I did, I should cease to assume that my apparent memories must be about my own experiences. I should come to assess an apparent memory by asking two questions: (1) Does it tell me about a past experience? (2) If so, whose? Moreover (and this is a crucial point) my apparent memories would now come to me *as* q-memories. Consider those of my apparent memories which do come to me simply as beliefs about my past: for example, "I did that." If I knew that I could q-remember other people's experiences, these beliefs would come to me in a more guarded form: for example, "Someone -- probably I -- did that." I might have to work out who it was.

[52] I have suggested that the concept of q-memory is coherent. Wiggins' case provides an illustration. The resulting people, in his case, both have apparent memories of living the life of the original person. If they agree that they are not this person, they will have to regard these as only q-memories. And when they are asked a question like "Have you heard this music before?" they might have to answer "I am sure that I q-remember hearing it. But I am not sure whether I remember hearing it. I am not sure whether it was I who heard it, or the original person."

[53] We can next point out that on our definition every memory is also a q-memory. Memories are, simply, q-memories of one's own experiences. Since this is so, we could afford now to drop the concept of memory and use in its place the wider concept q-memory. If we did, we should describe the relation between an experience and what we now call a "memory" of this experience in a way which does not presuppose that they are had by the same person. This way of describing this relation has certain merits. It vindicates the "memory criterion" of personal identity against the charge of circularity. And it might, I think, help with the problem of other minds.

[54] But we must move on. We can next take the relation between an intention and a later action. It may be a logical truth that we can intend to perform only our own actions. But intentions can be redescribed as q-intentions. And one person could q-intend to perform another person's actions.

[55] Wiggins' case again provides the illustration. We are supposing that neither of the resulting people is the original person. If so, we shall have to agree that the original person can, before the operation, q-intend to perform their actions. He might, for example, q-intend, as one of them, to continue his present career, and, as the other, to try something new. (I say "q-intend *as one of them*" because the phrase "q-intend *that one of them*" would not convey the directness of the relation which is involved. If I intend that someone else should do something, I cannot get him to do it simply by forming this intention. But if I am the original person, and he is one of the resulting people, I can.)

[56] The phrase "q-intend *as one of them*" reminds us that we need a sense in which one person can survive as two. But we can first point out that the concepts of q-memory and q-intention give us our model for the others that we need: thus, a man who can q-remember could q-recognize, and be a q-witness of, what he has never seen; and a man who can q-intend could have q-ambitions, make q-promises, and be q-responsible for.

[57] To put this claim in general terms: many different relations are included within, or are a consequence of, psychological continuity. We describe these relations in ways which presuppose the continued existence of one person. But we could describe them in new ways which do not.

[58] This suggests a bolder claim. It might be possible to think of experiences in a wholly "impersonal" way. I shall not develop this claim here. What I shall try to describe is a way of thinking of our own identity through time which is more flexible, and less misleading, than the way in which we now think. This way of thinking will allow for a sense in which one person can survive as two. A more important feature is that it treats survival as a matter of degree.

IV

[59] We must first show the need for this second feature. I shall use two imaginary examples. The first is the converse of Wiggins' case: fusion. Just as division serves to show that what matters in survival need not be one-one, so fusion serves to show that it can be a question of degree.

[60] Physically, fusion is easy to describe. Two people come together. While they are unconscious, their two bodies grow into one. One person then wakes up. The psychology of fusion is more complex. One detail we have already dealt with in the case of the exam. When my mind was reunited, I remembered just having thought out two calculations. The

one person who results from a fusion can, similarly, q-remember living the lives of the two original people. None of their q-memories need be lost.

[61] But some things must be lost. For any two people who fuse together will have different characteristics, different desires, and different intentions. How can these be combined? We might suggest the following. Some of these will be compatible. These can coexist in the one resulting person. Some will be incompatible. These, if of equal strength, can cancel out, and if of different strengths, the stronger can be made weaker. And all these effects might be predictable.

[62] To give examples -- first, of compatibility: I like Palladio and intend to visit Venice. I am about to fuse with a person who likes Giotto and intends to visit Padua. I can know that the one person we shall become will have both tastes and both intentions. Second, of incompatibility: I hate red hair, and always vote Labour. The other person loves red hair, and always votes Conservative. I can know that the one person we shall become will be indifferent to red hair, and a floating voter.

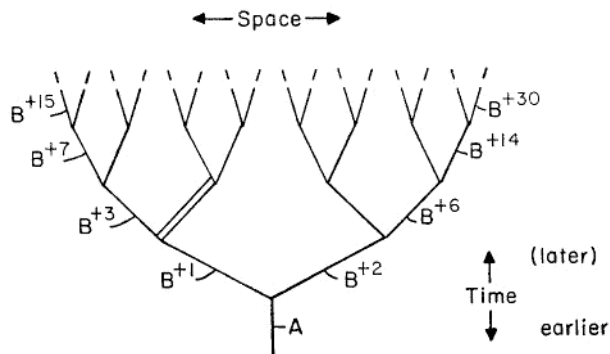
[63] If we were about to undergo a fusion of this kind, would we regard it as death? Some of us might. This is less absurd than regarding division as death. For after my division the two resulting people will be in every way like me, while after my fusion the one resulting person will not be wholly similar. This makes it easier to say, when faced with fusion, "I shall not survive," thus continuing to regard survival as a matter of all-or-nothing. This reaction is less absurd. But here are two analogies which tell against it.

[64] First, fusion would involve the changing of some of our characteristics and some of our desires. But only the very self-satisfied would think of this as death. Many people welcome treatments with these effects.

[65] Second, someone who is about to fuse can have, beforehand, just as much "intentional control" over the actions of the resulting individual as someone who is about to marry can have, beforehand, over the actions of the resulting couple. And the choice of a partner for fusion can be just as well considered as the choice of a marriage partner. The two original people can make sure (perhaps by "trial fusion") that they do have compatible characters, desires, and intentions.

[66] I have suggested that fusion, while not clearly survival, is not clearly failure to survive, and hence that what matters in survival can have degrees. To reinforce this claim we can now turn to a second example. This is provided by certain imaginary beings. These beings are just like ourselves except that they reproduce by a process of natural division. We can illustrate the histories of these imagined beings with the aid of a diagram. The lines on the diagram represent the spatiotemporal paths which would be traced out by the bodies of these beings. We can call each single line (like the double line) a "branch"; and we can call the whole structure a "tree." And let us suppose that each "branch" corresponds to what is thought of as the life of one individual.

[67] These individuals are referred to as "A," "B+1," and so forth. Now, each single division is an instance of Wiggins' case. So A's relation to both B+1 and B+2 is just as good as survival. But what of A's relation to B+30? I said earlier that what matters in survival could be provisionally referred to as "psychological continuity." I must now distinguish this relation from another, which I shall call "psychological connectedness."



[68] Let us say that the relation between a q-memory and the experience q-remembered is a "direct" relation. Another "direct" relation is that which holds between a q-intention and the q-intended action. A third is that which holds between different expressions of some lasting q-characteristic.

[69] "Psychological connectedness," as I define it, requires the holding of these direct psychological relations. "Connectedness" is not transitive, since these relations are not transitive. Thus, if X q-remembers most of Y's life, and Y q-remembers most of Z's life, it does not follow that X q-remembers most of Z's life. And if X carries out the q-intentions of Y, and Y carries out the q-intentions of Z it does not follow that X carries out the q-intentions of Z.

[70] "Psychological continuity," in contrast, only requires overlapping chains of direct psychological relations. So "continuity" is transitive.

[71] To return to our diagram. A is psychologically continuous with B+30. There are between the two continuous chains of overlapping relations. Thus, A has q-intentional control over B+2, B+2 has q-intentional control over B+6, and so on up to B+30. Or B+30 can q-remember the life of B+14, B+14 can q-remember the life of B+6, and so on back to A.

[72] A, however, need *not* be psychologically connected to B+30. Connectedness requires direct relations. And if these beings are like us, A cannot stand in such relations to every individual in his indefinitely long "tree." Q-memories will weaken with the passage of time, and then fade away. Q-ambitions, once fulfilled, will be replaced by others. Q-characteristics will gradually change. In general, A stands in fewer and fewer direct psychological relations to an individual in his "tree" the more remote that individual is. And if the individual is (like B+30) sufficiently remote, there may be between the two *no* direct psychological relations.

[73] Now that we have distinguished the general relations of psychological continuity and psychological connectedness, I suggest that connectedness is a more important element in survival. As a claim about our own survival, this would need more arguments than I have space to give. But it seems clearly true for my imagined beings. A is as close psychologically to B + 1 as I today am to myself tomorrow. A is as distant from B + 30 as I am from my great-great-grandson. Even if connectedness is not more important than continuity, the fact that one of these is a relation of degree is enough to show that what matters in survival can

have degrees. And in any case the two relations are quite different. So our imagined beings would need a way of thinking in which this difference is recognized.

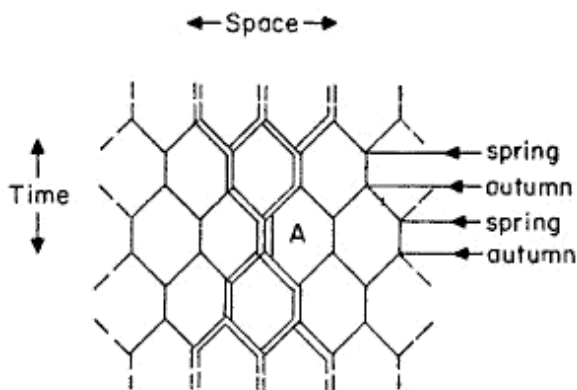
V

[74] What I propose is this. First, A can think of any individual, anywhere in his "tree," as "a descendant self." This phrase implies psychological continuity. Similarly, any later individual can think of any earlier individual on the single path which connects him to A as "an ancestral self." Since psychological continuity is transitive, "being an ancestral self of" and "being a descendant self of" are also transitive. To imply psychological connectedness I suggest the phrases "one of my future selves" and "one of my past selves."

[75] These are the phrases with which we can describe Wiggins' case. For having past and future selves is, what we needed, a way of continuing to exist which does not imply identity through time. The original person does, in this sense, survive Wiggins' operation: the two resulting people are his later selves. And they can each refer to him as "my past self." (They can share a past self without being the same self as each other.)

[76] Since psychological connectedness is not transitive, and is a matter of degree, the relations "being a past self of" and "being a future self of" should themselves be treated as relations of degree. We allow for this series of descriptions: "my most recent self," "one of my earlier selves," "one of my distant selves," "hardly one of *my* past selves (I can only q-remember a few of his experiences)," and, finally, "not in any way one of *my* past selves — just an ancestral self."

[77] This way of thinking would clearly suit our first imagined beings. But let us now turn to a second kind of being. These reproduce by fusion as well as by division. And let us suppose that they fuse every autumn and divide every spring. This yields the following diagram:



[78] If A is the individual whose life is represented by the three-lined "branch," the two-lined "tree" represents those lives which are psychologically continuous with A's life. (It can be seen that each individual has his own "tree," which overlaps with many others.)

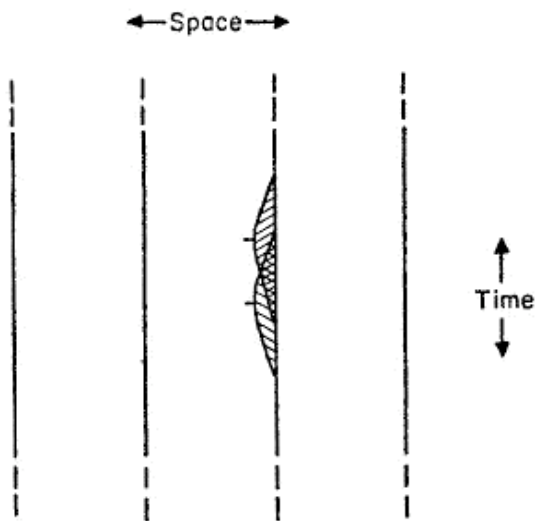
[79] For the imagined beings in this second world, the phrases "an ancestral self" and "a descendant self" would cover too much to be of much use. (There may well be pairs of dates such that every individual who ever lived before the first date was an ancestral self of

every individual who ever will live after the second date.) Conversely, since the lives of each individual last for only half a year, the word "I" would cover too little to do all of the work which it does for us. So part of this work would have to be done, for these second beings, by talk about past and future selves.

[80] We can now point out a theoretical flaw in our proposed way of thinking. The phrase "a past self of" implies psychological connectedness. Being a past self of is treated as a relation of degree, so that this phrase can be used to imply the varying degrees of psychological connectedness. But this phrase can imply only the degrees of connectedness between different lives. It cannot be used within a single life. And our way of delimiting successive lives does not refer to the degrees of psychological connectedness. Hence there is no guarantee that this phrase, "a past self of," could be used whenever it was needed. There is no guarantee that psychological connectedness will not vary in degree within a single life.

[81] This flaw would not concern our imagined beings. For they divide and unite so frequently, and their lives are in consequence so short, that within a single life psychological connectedness would always stand at a maximum.

[82] But let us look, finally, at a third kind of being. In this world there is neither division nor union. There are a number of everlasting bodies, which gradually change in appearance. And direct psychological relations, as before, hold only over limited periods of time. This can be illustrated with a third diagram. In this diagram the two shadings represent the degrees of psychological connectedness to their two central points.



[83] These beings could not use the way of thinking that we have proposed. Since there is no branching of psychological continuity, they would have to regard themselves as immortal. It might be said that this is what they are. But there is, I suggest, a better description. Our beings would have one reason for thinking of themselves as immortal. The parts of each "line" are all psychologically continuous. But the parts of each "line" are not all psychologically connected. Direct psychological relations hold only between those parts which are close to each other in time. This gives our beings a reason for *not* thinking of each "line" as corresponding to one single life. For if they did, they would have no way of implying these direct relations. When a speaker says, for example, "I spent a period doing

such and such," his hearers would not be entitled to assume that the speaker has any memories of this period, that his character then and now are in any way similar, that he is now carrying out any of the plans or intentions which he then had, and so forth. Because the word "I" would carry none of these implications, it would not have for these "immortal" beings the usefulness which it has for us.

[84] To gain a better way of thinking, we must revise the way of thinking that we proposed above. The revision is this. The distinction between successive selves can be made by reference, not to the branching of psychological continuity, but to the degrees of psychological connectedness. Since this connectedness is a matter of degree, the drawing of these distinctions can be left to the choice of the speaker and be allowed to vary from context to context.

[85] On this way of thinking, the word "I" can be used to imply the greatest degree of psychological connectedness. When the connections are reduced, when there has been any marked change of character or style of life, or any marked loss of memory, our imagined beings would say, "It was not I who did that, but an earlier self." They could then describe in what ways, and to what degree, they are related to this earlier self. This revised way of thinking would suit not only our "immortal" beings. It is also the way in which we ourselves could think about our lives. And it is, I suggest, surprisingly natural.

[86] One of its features, the distinction between successive selves, has already been used by several writers. To give an example, from Proust: "we are incapable, while we are in love, of acting as fit predecessors of the next persons who, when we are in love no longer, we shall presently have become" [*Within a Budding Grove* (London, 1949), I, 226 (my own translation).]

[87] Although Proust distinguished between successive selves, he still thought of one person as being these different selves. This we would not do on the way of thinking that I propose. If I say, "It will not be me, but one of my future selves," I do not imply that I will be that future self. He is one of my later selves, and I am one of his earlier selves. There is no underlying person who we both are.

[88] To point out another feature of this way of thinking. When I say, "There is no person who we both are," I am only giving my decision. Another person could say, "It will be you," thus deciding differently. There is no question of either of these decisions being a mistake. Whether to say "I," or "one of my future selves," or "a descendant self" is entirely a matter of choice. The matter of fact, which must be agreed, is only whether the disjunction applies. (The question "Are X and Y the same person?" thus becomes "Is X at least an ancestral [or descendant] self of Y?")

VI

[89] I have tried to show that what matters in the continued existence of a person are, for the most part, relations of degree. And I have proposed a way of thinking in which this would be recognized. I shall end by suggesting two consequences and asking one question.

[90] It is sometimes thought to be especially rational to act in our own best interests. But I suggest that the principle of self-interest has no force. There are only two genuine competitors in this particular field. One is the principle of biased rationality: do what will best achieve what you actually want. The other is the principle of impartiality: do what is in the best interests of everyone concerned.

[91] The apparent force of the principle of self-interest derives, I think, from these two other principles. The principle of self-interest is normally supported by the principle of biased rationality. This is because most people care about their own future interests.

[92] Suppose that this prop is lacking. Suppose that a man does not care what happens to him in, say, the more distant future. To such a man, the principle of self-interest can only be propped up by an appeal to the principle of impartiality. We must say, "Even if you don't care, you ought to take what happens to you then equally into account." But for this, as a special claim, there seem to me no good arguments. It can only be supported as part of the general claim, "You ought to take what happens to everyone equally into account."

[93] The special claim tells a man to grant an *equal* weight to all the parts of his future. The argument for this can only be that all the parts of his future are *equally* parts of his future. This is true. But it is a truth too superficial to bear the weight of the argument. (To give an analogy: The unity of a nation is, in its nature, a matter of degree. It is therefore only a superficial truth that all of a man's compatriots are *equally* his compatriots. This truth cannot support a good argument for nationalism.)

[94] I have suggested that the principle of self-interest has no strength of its own. If this is so, there is no special problem in the fact that what we ought to do can be against our interests. There is only the general problem that it may not be what we want to do.

[95] The second consequence which I shall mention is implied in the first. Egoism, the fear not of near but of distant death, the regret that so much of one's only life should have gone by -- these are not, I think, wholly natural or instinctive. They are all strengthened by the beliefs about personal identity which I have been attacking. If we give up these beliefs, they should be weakened.

[96] My final question is this. These emotions are bad, and if we weaken them we gain. But can we achieve this gain without, say, also weakening loyalty to, or love of, other particular selves? As Hume warned, the "refined reflections which philosophy suggests ... cannot diminish ... our vicious passions ... without diminishing... such as are virtuous. They are ... applicable to all our affections. In vain do we hope to direct their influence only to one side."

[97] That hope is vain. But Hume had another: that more of what is bad depends upon false belief. This is also my hope.