

NO SELF TO BE FOUND

The Search for Personal Identity

James Giles

PhD

University of Edinburgh

1989



ABSTRACT

This study begins by seeing what is involved in the notion of identity and then moves on to argue that the problem of personal identity is an empirical problem. From here an attack is launched on Butler's and Reid's idea of a persisting self which is strictly identical over time. Other attempts to support this view of personal identity, such as those of McTaggart, Madell, and Chisholm are also taken to task. The next targets are the memory theory of Locke and the psychological theories of Quinton and Parfit. All of these ultimately fail, it is argued, because they ignore the structure of experience and continue to cleave to the model of strict identity. Proceeding to the bodily theory we find both Williams' and Penelhum's effort to disparage the possibility of bodily transfer are of no avail. Williams' further attempt to acclaim the bodily theory on the grounds that to love a person is basically to love a body proves no better. The final reason for the collapse of the bodily theory is because it overlooks the first-person perspective on the body. With this we turn to the no-self account which, being an eliminativist rather than a reductive view, is a rejection of the idea of personal identity. Hume's doctrine is shown to be such an account and is defended against Penelhum, Noxon, and Ashley and Stack. This involves an appeal to the Buddhist theory of the two levels of truth which, it is argued, is already implicit in Hume. To defend Hume against Kemp Smith's charge of inconsistency a phenomenological enquiry into the nature of the constructed or condensed self-image is carried out. The study ends by suggesting the path that related ethical enquiries might pursue.

If you think the mind
That attains enlightenment
Is 'mine'
Your thoughts will wrestle, one with
the other

These days I'm not bothered about
Getting enlightenment all the time
And the result is that
I wake up in the morning feeling
fine!

Bankei Zenji
(1622-1693)

CONTENTS

PREFACE	1
1. THE PROBLEM OF PERSONAL IDENTITY	5
2. THE INDEFINABLE SELF	20
3. THE PSYCHOLOGICAL THEORY	65
4. THE BODILY THEORY AND THE THEORY OF THE BODY ...	129
5. NO SELF TO BE FOUND	178
BIBLIOGRAPHY	249

PREFACE

This study is an attempt to address one of the central issues in philosophical psychology: the problem of personal identity. This problem is often said to be one of accounting for what it is that gives persons their identity. However, once the problem has been construed in these terms it is plain that too much has already been assumed. For what has been assumed is just that persons do have an identity. To the philosophers who approach the problem with this supposition firmly ensconced in their minds the possibility that there may be no such thing as personal identity is scarcely conceived. As a result the more fundamental question - whether or not personal identity exists in the first place - remains unasked. The inadequacy of one theory of identity is then taken as evidence for the truth of another theory, rather than as evidence for the rejection of such identity altogether. And this devotion to the idea of identity persists even when all of the competing accounts turn

out to be clearly inadequate.

One of the reasons for this tenacious clinging to the idea of personal identity is perhaps the (philosopher's) belief that the ordinary (non-philosophical) person has some sort of deep conviction concerning his own identity. This supposed 'common sense' view is then revered as the ultimate court of appeal in which each philosophical theory must finally present its case. It shall be my contention, however, that the idea of a common sense commitment to this or that theory of personal identity, or even to the idea of personal identity itself, is nothing but an offspring of the philosopher's imagination. For not only do ordinary people disagree and argue about the nature of identity (just like ordinary philosophers do), but many such people openly dismiss the idea as sheer nonsense.

In what follows, therefore, my concern shall not be with the question of which theory of personal identity is correct, but rather with the question of whether there is personal identity at all. This approach has the added benefit of freeing us from the urge to invent a principle of identity where none can be found.

First, however, a word about my use of the third-person singular pronoun is in order. In discussing a topic like persons it is difficult to avoid using these pronouns. Where

the gender of the person is left unspecified one therefore has to make a choice between the masculine, the feminine, or the masculine/feminine disjunction. To avoid both awkwardness of style and non-standard English I have decided in such cases to stick with accepted usage and, except where it reads better, to employ the masculine pronoun generically to indicate both male and female. I am of course sensitive to the inadequacies that this involves but hope the reader will accept my good faith in this matter.

Although many people have helped me with this research, there are some who deserve special acknowledgement. First, I would like to thank my supervisor Geoffrey Madell for our many meetings and discussions, for all his help, and especially for his sense of humour. Without his guidance this work could not have been completed. Thanks also are due to John Llewellyn and Peter Lewis, both of whom gave freely of their time and learning whenever I was in need. John Llewellyn's comments on Chapter Four were particularly helpful. I would also like to express my appreciation to Vincent Hope for both the interest he took in my work and his constant support. I learned much from him over our numerous cups of tea at Buccleuch Place. A special thanks must go to Timothy Sprigge who was always willing to see me and help me out of the philosophical quandaries in which I would often become enmeshed. I owe much to him because he took the time to read these pages and offer

his comments and encouragement. Here too I must give my thanks to Mairi McLeod whose knowledge of Eastern thought and the selflessness of all things has been a continual inspiration to me. I would also here like to thank Andrew Brennan for kindly agreeing to be my external examiner. Finally, I must express my gratitude to the University of Edinburgh and the Committee of Vice-Chancellors and Principals of the Universities and Colleges of the United Kingdom, both from whom I received the scholarships that made this research possible.

CHAPTER ONE

THE PROBLEM OF PERSONAL IDENTITY

Like many problems in philosophy the problem of personal identity arises from a discrepancy between what at first glance seems to be a reasonable belief, and the results of careful examination of that belief. It is one of our apparently reasonable beliefs that each of us is in possession of a personal identity, that is, something which distinguishes us from others and imparts some form of unity to our existence. Yet, when we bring this belief into the open air and begin to examine it, strangely enough, it seems a belief with little foundation. To appreciate the difficulties here let us begin by seeing what is involved in the notion of identity.

The first thing to notice is that to assert the existence of identity is to assert the existence of a certain type of relation. But a relation must relate something to something.

Thus to say that there is an instance of identity is the same as saying there are certain things that have the relation of identity to one another. But what sort of relation is identity? To this we can only answer it is the relation of 'one and the same'. And so to say that one thing is identical with another is to say that one thing is one and the same as another. But no sooner has this been accepted than we realize we have come upon a peculiar problem. For if the two items at either end of the identity relation are one and the same thing, then in fact there are not two items at all but only one. And if there is only one item, then it seems we cannot have relation; for as we have just said a relation must relate something to something.

Some philosophers have tried to deal with this problem by saying identity is the relation that every object has to itself. But this will get us no further. For what could it possibly mean to say 'This object is identical with itself'? If we consider that the function of the word 'itself' is to automatically refer to the same thing that was just mentioned, then 'this object' and 'itself' will obviously refer to the same thing. And as is evident, if the idea referred to by the word 'object' is undistinguishable from the idea referred to by the word 'itself', then the statement 'This object is identical with itself' is empty, i.e. it says nothing. (1)

Another way of putting this is to say that such a statement is

little more than an obvious variation of 'This object is identical with this object', which is not really an identity-statement because it is not really a statement at all. It is easy to be fooled here and think that statements of this form are genuine statements; they do after all display the typical structure of a statement where information is conveyed by predicating something of something else. But, as is clear, in this type of statement there is no something else. As a result, 'this object is identical with this object' conveys no information. This is corroborated by a moment's reflection; for if we think about what it is we learn from being told that this object is identical with this object, we must confess we have learned nothing. This is not to say that formulas like 'A=A' do not have a function to fulfill in certain systems of logic, but only that whatever function they have it cannot be one of conveying information.²

We are not, however, stranded. For in translating 'This object is identical with itself' into 'This object is identical with this object' I made use of the fact that 'itself' is obviously referring to the same thing as 'this object'. But what if the second term in an identity statement is not obviously referring to the same thing as the first term? What if, for example, we make the claim 'The Vernal Equinox is identical with the 20th of March', or 'Renoir's favourite renaissance sculptor is identical with the sculptor of the

reliefs of the Fountains des Innocents', or again 'The city whose heart is Dam Square is identical with the Venice of the North'? In such identity statements it is not obvious that the two terms refer to the same thing. This is because the two terms on either side of the identity relation display a different sense from each other, i.e. what Frege describes as a different 'mode of presentation'.(3) We might also say, they are two different ways of denoting the same thing. Thus because the term 'The city whose heart is Dam Square' has a different sense from 'The Venice of the North', we learn something when we are told that the city whose heart is Dam Square is identical with the Venice of the North.

But what sort of thing is it we learn? To answer this we must look to what it is that distinguishes identity statements of this sort from empty claims like 'This object is identical with this object'. And what distinguishes them, as we have noted, is that in the former case the terms of the identity have different senses. What we learn therefore is that two terms with different senses apply to the same thing. But if this is what is conveyed to us by such identity statements then it seems they teach us little more than how we are to use certain terms. That is, they do not convey to us anything about an instance of identity existing in the world.

This is a point well-made by T. L. S. Sprigge.(4)

According to Sprigge, identity statements of the form we are now considering fall short of being full-blooded identity statements for two reasons. First, they are creatures of language who owe their existence to the linguistic fact that there exists more than one term to denote the same thing. Secondly, their sole purpose is to dispel ignorance which, once being done, leaves us with no interesting fact of identity to contemplate. Thus says Sprigge, if 'x is identical to y' is a statement of this form, the only reason for making it is because people do not or might not be aware that x and y are two different ways of denoting the same thing, say z. However, once this is realized then there is no longer any identity relation to report. The natural way of reporting the fact contained in this statement then becomes 'z alone is x and z alone is y' or 'z alone is both x and y'. And here there is no fact of identity, only the fact that one thing plays two different roles. Applying this analysis to one of the above examples we can say that once the relevant ignorance has been dispelled about Amsterdam, i.e. once we know that the city whose heart is Dam Square is identical with the Venice of the North, the natural way of expressing this is not to make an assertion of identity but merely to say 'Amsterdam alone is both the city whose heart is Dam Square and the Venice of the North'.

There is of course the possible protest that the term

'alone' conceals an implicit identity relation. So even if the statement 'z alone is both x and y' does not display an overt assertion of identity, such an assertion will become explicit once the statement has been analyzed. Sprigge's reply here is simply that any such analysis will be artificial. For to hold that the statement 'Jane alone is the woman in the room' is really an identity statement we will have to invoke a logician's analysis and say something like 'for every x, x is a woman in the room if and only if x is identical to Jane'. But, as Sprigge remarks, this suggests our recognition of some individual's having a unique property involves our recognizing of every single thing that it has that property if and only if it is identical with that individual. and this is plainly what we do not do. What we recognize here is that this individual alone has a certain property.

So once again we are no further ahead in understanding just how identity can be a relation and yet involve only one object. Sprigge's solution to this dilemma is to point to the existence of identity statements that involve the notion of pattern or quality recognition. An example of this type of statement might be 'The colour of my jacket is identical with the colour of your scarf'. What is asserted here is the identity of a quality (a colour) which appears in two different instances (my jacket and your scarf). Statements of this sort, says Sprigge, are genuine statements of identity because,

first, they assert a real identity, secondly, they are not vacuous, and thirdly, they are not merely ignorance dispelling; that is, even when we have been suitably enlightened there is still something of interest we can contemplate.

I agree with Sprigge that these sort of statements are neither vacuous nor merely ignorance dispelling, but I am not so sure that they assert a real identity. That is, I am not convinced that statements of pattern or quality recognition are not merely asserting an exact similarity rather than a real identity. On this view the word 'identical' in the statement 'The colour of my jacket is identical with the colour of your scarf' would have the same meaning as it does in the phrase 'identical twins'. For identical twins are not really identical (i.e. one and the same), they are only exactly similar. Consequently, the above statement would only be an assertion of an exact similarity of colour, not of a real identity of colours. Anyway, to hold that pattern or quality recognition statements do assert a real identity would be to commit ourselves to the existence of universals that manifest themselves in particular instances. I think, however, there is another way out of this impasse which enables us to avoid making such commitments. And this is by appealing to the idea of time. For once we introduce the idea of time into the discussion, then it becomes clear how identity can be a relation which nevertheless involves only one thing. For now we

can say that the relation of identity is what occurs when an object at one time is identical with an object at another time.(5) Were we to make a statement about such a relation we would have an identity statement that is neither vacuous nor merely ignorance dispelling. We would have instead an assertion of a real instance of identity. An example of such a statement might be 'The sundial now in the garden is identical with the one put there last year'. Here we have a reference to only one object (the sundial in the garden) and yet, because of the introduction of a temporal dimension, two genuinely different contexts (now and last year) in which the same object can be said to exist. It is clear therefore why this statement is not vacuous: it is making a definite claim about the world which may or may not be true. Further, it is not merely ignorance dispelling; for the two terms of the identity, namely 'the sundial in the garden now' and 'the sundial in the garden last year', are not merely different ways of referring to the same thing, rather they refer to genuinely distinct contexts in which the same thing is said to exist.

We can conclude then that what makes an identity statement a statement about a genuine fact of identity is that it refers to an instance of identity over time. The problem of personal identity then becomes the problem of what, if anything, makes a person at one time identical with himself at another time.(6) This does not mean that questions about non-temporal aspects of

persons will have no bearing on the issue of personal identity over time. For if we can show that what we call a person is at any one moment merely a collection of related parts, then it will prove more likely that a person considered over time is merely a succession of related parts. And this in turn will have implications about whether or not a person has identity over time. It is because of this point that, in what follows, we will have occasion to examine the doctrine of the unity of consciousness, i.e. the view that consciousness is, at any one instance, necessarily a unified entity.

Before, however, we press on to these issues there remains a preliminary question that must be addressed. And this is the question of how the problem of personal identity ought to be regarded. For the most part, philosophers have tended to regard the problem as a conceptual one to be solved by conceptual analysis. This approach would involve analyzing the concept of personal identity in the hope of discovering whether it is a basic concept, or whether it is built up from or dependent on simpler concepts. Thus it might be found that concept of personal identity is dependent on the concept of memory. If this were the case, it would mean that statements about personal identity would logically entail statements about memory, and consequently, it would be argued, that personal identity is somehow constituted by memory.

But this sort of approach cannot but leave us dissatisfied. For if we are genuinely concerned about the problem of personal identity, then arguments that invoke premises about concepts and language will only appear to us as so much mist before our eyes. What we want to know is how persons undergo their existence and whether they experience themselves as possessing an identity through time; and here the analysis of concepts will be of little avail.

What I am claiming then is that the issue of personal identity cannot be divorced from the issue of psychology. This claim, no doubt, will be met with dismay from various quarters. The problem of personal identity, it will be said, is the problem of what entitles us to apply identity- terms to persons. And this question is to be solved by discovering the criteria we employ in making statements of identity about persons, not by conducting psychological enquiries. The difficulty with this sort of response, however, is that it ignores the fact that the problem at hand is one which deals with the nature of persons. Why, we ought to ask, is there a problem about personal identity in the first place? Why has this particular problem been ear-marked as worthy of a separate enquiry in the way that other problems of identity - e.g. the identity of stones - have not? The answer to this, it seems, is that although persons, like stones, have a physical form, they nevertheless, undergo their existence in a way that

things like stones do not. That is, persons have awareness, memories, beliefs about their own identity, and so on. consequently, any enquiry into the nature of persons - such as an enquiry into whether they have identity over time - will have to come to terms with the way in which persons experience themselves. It will be no good to assure someone that his identity is maintained, for example, by the existence of a persisting self or the continuity of his body if he has no experience of this. For this puts the principle of his identity beyond the realms of his own awareness. But for a theory of personal identity to have any relevance to a person's life it must address itself to the way in which a person is aware of himself, i.e. it must address itself to his existential situation. A theory which makes no attempt to do this is one that has lost its point of contact with the world.

• 4

A perfect example of this type of ignoring experience in favour of applying criteria is displayed by P. F. Strawson.⁽⁷⁾ After claiming that there is no principle of unity or identity which any of us employ to decide whether a certain contemporary experience is ours or someone else's, Strawson goes on to say that he is not denying that someone might have doubts about his own identity. A person might, he admits, be uncertain about whether a particular history was his own. In this case, says Strawson, a person can resolve the doubts he harbours about himself in the same way that others

would resolve similar doubts about him: he would merely apply to himself the ordinary criteria for personal identity. Thus if someone else was unsure about whether it was me he saw at the opera last night he would set about solving this problem by enquiring into where my body was last night, testing the accuracy of the memories I have of last night, and so on. Likewise, according to Strawson, if I have doubts concerning whether I myself was at the opera last night, I merely find out where my body was last night, test the accuracy of my memories, and so on.

But there is something eminently odd about me trying to satisfy my doubts concerning my identity by carrying out such investigations. And what is odd about it is this: if I am unsure of whether it was me who was at the opera last night, then the discovery that my body was in fact there, that my apparent memories of being elsewhere are delusive, and any number of other findings, is hardly going to settle things for me. This is simply because no matter what I discover from the application of these criteria to 'myself', the fact remains that within my experience I lack a sense of identity with the person who was at the opera last night. It is possible, of course, that after contemplating the results of this 'application of criteria' I might eventually come to acquire the 'feeling' that I had been at the opera last night. But it should be clear that such a feeling, whether correct or not,

would be a delusive one; that is, it would not have sprung from my direct awareness of having been there.

The upshot of this is that the theory of personal identity cannot proceed in isolation from how we experience ourselves. Thus our lodestar throughout this study will always be experience itself, and, as a result, our enquiry will be foremost a phenomenological one. This approach, however, does not make the way that lies ahead an easy one. For the more we look to our experience the more we shall see the foundations for the belief in personal identity dissolving before our eyes. One of the first figures in Western thought to appreciate the immense difficulties that beset the notion of personal identity was the Eighteenth-century Scottish philosopher, David Hume. When Hume looked into himself all he could find was some perception or other, never anything which could be called a self. Within the mind, Hume was led to say, there is never any simplicity at one time nor identity at two different times. One of his conclusions, therefore, was that personal identity was a fiction; that is, something which we can only mistakenly attribute to ourselves. We shall return to the ideas of Hume, but for now let us embark on an examination of the view that each of us is in possession of a personal identity.

Those who defend the notion of personal identity tend to

support either a strict non-reductive view: one in which personal identity is something simple and indefinable, or a reductive view, where our identity is analyzable into various empirical features such as memory or bodily continuity. We shall start by having a look at the strict view.

NOTES TO CHAPTER ONE

1. See David Hume, A Treatise of Human Nature, edited by J. A. Selby-Bigge (Oxford, 1978), p. 200.

2. For an illuminating discussion on how identity statements work see Michael Lockwood, 'Identity and Reference', in M. Munitz (editor), Identity and Individuation (New York, 1971), and P. F. Strawson, 'Chisholm on Identity Through Time', in H. E. Kiefer and M. Munitz (editors), Language, Belief, and Metaphysics (Albany, New York, 1970).

3. See 'On Sense and Reference', translated by Max Black in Translations From the Philosophical Writings of Gottlob Frege, edited by Peter Geach and Max Black (Oxford, 1960), p. 57.

4. 'Personal and Impersonal Identity', Mind, XCVII, 1988, pp. 29-49.

5. I should mention here that Sprigge agrees that statements of identity-over-time are genuine statements of identity. It is just that he feels such statements are identity statements in virtue of their referring to the reactualization of the same universal from moment to moment.

6. It must be noted that the appearance of the word 'himself' in this sentence does not beg the question in favour of personal identity. I could have just as easily said that the problem is 'what, if anything, makes a person at one time identical with a person at another time' and thereby omitted the term 'himself'. However, the continual omission of the personal reflexive-pronoun in order to avoid the appearance of begging the question would not only prove cumbersome, but is also philosophically unnecessary. This is because at the level of linguistic convention personal terms (including personal nouns like 'I', 'self', 'one', etc.) are merely grammatically convenient devices that need carry no metaphysical implications. We shall return to this point in Chapter Five.

7. Individuals: An Essay in Descriptive Metaphysics (London, 1971), pp. 133-134.

CHAPTER TWO

THE INDEFINABLE SELF

One of the most famous statements of the strict theory of personal identity comes from the writings of The Bishop of Bristol, Joseph Butler. Although Butler is well-known for his views on personal identity, he is also remembered for a somewhat singular remark that G. E. Moore chose to quote on the fly leaf of his book Principia Ethica. This remark of Butler's is 'Everything is what it is and not another thing'. Although we may doubt whether this is a genuine statement, it nevertheless serves to convey the anti-reductionist attitude that is so much part of Butler's philosophy. This 'statement' is made by Butler in the preface to his ethical sermons and is meant to 'explain' to us why ethical notions cannot be reduced to non-ethical notions. And yet, such a remark would not, as we shall see, be out of place in Butler's discussion of personal identity.

In a brief dissertation appended to The Analogy of Religion (1736) Butler says that although any attempt to define personal identity will end in perplexity, the notion is nevertheless easily understandable. For by reflecting on myself now, and myself of twenty years ago, I can readily perceive that they are not distinct but identical selves. However, says Butler, we should not conclude from this that consciousness of personal identity is the basis of personal identity; for 'one should really think it self-evident that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes'.(1)

The origins of this 'wonderful mistake', Butler suggests, might be that since being conscious and being a person are inseparable notions, one might be led to the conclusion that consciousness is the essence of personal identity. And Butler admits that our consciousness of our present feelings and actions is undoubtedly necessary for our being the person we are now. But he does not allow that our memory of past feelings and actions is a necessary element in our being the same person who once had those feelings or did those actions. My remembrance of myself performing a certain action twenty years ago is not a necessary element in my remaining, twenty years on, one and the same person who performed that action.

This persisting strict identity that informs a person throughout his existence is something which is lacking in plants. It is true that we often talk about a tree as though it were the same tree over the many years of its life, but strictly speaking this is incorrect. 'For when a man swears to the same tree, as having stood fifty years in the same place, he means only the same as to all the purposes of property and common life, and not that the tree has been all that time the same in the strict and philosophical sense of the word' (pp.100-101). This is because he does not know whether any particle of the present tree is identical with any particle of the earlier tree. And if they have no parts in common, then they cannot, strictly speaking be the same tree. Thus Butler concludes the sameness of the tree over time can only be sameness in a 'loose and popular' sense. In the 'strict and philosophical' sense, nothing is the same that changes.

It is in this latter sense that the notion of sameness is applied to a person throughout his life. This is one of the reasons why, for Butler, personal identity cannot consist of consciousness; for consciousness is ever-changing. And if a person's consciousness is at every moment different than it was before, then that person is never, from moment to moment, strictly the same. Still one might reasonably wonder what is wrong with not being strictly the same from moment to moment.

If we accept that consciousness is continuously changing, why cannot we accept that personal identity itself is continuously changing. Butler feels that the best way of refuting this view is simply by making explicit its implications. For on this view, asserts Butler, it must follow that what happened to us yesterday, and what will happen to us tomorrow, is of little concern to our present self: 'for if the self or person of today, and that of tomorrow, are not the same, but only like persons, the person of today is really no more interested in what will befall the person of tomorrow, than in what will befall any other person' (p.102). And the same, it would appear, holds for our present self's concern with our self of yesterday. With this much concluded, Butler thinks it plain that anyone who accepted the implications of the consciousness, or memory, theory of personal identity could only do so 'owing to an inward unfairness, and secret corruption of the heart'.

The first thing one notices about this account of personal identity is its lack of any positive arguments. It is true that, in addition to his mere assurances of the certainty we harbour regarding our own identity, Butler also produces some interesting criticisms of the memory theory. And to be sure, herein lies his fame. But other than this he does little to support his own theory. One reason for this omission is probably because he sees the memory theory and his own view of

identity as being jointly exhaustive. Thus all that is required to establish the strict theory is to demonstrate the inadequacies of the memory theory. Such an assumption is, as can be noted from Hume's rejection of personal identity, premature. Another reason for Butler's non-productivity here seems to be his belief in the obviousness of his own doctrine. So self-evident is the strict theory of personal identity that all Butler need do is to appeal to the certainty we have concerning our own identity. And so, proclaims Butler, 'when any one reflects on a past action of his own, he is just as certain of the person who did that action, namely himself, the person who now reflects upon it, as he is certain that the action was done at all' (p.104). This, however, as a moment's reflection will show, is plainly false. It is a common experience that we may be certain of the occurrence of a past action but not certain of its authorship, even if it was our action (assuming for now that actions do have authors). Thus I can clearly remember a particular story being told over lunch six years ago, but I cannot recall whether it was Mary who told the story to me or I who related it to her. It might be thought that such instances are too minor to offer a serious challenge to our sense of identity. Whether or not this is so, it must be admitted that there are instances in which the uncertainty of one's identity is more than a secret corruption of the heart. And with this admission we have exposed what might be called a fracture in the structure of personal identity.

A possible reply here could be that such occurrences do not present a problem because they are merely errors of memory, not errors of identification. This is the line taken by Sydney Shoemaker in his paper on 'Persons and Their Pasts'.(2) Here Shoemaker says that although it is possible to misidentify other persons, and even in some instances oneself, there remains an important class of first- person statements that are immune to error through misidentification. Thus were I to say, on the basis of a full and accurate memory, 'John shouted X' it is still possible that my statement could be false. For it might be that I have misidentified someone else as John, (either when I first heard the shouting, or later when I subsequently recalled it). That is, although I might remember that the person who shouted X exactly resembled John, it is possible that he was nevertheless someone else. However, were I also to claim on the basis of a full and accurate memory 'I shouted X', here it is not possible that I could be wrong. Shoemaker grants that I could make an error if I happened to misremember the incident, but this would be an error of memory not of misidentification.

Still, not all first-person claims, even if correctly remembered, enjoy this immunity from error. If, for instance, I make a claim about myself on the basis of what I see in a mirror (e.g. 'I went red'), it is conceivable that the person

whom I took for myself in the mirror was in fact someone else, say, my identical twin or double. This would be a case of misidentification relative to the first-person pronoun.(3)

Unfortunately, Shoemaker gives us no precise way of deciding which first-person claims are immune to the error of misidentification and which are not. All we are told is that 'In general, if at some past time I could have known of someone that he was ϕ , and could at the same time have been mistaken in taking that person to be myself, then the subsequent memory-claims I make about the past occasion will be subject to error through misidentification with respect to first-person pronouns'. But, Shoemaker continues, if I could not have been mistaken this way in the past, then my subsequent memory-claims will be immune to the error of misidentification with respect to the first-person pronoun. This could be expressed, says Shoemaker, by stating that 'where the present-tense version of a judgement is immune to error through misidentification relative to the first-person pronouns contained in it, this immunity is preserved in memory'.

Let us start by asking whether it makes sense to say that I can have a full and accurate memory of something that never happened; that is, whether I can veridically remember that John shouted X when in fact he did not. Shoemaker thinks it makes sense because of his alleged distinction between misremembering

and misidentifying. He thinks it is possible that I could have the veridical memory 'John shouted X' when in fact the person that shouted X was not John but only someone whom I took to be John; that is, someone whom I misidentified as John. But if the person who shouted X was not John, then it is patently false to say I have a full and accurate memory of John shouting X. This is because I cannot have a veridical memory which is at the same time a delusive one. Of course I may have a veridical memory that someone who exactly resembled John shouted X, but that is not the same as having a veridical memory that John shouted X. This is evident from the fact that memory is, among other things, a knowledge of the past. And just as I cannot know P if P is false - for one of the conditions of knowing P is P is true - likewise I cannot remember P if P is false. And as it is senseless to say that I can have a full and accurate knowledge that John shouted X even if John did not shout X, so it is senseless to say the same of memory.

Shoemaker has become misled, it seems, because he thinks that the process of identification can be prised off from the process of memory. To see the untenability of this procedure consider my memory 'John offered Jane a piece of cake in the cafe'. Now, let us say that this memory involves the misidentification of not only someone else for John, say Paula, but also the misidentification of Peter for Jane, 'threatening with a revolver' for 'offering a piece of cake', and the bank

for the cafe. On Shoemaker's account it would make good sense to say that I have the full and accurate memory 'John offered Jane a piece of cake in the cafe' even though the memory is fraught with misidentifications and what really happened is 'Paula threatened Peter with a revolver in the bank'. But if this is what really happened, it then becomes absurd to insist that my memory is in any meaningful sense full and accurate. For what is it that I am supposed to be fully and accurately remembering?

One response here might be to claim that although I may have misidentified the various elements of my memory, my memory could still be said to be full and accurate so far as I correctly remember one person doing something to another. But even this is too much; for it still depends on my identifying something as a person, something else as an action, and so on. And all these could turn out to be misidentifications. What this objection needs to be sustained is for my memory of the event to be devoid of any identifications, i.e. for me to remember the event without identifying any of the elements of my memory. But if I do not identify any of the elements in my memory, then I cannot know what it is I am remembering; for to know what it is I am remembering I at least have to identify what I am remembering. And since to remember P is at least to know P, it seems to follow that memory is dependent on the process of identification. We can conclude therefore that it is

false to hold that I can have the full and accurate memory 'John shouted X' if I was mistaken in thinking the person who shouted X was John.

If we now turn to Shoemaker's claim that if I were to say on the basis of a full and accurate memory 'I shouted X', I could not be in error relative to 'I', we see this is true not because of any special status of the first-person, but simply because full and accurate memory-claims cannot be false.

It might be thought that although first-person memory-claims lack the special status that Shoemaker imputes to some of them, an immunity from error through misidentification of 'I' is still to be found in a certain class of first-person present-tense claims. Thus were I to say on the basis of a present experience 'John is shouting X' it is possible that I may have misidentified the person shouting X as John and so have made a false claim. But were I also to say on the basis of a present experience 'I am shouting X' there seems to be no way in which I could make an error of misidentification with respect to 'I'.

Yet I can see no reason why I could not make such an error. Imagine a situation in which other people are shouting and screaming so much that were I to shout I would hardly be able to hear myself. Further, imagine that just before I

attempt to join in the shouting at this noisy gathering I am, unbeknownst to myself, suddenly struck with a type of aphasia which renders vocal production impossible. Now when I attempt to shout X it just could be that someone near me shouts X at exactly the same time. And because of my faulty hearing, or because of the profusion of noise, I may think the 'X' being shouted by someone else is in fact being shouted by me. That is, I could make an error of misidentification relative to the first-person pronoun.

It might still be that there exists some special class of first-person claims that enjoy the sort of immunity to error that Shoemaker is arguing for. All we can say, however, is that Shoemaker has given us no reason to believe in their existence.

Returning now to Butler's criticism of the memory theory, we see it is aimed at John Locke, an early proponent of the view that personal identity consists of the ability to be conscious of, or remember, the past experiences of one's life. Butler's charge, which is one of circularity, is often put in the following way: to have a personal identity through time I must have memories of myself experiencing certain events in the past. But this implies the previous existence of my personal identity. Therefore memory cannot be the basis of personal identity.

This argument is generally accepted as having dealt a fatal blow to the memory theory. It is only fatal however if we are not careful about how we interpret the theory. In his Essay Concerning Human Understanding (1694) what Locke proposes is that if I who am writing now had the same consciousness when I saw the Thames overflow last winter as I did when I saw Noah's flood, then I who am now writing must be the same person as the one who saw the Thames overflow, and as the one who saw Noah's flood.(4) If this begs the question it is because Locke might be interpreted as saying 'I remember myself seeing the Thames overflow' and 'I remember myself seeing Noah's flood'. But there is no need to insert the 'myself' in either case. All that need be said is I who am now writing remember seeing the Thames overflow and remember seeing Noah's flood. It is not necessary that my memory-claims involve any reference to 'I', or, as Butler says, to a consciousness of personal identity.

There are those who would no doubt protest that memory experiences are not had in an impersonal way, and that part of what it is for me to remember witnessing a past event is to remember that it was I who witnessed it. But this is simply dogmatic. For we do have memories in which the self does not appear. I know that many of my memories, especially those of childhood, are of this sort. I can, for example, remember some children playing in a field, but have no memory of myself being

there at the time. To this it could be objected that if I do actually remember children playing in a field, then I must have been there to see it. But whether or not this is true, the fact remains that I have no memory of myself being there.

A defender of Butler, however, would not give in so easily. For even if it is true, he would continue, that our memories are not necessarily remembered in a personal way, there remains the problem of what distinguishes a veridical memory from a delusive one. And this distinction, it will be proclaimed, can only be made by reference to the presence of an I. Thus if I claim to remember some children playing in a field, what renders my memory veridical rather than delusive is that, in the former case I was actually present to witness the event, while in the latter case I was not. And so we see that Butler's charge of circularity - that memory presupposes personal identity - holds after all.

It is curious that this argument has held sway over so many philosophers; for it easily collapses under a little scrutiny. Of course if we were to accept that what makes a memory veridical is that the person who has the memory was actually there to witness the event he claims to remember, then to assert that personal identity is constituted of memory would indeed be fallacious. But why do we need to accept this? Plainly, such an account of the veridicality of memory can do

nothing but support the strict theory of personal identity, and for us to acquiesce to it would be to jump straight into the trap. A less partisan account of what makes a memory veridical is therefore in order.

What makes my memory of some children playing in a field veridical, I propose, is simply that it is causally connected in the proper way with a veridical perception of children playing in a field. What I mean by 'proper way' is something like this: that after the initial occurrence of the veridical perception, the perception is transferred directly to the sensory-buffer, then to short-term memory, and from there to long-term memory from where it is directly retrieved to now be experienced as a memory. This is, it will be allowed, a rough scenario from which there may be various deviations, e.g. some perceptions may bypass short-term memory and travel directly to long-term. The important point is that for a memory to be veridical, the causal chain must run from the veridical perception to the memory experience in accordance with the principles of the psychology of memory: it must not include things like imagining, dreaming, or being told about the perception.

Butler's other complaint with the memory theory is that if it is true, then a person is really no more interested in what happened to himself yesterday, or will happen to himself

tomorrow, than he is in what happens to someone else. What Butler is saying is that since I do care more about both my past and future selves than I do about others, the memory theory must be false. But this is like saying because people behave as if God exists, atheism must be false. Our inclination to embrace or discard a theory does not thereby render that theory true or false. For plainly there are many reasons for being inclined towards or against a particular theory which have nothing to do with its truth or falsity. What Butler should have said is if the memory theory is true, then people ought to be no more concerned about their past or future selves than they are about someone else. And this may well be true, but once put this way it is no longer a criticism of the memory theory. Rather, it is merely an observation about what sort of behaviour would be rational under certain circumstances. It could always be argued that it would be impossible for us to act in a way that placed our self of tomorrow on the same level of concern as another person. But even if this argument were feasible, and I doubt that it is, it would still miss the mark. All it would show is simply that were the memory theory true, we could never live up to it.

Despite the inadequacies of Butler's account, Thomas Reid, a contemporary of Butler's, thought it to be wholly correct. Reid's work on personal identity, differing little from Butler's, is therefore not terribly original, much of what he

says being merely a re-phrasing of his mentor's ideas. Nevertheless, he does give us a more detailed exposition of the strict theory than does Butler, and is therefore worthy of our notice.

For Reid personal identity is something which is simple, indefinable, and admits of no degrees. A person is thus either identical with his earlier self or he is not; he is never more or less the same. Personal identity 'implies the continued existence of that indivisible thing which I call myself'.(5) I am not my thoughts, actions, or feelings, says Reid, for these being merely successive have no continued existence. I am that permanent self to which these thoughts, actions, and feelings belong. The belief in this identity, we are told, is so deeply rooted that it is beyond the possibility of being seriously questioned: 'The conviction which every man has of his own identity, as far back as his remembrance reaches, needs no aid of philosophy to strengthen it; and no philosophy can weaken it, without first producing some degree of insanity' (p.107). It should be noted, however, that like Butler's assurances of our sense of identity, this assertion is not an argument for personal identity. The fact that I may cling tenaciously to a particular conviction in no way demonstrates the truth of my conviction. And indeed, as is commonly known, most deeply held convictions - prejudices most of all - are quite immune to the assails of philosophy.

Further, if Reid's claim about the insanity of disbelief in personal identity is to be philosophically interesting, it must be the case that because someone goes insane upon the loss of a particular belief, that belief must be true. And this of course is false. I may well believe, for example, that the one whom I love also loves me, and upon coming to see the falseness of this belief fall into despair and go insane.

It might be felt that I am perhaps being unfair to Reid and that assertions like the one I am now attacking are to be taken in a rhetorical rather than a straightforward way. I do not think this is so. Reid is after all the philosopher of common sense, and quite often he sees a position's deviation from common sense as being sufficient grounds for its dismissal. In saying this, however, I am not agreeing that the belief in one's own personal identity is a matter of common sense, I am only saying that Reid himself takes such a belief to be common sense. The reason I do not think the belief in a strictly persisting personal identity is one of common sense is simply because I find that surprisingly few people hold it. When I ask the ordinary person if he is the exact same person he was twelve years ago, the most common answer I receive is an unqualified 'No'. Now as a philosopher it might be easy, and perhaps satisfying, to convince myself that my interlocuter has not fully grasped the depth of my question. But even after I have clearly explained that what I am asking is not whether his

ideas, feelings, or plans are the same, but rather, whether the person to which these belong is the same, I still frequently receive a 'No'. I admit, of course, that there are those who upon hearing the fuller version of the question see me as asking about what they call their 'Inner Being' or 'True Self' and reply that yes this is something that has never changed. But the person who gives this sort of answer is certainly no more common than the one who quite firmly stands behind his claim to be essentially a different person than he was twelve years ago. None of this is to say that the ordinary person believes himself to be lacking in selfhood. Most people, it would seem, do seem to believe themselves to be in possession of some sort of a self. However, from my various discussions I have come to see that there is no one particular conception of the self which is commonly held. All the reader need do to confirm this for himself is merely ask those around him. He will quickly see that while for one person the self is the personality, for another it is the brain, while yet for another it is a glowing mass of energy that somehow surrounds the body. Just why there is such a diversity of views here is difficult to say. One reason however might be that, as Andrew Brennan has suggested, we simply have no clear conception of what a person is. As a result, each person is free to construct a notion of himself in the way he chooses. This has the interesting effect of even allowing us to entertain quite inconsistent views about ourselves, leaving us, as Brennan puts it, at the mercy of

somewhat inconsistent expectations.(6)

Fortunately, however, we need not linger on this issue, for Reid does supply us with an argument for the non-reductive view of personal identity. Asking himself why he thinks there is a permanent I, Reid gives the answer that it is remembrance which is the proper evidence. Thus referring to a dialogue of twenty years ago, he says 'I remember several things that passed in that conversation; my memory testifies not only that this was done, but that it was done by me who now remember it. If it was done by me I must have existed at that time, and continued to exist from that time to the present' (p.110). It is important to see here that Reid is not claiming memory to be the basis of personal identity, for this would be the Lockean thesis. Rather, he is only holding memory to be the evidence for personal identity.

At first glance it might seem that Reid's argument is a plausible line to take. But upon looking further some intolerable difficulties become apparent. First, there is the question of how it is that my memory testifies that a certain thing was done by me. As Reid himself says, I am not thought, action or feeling. But what is it then that I remember when I remember that something was done by me? Is it the permanent and indivisible self that I remember? Certainly such a self must come into the memory in some way, or how

could I remember it was done by myself. And yet for my self to enter into my memory it must enter in some form. But as just mentioned it cannot take the form of thoughts, feelings, or actions. So what is it I am supposed to be remembering when I remember myself? Reid's only reply here would seem to be that such a memory cannot be analyzed any further due to the simplicity of the self - I just remember myself doing something and that is the end of it. But if this is the nature of my remembrance of myself, then remembrance cannot, without begging the question, be used as evidence for the existence of the self. Far from providing evidence for a strictly identical self, it already assumes such a thing.

A second problem with Reid's argument is that even if we waive these difficulties and accept that I do in fact have a veridical memory of myself having a conversation twenty years ago, it in no way follows that I must have existed from then till now. For it is possible that after my conversation I ceased to exist for twenty years, and only now have started to exist again. Indeed, I may be something whose very nature is to slip in and out of existence at various times.

Reid's reply here is that the concept of identity presupposes an uninterrupted continuance of existence. To allow otherwise, says Reid, would be to suppose that something could exist after it has ceased to exist, and also that

something could exist before it came into existence, which are 'manifest contradictions'. I do not feel Reid is being fair here. It is certainly true that once something has ceased to exist forever, it cannot start to exist again. But I see nothing in the notion of ceasing to exist that implies that the cessation must be forever. If such an implication was carried in the notion of ceasing to exist, then it would be a contradiction of terms to say that something has ceased to exist for twenty years. I do not see that it is a contradiction. Indeed, I can clearly imagine a state of affairs in which someone suddenly vanishes from the face of the Earth only to reappear sometime later with no awareness of what transpired in the interim. For us to approach such a person and assure him that because of his one-time cessation of existence he can no longer be the same person, seems little more than a metaphysical prejudice. Likewise, to take the other route and assume that because he is the same person he must have somehow existed during his absence, is equally unfounded.

Despite the failure of Butler and Reid, there are nevertheless some philosophers who uphold the idea of strict personal identity. One such philosopher is John McTaggart.(7) Although McTaggart has similar views to Butler and Reid on personal identity, his argument for his position is quite different. Borrowing Russell's distinction between knowledge by acquaintance and knowledge by description, McTaggart argues

that because I can judge that I am aware of certain things, and because I can know the meaning of judgements containing the word I, I must know I. And because I cannot be known by description it must be known by acquaintance, that is, I must be directly aware of it. McTaggart's reason for thinking I cannot be known by description is simply that no matter how extensively I describe any fact about myself I am never entitled to conclude that the description refers to me. For if I do not know myself by acquaintance, the only thing I can know about the person who makes the judgement is just 'the person who makes the judgement', I can never know that the person who makes the judgement is me.

This is an interesting argument and it appears to arrive at the same conclusion as Butler and Reid while avoiding their difficulties. Nevertheless, the argument is still confused. The first thing to notice is that there are at least two interpretations of the conclusion. In concluding that I must know I by direct awareness, does McTaggart mean I must at every moment of awareness be aware of I, or does he mean only that at least sometimes I am directly aware of I? If he means the former, then he is standing on dubious ground; for there are many instances of awareness which do not contain the awareness of I. Take, for example, my awareness in the activity of reading. It is often the case in reading that one is engrossed enough in the story to be quite unaware of oneself. And reading

is hardly an exceptional case. Other activities, such as day-dreaming, playing chess, and making love, all provide instances in which one can go on being thoroughly unaware of oneself. If, on the other hand, McTaggart means only that I am at least sometimes aware of I, then he has allowed that self-awareness is not a necessary component of awareness. But now there arises the question of what it is that entitles me to say an experience, which does not involve a direct awareness of I, belongs to me. I cannot know by description that such an experience is mine: for as McTaggart has shown, no matter how extensively I describe the experience, I will never be able to conclude that it is mine. And yet I cannot know by direct acquaintance that it is mine; for the experience in question involves no direct acquaintance with I. McTaggart could always reply that none of this affects his argument because he is only referring to those moments of awareness in which I do judge that I am aware. However, from the fact that one need not be aware of oneself in order to be aware, I shall eventually be drawing some far-reaching conclusions. I shall, for example, argue that the self, not being a necessary element in experience, is something which experience itself has fabricated. This fabricated self therefore has no actual identity and is merely a collation of transient images. It seems that McTaggart himself is cognizant of the possibility of the self being a fabrication. This is intimated at the outset of his argument where he says that the establishment of our

perception of the self would not alone prove selves exist, but only that something was perceived as being a self. This is, however, to anticipate. For now we will continue in our appraisal of McTaggart's argument.

The error that has been exposed in McTaggart's reasoning is one which is easy to overlook. This is because the way the argument is presented makes it seem as though the fundamental issue is whether we know the self by acquaintance or by description, and not whether we know the self at all. And so McTaggart spends page after page showing why the self cannot be known by description, but only a line or two in arguing that the self is known in the first place. This is, I think, merely symptomatic of depth of McTaggart's belief in of personal identity. When one is under the spell of such a conviction, the possibility of the non-existence of the self is barely raised. Even McTaggart's commentators do not seem to have noticed this shortcoming. Thus Geoffrey Madell, a recent exponent of the non-reductive view, says of McTaggart's argument, 'To put it another way, since no description of any body or experience can entail, or otherwise support the claim, that it is mine, the property of being mine is unanalysable'.(8) Clearly, there is an important premise missing here. For this argument can only go through if we assume there is a property of being mine: the claim that no description of any body or experience can demonstrate that what

is being described is mine is perfectly compatible with the view that there is no self, i.e. there is no property of being mine. Indeed, it is just what one should expect in a world where selves do not exist.

A more general complaint that can be lodged against McTaggart is that although he is hoping to establish an empirical conclusion, i.e. that we are directly aware of the self, his method of reasoning is a priori. Thus, rather than turning to experience itself to see if and how we know I, McTaggart argues that we must know I because we know the meanings of certain judgements, and further that because we cannot know I in one way we must know it in another. This sort of a priori arguing about what experience must be like seems somewhat dubious to say the least. If we were directly aware of I, it seems hardly likely that we would need to deduce this awareness from some purported fact about our knowledge of meanings. One wonders why McTaggart does not simply direct our attention to this perception of the self which he claims we all have. If such an awareness did exist then plainly there would be no difficulty in us knowing immediately what McTaggart was referring to and acknowledging his point. And yet what we find is McTaggart trying to convince us of a direct awareness by an indirect argument. If we remind ourselves here of Hume's attending to his own experience only to discover the complete lack of anything that could be dubbed the self, then it becomes all the more

understandable that McTaggart does not appeal to experience to establish his conclusion. And yet his failure to do so can only raise our suspicions. For the final test of whether or not I am directly aware of my self will always be an examination of my own awareness, and if such an examination turns up negative, then no amount of a priori arguing will prove otherwise.

There have, however, been recent attempts to argue for a persisting self by referring directly to awareness. Thus to the question of how we acquire the concept of a self, Roderick Chisholm answers that we acquire it in the same way we acquire the concept of sadness: by having it. This is because both sadness and being a self are 'self-presenting properties'. A self-presenting property, we are told, is one which is such that, 'if while having it, you consider your having it, then you will believe yourself to have it'.(9) But this only explains how once having a self we come to believe we have it, namely by considering it. It says nothing about the self itself or how we have it (which is why, of course, Chisholm underlines 'consider' and 'believe', but not 'having'). Chisholm seems to think he can tighten-up his case for the existence of the self by establishing a relationship between self-presentation and certainty. He does this by asserting that if certainty is the 'highest degree of epistemic justification', then a person's self-presenting properties are objects of certainty for that person. But this amounts to

little more than a needlessly verbose way of saying a person is certain of his self (which is, unhelpfully, a mere repeat of Butler's assurance). And so we are back at the beginning. For even though I now know I am supposed to have the highest degree of epistemic justification of my self-presenting property of having a self, I still do not know what this self is, what the nature of my having it amounts to, or where it exists in my awareness.

From here Chisholm moves on to the problem of the unity of consciousness. Many philosophers have thought that the unity of consciousness is something beyond dispute. Thus the problem was seen to be how consciousness, with all its perceptions, thoughts, and feelings, is unified. But even to try to answer this question is already to cast one's vote in favour of the unity of consciousness. Chisholm is an example of someone who does not seriously question the doctrine of the unity of consciousness. In response to the question 'Is it possible that what I call my experience has several subjects?' (i.e., one subject who hears, one who sees, and so on), Chisholm quotes with approval a passage from Brentano - a passage which makes the bare assertion that when a person is aware of seeing and aware of hearing, he is also aware that he is doing both at the same time, and then goes on to conclude that the two perceptions 'must belong to the same real unity.' Chisholm then asks if we could settle for less. Could it be, for example, that a person who is aware that he sees something and aware

that he hears something, is aware of his seeing and hearing as being 'parts of the same consciousness' or perhaps as being 'compresent in consciousness'? Chisholm's defense here is merely 'I think not. What could it mean to say they are 'parts of the same consciousness' or 'co-present in consciousness, other than that the same person is aware of both?' (10)

It should be clear that this is not acceptable as an argument. All Chisholm has done is to quote the assertion that consciousness must be a unity, and then make the unsupported claim that to describe two perceptions of a person's consciousness as being 'parts of the same consciousness' or 'co-present in consciousness' could mean nothing other than 'the same person is aware of both'. In all of this Chisholm has left untouched the most basic contention: Is consciousness a unity? Lest it be thought that Chisholm's remarks do constitute a viable argument for the unity of consciousness, it is perhaps worthwhile to point out some confusions they contain. We can start by examining Brentano's assertion that when a person is aware of seeing and aware of hearing, he is also aware that he is doing both at the same time. The apparent reasonableness of this claim dissolves as soon as we turn to specific instances of awareness.

Now although Brentano is referring to perceptions of a second order, that is, awareness of seeing and awareness of hearing, rather than simply seeing and hearing, it will

nevertheless be useful for us to start by examining perceptions of the latter sort. Let us say I am sitting at the beach watching the waves break on the shore. It is true that I am seeing the waves, and it is true that I am hearing them. I am not, however, performing any complex cognitive activity. I am not, for instance, trying to be aware of my observing the waves. It is therefore false that in seeing and hearing the waves I am aware that I am both seeing and hearing at the same time. To be so aware I would have to make the conscious move to enter a reflective state of awareness: I would have to start being aware of my process of observing the waves. To make this point in another way, let us consider a pigeon who is also aware of seeing the waves, and aware of hearing them. Because this animal lacks the ability to enter reflective states of awareness, it would be wrong to maintain that it is or could be aware that it is both seeing and hearing the waves at the same time. But now let us say I suddenly become aware of seeing the waves and aware of hearing them. According to Brentano, I must now be aware that both these perceptions are being had by the same person. But how is it the the notion of 'the same person', or of any person for that matter, suddenly sneaks in here? All that has changed is now I am aware of seeing and aware of hearing rather than merely seeing and hearing. In order for me to be aware that it is me who is both aware of seeing and hearing, I must first have an awareness that there is a me, that is, a self who is doing both the seeing and hearing. But I

have no such awareness. And Brentano, at least in the passage upon which Chisholm relies, has no argument to offer in favour of this awareness. Rather, it is quite clear that the existence of self-awareness, and so of a self, is assumed throughout.

It is pertinent to note that even if the awareness of different perceptions somehow guaranteed the awareness that it was the same person who was having these perceptions, we would still be faced with those acts of perception (which are by far the vast majority) which are not themselves objects of awareness. What is there to unify consciousness here? One could always reply that it is the possibility of becoming aware of the different acts of perception that bestows a unity on consciousness. But this would not do for various reasons. First, it implies that where such a possibility is absent, so is the unity of consciousness. By this rule, then, pigeons, dogs, and horses - any creatures who do not have the ability to reflect on their own states of awareness - all fail to pass the test. But this should arouse our suspicions. For if a dog has no unity of consciousness, it certainly does not suffer from its handicap. It sees, hears, and goes about its life quite happily. It even performs complex goal-directed behaviours which depend on, among other things, the ability to both see and hear at the same time. Of course it cannot discuss the question of its own identity, but I doubt this can be blamed on a disunified consciousness.

Secondly, we are bound to ask 'What is it that determines whether or not it is possible for an act of awareness to become an object for my awareness?'. For the reply now under examination is plainly dependent on the tacit supposition that only certain acts of awareness - namely, those which are called mine - could ever become objects of my own awareness. But why could not those acts of awareness which are called yours become objects of my awareness? Why could I not suddenly start perceiving your perceptions? To say that they would then be mine is unhelpful because it makes the issue of what is mine and what is yours into a logical point: any perception of which I am aware is, by definition, mine. In doing so, the claim that 'it is the possibility of my becoming aware of different acts of perception that makes them mine' becomes a mere stipulation about how the word 'mine' is to be used, and not, as it ought to be, a statement about the nature of the unity of consciousness.

There is yet a further confusion in this reply, a confusion which takes us straight to the heart of what is wrong with the doctrine of the unity of consciousness. The problem here arises from the illegitimate appearance of the word 'my'. I cannot say that what makes my perceptions unified in consciousness is the possibility of my becoming aware of them, without having severely begged the question. For by calling them mine I have already taken the liberty of assuming them to

be unified in consciousness. This is why Kant's theory of the original synthetic unity of apperception must ultimately fail. If we turn to the section of this title in The Critique of Pure Reason, we find Kant saying 'It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying would be impossible, or at least would be nothing to me'.(11) The problem here is that even if it were not possible for the 'I think' to accompany all my representations, they would still be something to me as they would still be, as Kant himself calls them, my representations. To make this point another way, if it is the possibility of the 'I think' accompanying all my representations which makes them mine, then 'my representations' just means 'those representations which it is possible for the 'I think' to accompany'. Kant's claim is then reduced to 'It must be possible for the 'I think' to accompany all those representations which it is possible for the 'I think' to accompany', which is manifestly, but uselessly, true.

It is tempting to think that this problem could be readily solved by removing the word 'my', which Kant has, inadvertantly no doubt, slipped into his proposition. But no sooner has this alteration been made than we see the futility in such a tactic. For in omitting to mention that the representations being referred to are all my representations, the proposition now refers to all representations indiscriminately and so fails to



provide a criterion for singling out, or, as we may also say, unifying, one consciousness amongst others.

In addition to this internal confusion there are also other considerations, such as the possibility of the splitting of consciousness and certain mental 'disorders', which bar our acceptance of the doctrine of the unity of consciousness. However, so as not to lead us astray from our present purposes, their examination will be postponed for later chapters.

We can now return to Chisholm's defense of Brentano. Chisholm's reason for accepting Brentano's thesis is that the attempt to describe the different perceptions of consciousness as being, say, 'parts of' or 'co-present in' consciousness, must fail because such descriptions can only mean that it is the same person who is aware of the different perceptions. However, if one takes a moment to examine the descriptions in question, it will readily be seen that they do not have the same meaning. The description 'these two perceptions are parts of the same consciousness' does not mean 'the same person is aware of both' simply because 'consciousness' and 'person' do not mean the same thing. Consciousness may be an aspect of persons, but not all things that have consciousness are persons. Pigeons have consciousness, but are not persons.

Further, this cannot be remedied by inserting 'human' in front of 'consciousness'. For it is allowed that persons may lack consciousness, such as when they are deeply asleep or in a

coma. But it does seem strange, if not contradictory, to say that consciousness, even human consciousness, may lack consciousness. One way for Chisholm to escape this difficulty, it might seem, would be to exchange the term 'person' in 'the same person is aware of both' for the term 'consciousness'. His claim would then run: the awareness of seeing and the awareness of hearing are 'parts of the same consciousness' or 'co-present in consciousness' means the same as 'the same consciousness is aware of both'. Although this gets round the difficulty of the person/consciousness distinction, the identity of meaning Chisholm desires would still not be achieved. If I am aware of seeing at the same time that you are aware of hearing, then my awareness of seeing and your awareness of hearing are co-present. And, since both my awareness of seeing and your awareness of hearing take place in consciousness, it may rightly be said that both our instances of awareness are 'co-present in consciousness'; which is not at all to say that 'the same consciousness is aware of both'. This problem does not occur with the first phrase suggested by Chisholm, i.e. 'parts of the same consciousness', because here he has mindfully inserted 'the same' before 'consciousness'. If he had done likewise with the second phrase and had said 'co-present in the same consciousness', then the charge we are now considering could not be brought against him. Unfortunately, however, had Chisholm said 'co-present in the same consciousness', he would have exposed himself to a further charge, one which he

already must face for having used 'the same' in 'parts of the same consciousness'. This is the charge of circularity. For in saying that what is co-present in consciousness is co-present in the same consciousness, or that the parts of consciousness are parts of the same consciousness, Chisholm has already begged the question in favour of the unity of consciousness, i.e. he has allowed that there is such a thing as a consciousness which, being the same, has a unity. With this we can now see that Chisholm's purported lesser claims are not lesser claims at all. Of course, the words 'parts of' and 'co-present in' have the appearance of describing a disunity, but once it is granted that they are describing things which are of or in one (i.e. the same) consciousness, then, because consciousness is awareness, it immediately follows that the same consciousness is aware of both.

Luckily, the discussion does not stop here but moves on to the nature of self-consciousness. According to Chisholm, there is a sense of self-consciousness which means simply attributing a property to oneself, or self-presentation. But the further and supposedly more important sense, is where one must 'know and believe that he, himself, is the one to whom he attributes such properties'; he must recognize these attributes 'as his own'.(12) This later form of self-consciousness, says Chisholm, is often referred to in literature as the discovery of 'I am me'. As an example of this discovery he then quotes

the following passage from the German romantic Jean Paul Richter:

I shall never forget what I have never revealed to anyone, the phenomenon which accompanied the birth of my consciousness of self [Selbstbewusstsein] and of which I can specify both the place and time. One morning, as a very young child, I was standing in front of our door and was looking over to the wood pile on the left, when suddenly the inner vision 'I am a me' [Ich bin ein Ich] shot down before me like a flash of lightning from the sky, and ever since it has remained with me luminously: at that moment my ego [I] had seen itself for the first time, and for ever. One can hardly conceive of deceptions of memory in this case, since no one else's reporting could mix additions with such an occurrence, which happened merely in the curtained holy of holies of man and whose novelty alone had lent permanence to such everyday concomitants. (13)

This discovery, says Chisholm, is the discovery that is made when one first realizes that all of one's properties belong to the same thing; that is, when one becomes aware of the unity of consciousness. How does one realize this? Chisholm answers that one only has to consider it to see it is true.

Yet the passage in question is hardly a mere account of someone's recognition that all his attributes are his own. This is clear from Richter's lack of reference to anything about attributes belonging to a self. Secondly, the episode is described as something which was never revealed, a sudden inner vision of 'I am an I' (14) which 'shot down before me like a flash of lightning from the sky', something which happened in the 'curtained holy of holies of man'. There is plainly

something of a mystical experience going on here, something which Chisholm chooses to ignore. Thirdly, Chisholm says that one comes to see one's own unity of consciousness by simply considering it. But there is nothing in the passage which suggests that the young Richter is considering the unity of consciousness at the moment of his vision; rather, he is simply standing in front of his door looking at a pile of wood. We may safely conclude therefore that Richter's childhood experience is not, as Chisholm would have it, an instance of someone recognizing all his properties as his own.

This, however, is not the end of the theory of strict personal identity; for returning to the work of Madell we find that there are other arguments to come. According to this philosopher the case for a strict and unanalyzable account of the self is powerfully supported by an investigation of the imagination. He argues that since I can imagine a world in which I possess totally different objective features than I now have, and since I can imagine a world in which, rather than myself existing, someone exactly similar to me exists instead, my identity cannot be analyzed in terms of any objective features of my existence, and therefore my identity must be strict and unanalyzable.

Remarkably enough, the considerations which Madell here adduces in support of his theory are the same ones used by Bernard Williams to demonstrate the incoherence of that theory.

In his 'Imagination and the Self' (15), Williams draws our attention to the thought that 'I might have been somebody else'. Now although this thought may intuitively seem to make sense, once it is examined some important difficulties emerge. If we turn, for instance, to the idea that I might have been Napoleon, what we are imagining, says Williams, is that there could have been a world exactly like this one except that I would have been Napoleon. But in order to imagine that I could have been Napoleon, I cannot imagine that I, with all my properties and characteristics, could have been Napoleon; for to be Napoleon I must have all his properties and characteristics. Therefore the I of this imaginative exercise can only be some sort of 'attenuated' or propertyless self which I imagine to exist in the place of Napoleon's self. The problem here, says Williams, is that there would then be no real difference between this imagined world and the present one. The imagined Napoleon endued with my self would be exactly the same as the actual Napoleon.

Madell's reply to this is that, contrary to Williams, there is in fact a difference between the two worlds; for 'in the real world I possess a certain set of properties, while in the imagined world I possess quite a different set'. (16) Madell admits that this reply would be unacceptable if the notion of a propertyless Cartesian self did not make sense. But, he is quick to add, Williams has not demonstrated this.

Rather, in arguing from there being no real difference between the imagined and the actual world to the incoherence of a propertyless self, Williams has begged the question. That is, Madell would say, in asserting there to be no real difference between the two worlds, Williams has already assumed the impossibility of there being a propertyless self.

I do not think this is a correct assessment of the argument. For Williams has surely given us good reason to dismiss the notion of a propertyless I; namely that when I try to imagine myself being Napoleon, I am confronted with the discovery that there is nothing I can bring to this imaginary world that would be distinguishable from what is already there. There is no question begging here. Madell's remonstrance that there is an intelligible difference between the two worlds; that in one world I would have one set of characteristics, while in the other world I would have another, is most unilluminating. For what we want to know is just what this difference amounts to. Madell's assertion does not explain the difference, rather it is something which needs to be explained.

A further problem with Madell's argument is that, like his restatement of McTaggart's position, it also ignores the fundamental question: it assumes that I know what it is like to have a self in the first place. How could I possibly imagine myself as possessing a variety of different features if I do not already know or have some idea of my self? Similarly, to

be able to imagine a world in which rather than myself existing, someone very much like me exists, is to already be employing the concept of selves. I must have some awareness of my self before I can start to imagine that self in various circumstances. To this Madell might reply that the fact that I can imagine myself in different possible worlds demonstrates that I do have an awareness of my self. My rejoinder here is simply that if you cannot show that we have an awareness of the self in the present world - as the failure of both McTaggart's and Chisholm's arguments demonstrates - turning to the imagination will be of little avail. I admit quite happily that it is possible to imagine a world very different from this one. I can imagine, for instance, the existence of thoughts, feelings, and actions which do not now exist. And, equally, I can imagine the existence of thoughts, feelings, and actions which are very similar to those that now exist, and, further, I can imagine them existing instead of those which exist now. But as Madell would no doubt agree, this hardly supports the strict theory of personal identity. And yet, I fear, this is all the imagination can show us.

Another reason for holding that personal identity is unanalyzable is, according to Madell, because only on such an account can we make sense of our concern for our future selves. If we try to see personal identity as being composed of, say, memory or psychological continuity, then we shall be at a loss to explain why we should care about our future selves

in the way we do. This is of course just Butler's allegation against Locke, an allegation we have shown to be without substance. Madell, however, seems to think a case can be made for the strict theory of personal identity by turning to the example of the fear of future pain. If I fear that I shall suffer future pain, says Madell, what I fear is not that a person with a certain psychological continuity will undergo pain, but simply that the person suffering will be me. Madell can see no reason why he ought to particularly care about the fact that the one who is to be suffering will be the one with just these memories and interests. To the response that I ought to care simply because to have these memories and interests is just to be me, Madell replies that he is left with the feeling that no explanation has been offered.

There are some problems with this argument. First, It is plainly false that the only way the fear of future pain can be made sense of is by appealing to the strict theory of personal identity. There is a multitude of other accounts, e.g. socialization, conditioning, modelling, and so on, which can, without making any reference to the strict theory of personal identity, also explain this fear. Pigeons obviously fear future pain, and will go into quite a panic if it is apparent that they are to undergo pain. Madell would not, presumably, feel the need to invoke the strict theory of personal (or even pigeonal) identity here.

Secondly, the premise of this argument in no way supports the conclusion. I readily agree with Madell that there is nothing in a person's having thus and so psychological continuity that implies I should have a particular concern about the pain that person will have. The reason I agree, however, is because there is nothing in anything that implies I should fear pain other than in pain itself. This is because pain is an intrinsic evil. To say that what particularly disturbs me about future pain is that it will be mine, is to try to explain the the evilness of pain extrinsically. It is to think that the evilness of pain resides not in the pain itself, but rather in the fact that it is me who will be having it. One reply that Madell might be tempted to make to this is that the notion of a pain which does not belong to anyone does not make sense, and therefore that pain is an intrinsic evil only when it is had by someone. But he could only make this reply by overlooking what we established in our critique of McTaggart, namely, that the awareness of I is not a necessary element in awareness, and, consequently, that experiences do exist which are not had by an I. Pain, especially extreme pain, is, I think, a good example of a state of awareness which does not involve the awareness that it is I who am having the pain. What one experiences in extreme pain is its sheer terribleness. In such a state the various aspects of one's mental life are wiped away and there is nothing in one's awareness save the pure agony of the moment. In the instance of extreme pain there is

no room for the reflective and complicated awareness that it is this I to whom the pain belongs. This view that pain (or any experience for that matter) does not belong to anyone is one which was recognized by Buddhist philosophers hundreds of years ago. Thus in the Visuddhimagga Sutra we encounter the line 'Suffering alone exists, none who suffer'.(17)

We have now come to the end of our examination of the strict non-reductive theory of personal identity, a theory which, in light of the above considerations, we must reject. This does not allow us, however, to dispense with the notion of personal identity all together; for it is still possible that an account of personal identity can be formulated in terms of some non-strict criterion, such as psychological or bodily continuity. It is to these reductive theories that we shall now turn.

NOTES TO CHAPTER TWO

1. 'Of Personal Identity', in Personal Identity, edited by John Perry (Berkeley, California, 1975), p.100. Further references are included in the text.

2. in Identity, Cause, and Mind: Philosophical Essays, (Cambridge, 1984), Essay II.

3. The claim that there is a class of first-person statements that are immune to error through misidentification of 'I', which was first made by Shoemaker in 'Self-reference and self-awareness' (1968), is qualified considerably in his later paper 'Persons and their pasts' (1970). This is because of his recognition of the possibility of 'branching' or the fission of one person into two. We will explore this idea in Chapter Three.

4. 'Of Identity and Diversity' in Personal Identity, edited by John Perry (Berkeley, California, 1975). Further references to this edition are included in the text.

5. 'Of Identity', in Personal Identity, edited by John Perry (Berkeley, California, 1975), p.109. Further references are included in the text.

6. Conditions of Identity: A Study of Identity and Survival (Oxford, 1988), p.309. See especially Chapter 10.

7. The Nature of Existence, (Cambridge, 1927), II, Chap.XXXVI.

8. Geoffrey Madell, The Identity of the Self, (Edinburgh, 1981), p.25.

9. Roderick Chisholm, The First Person: An Essay on Reference and Intentionality (Brighton, 1981), p.81. Strictly speaking, although Chisholm claims that one is directly acquainted with a self, he does not claim that being a self is a self-presenting property. The claim is that self-presenting properties (sadness, for example) pertain to a self in the sense that the fact that they obtain implies there is a self which has them. Because this is so, claims Chisholm, we can say that in being, for example, sad, one is directly acquainted with oneself.

10. It should also be pointed out that when Chisholm considers the weaker claims he changes Brentano's phrase from 'aware of seeing and hearing' to 'seeing and hearing something'. That Chisholm has exchanged these phrases inadvertently seems to follow from his not commenting on the

distinction between the two, and his use of the latter as is if meant the former. There is, of course, a crucial distinction to be drawn here: to see something is to apprehend something visually, while to be aware of seeing is to direct one's attention to the process by which it is so apprehended.

11. Immanuel Kant, Immanuel Kant's Critique of Pure Reason, translated by Norman Kemp Smith (London, 1950), pp.152-153.

12. Chisholm, p.90.

13. Chisholm, p.90.

14. The correct translation of Ich bin ein Ich is 'I am an I', not, as it is given here, 'I am a me'. The German for 'I am a me' would be Ich bin ein mich. Although this might seem a trivial point, I do not think this is so. Because 'me' is in the accusative while 'I' is in the nominative, the statement 'I am a me' seems to carry with it some sort of factual connotation. That is, it seems to be implying that I am something which can be acted upon. The statement 'I am an I', on the other hand, carries no such implication. Consequently, it is a more trivial claim.

15. Bernard Williams, Problems of the Self (Cambridge, 1973), Chapter 3.

16. Madell, p.97.

17. Bhadantacariya Buddaghosa, The Path of Purification (Vissuddhimagga), translated by Bhikkhu Nanamoli (Columbo, Ceylon, 1956), p.587. In this and other references to Pali and Sanskrit works I have omitted diacritical notations.

CHAPTER THREE

THE PSYCHOLOGICAL THEORY

In the last chapter we had occasion to touch on one version of reductionism; namely, the memory theory of personal identity as propounded by Locke. There we were able to show that Butler's charge of circularity was without foundation. Once we are free from this ill-starred criticism we find that the memory theory makes some definite advances on the strict theory of identity.

A major feature of the memory theory, for example, is because it defines the self in terms of memory it immediately disposes of the problem encountered throughout the last chapter; that is, the problem of locating the self within experience. For according to this theory it is experience which is located in the self, not the other way around. In other words, the self is something which is constructed out of experience. Consequently, looking for the self within

experience would be like looking for the gaggle within the goose.

A corollary of this is that we also elude Reid's problem of having to explain how I can remember that a certain action was done by me without referring to any particular experience. This is an intractable assignment for Reid because, as we have seen, there is nothing to remember but some experience or other. On the memory theory, however, it is just the ability to correctly remember an action that makes me able to remember it was done by me.

Although we may applaud these advances of the memory theory, there are nevertheless some serious difficulties which require examination. Before we look into these, let us return to Locke to get a fuller picture of what this theory involves.

Locke starts his account by distinguishing three different types of identity. These are the identities of substances, biological organisms, and persons. He points out that the notion of identity must be suited to the thing to which it is intended to apply. Thus what counts as identity with regard to substances will differ from the sort of identity that applies to biological organisms, and also from the identity that applies to persons. Consequently, Locke would reject Butler's claim that because the particles of a tree are continuously

changing, a tree can be said to have identity through time only in a loose and popular sense. Butler, Locke would say, has mistakenly applied the notion of substance identity to something that is not a substance, i.e. to a biological organism.

For while, in the case of material substance, identity depends on the unalterable persistence of particles of matter, in the case of plants it will depend on 'the organization of parts in one coherent body' which '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'.(1)

It is this same organized parts of a living body that defines the identity of the human organism. Both this 'identity of man', as Locke calls it, and the identity of the soul (i.e. finite immaterial substance) should be carefully distinguished from the identity of the person. Because every soul began to exist at a particular time and place, the identity of the soul will be determined by its continued relation to that time and place. The identity of a person, however, is to be decided differently. For a person, according to Locke, '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'(p.39). The word

'thinking' is for Locke a generic term referring to such mental functions as perception, sensation, cognition, and volition; all of which, it is claimed, are inseparable from consciousness. Since consciousness is the single connecting thread which runs throughout our mental life, it is in consciousness alone that our personal identity is to be found. Consequently 'as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person' (p.39).

Although it is sometimes unclear exactly what Locke means by 'consciousness', when he is discussing personal identity over time, as in the above quotation, it seems evident that he is equating the term with memory. For consciousness extended backwards to a past thought is just my memory of a past thought. He also speaks of consciousness as being an awareness of present experiences, e.g. pleasure and pain. In these instances the term 'consciousness' does not refer directly to memory but rather to the experiences which will, should their occurrence be recalled sometime in the future, become memories. Locke concludes, therefore, that it is by this consciousness, or memory, that 'everyone 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' (p.39).

That the identity of a substantial soul has nothing to do with one's personal identity can be seen to follow from Locke's contention that substance is 'void of consciousness'. It is true that Locke does sometimes speak of 'thinking substances' or of a substance as being that which 'thinks in' the person, but his general position makes it clear that thinking itself is not a substance, and that substance is only meant to function as some sort of metaphysical basis for thinking. And if this is the nature of substance, says Locke, then there is no apparent absurdity in the suggestion that the substance which supports my consciousness is in fact the same substance which at sometime in the past supported the consciousness of some other person. This however would be of no consequence to me as a person. What is of consequence to me as a person is just my consciousness and memories. And consciousness and memories are indifferent to the particular substance which supports them. Thus my being informed with the substance of another person would no more make me that person than would my having some particles of that person's body in mine.

The reason why personal identity and the identity of man are not to be confused is given in Locke's imagined case in which the soul of a prince, bringing with it the memories of the prince's past life, transmigrates into the freshly de-souled body of a cobbler. In such a case, says Locke, it is clear that the resulting person would be the prince, and so be

accountable for the prince's past actions. The resulting man, however, would still be the cobbler.

Despite Locke's construction of the example, it should be clear that the transmigration of the prince's soul - which is said to bring his consciousness with it - is irrelevant for the person of the prince to appear in the cobbler's body. For as we have just seen, it is consciousness alone and not consciousness plus the soul that makes the person. Locke's introduction of the soul into his story is possibly the result of a tendency to believe that if transmigration is possible, it could not be consciousness alone that transmigrates; that is, it could not be the prince's consciousness which enters the cobbler's soul and body. Nevertheless, as we have seen, for Locke there is no necessary connection between a particular soul and the consciousness it supports. He therefore could have just as easily supposed that it is the prince's consciousness which transfers to the (unconscious) soul and body of the cobbler. The resulting person would still, for Locke, be the prince.

One of the most popular objections to this account of personal identity is that people forget things. Thus if memory is the sole criterion of personal identity through time, and I forget the experiences of a certain part of my life, then it follows that I am not the same person who once had those now

forgotten experiences. And yet it may seem plain that this I of today is the same I who had those earlier experiences whether or not I remember them. Unbeknownst to many critics of the memory theory, Locke is quite aware of this objection and, accordingly, gives a reply. One must stop to notice, says Locke, what it is that the term 'I' here denotes. For in this instance 'I' denotes the man and not the person. And so far as the organization of my biology has continued from three years ago till today, then I am the same man three years ago as I am today even if I have forgotten what transpired back then. But, Locke persists, if I have forgotten the events of three years ago, I am no longer the same person.

Why then, one might ask, do we not allow the plea of amnesia to exculpate someone from his crimes? For the inebriate or somnambulist who cannot later, when sober or awake, remember his actions is nevertheless still held responsible for them. Locke's answer is that our holding someone thus responsible is strictly a pragmatic affair. For although it can be shown that someone has performed a certain act, it is not so easy to determine whether a claim to not remember performing the act is genuine or mere pretence. Although Locke concludes it is therefore just to punish amnesiac offenders, he nevertheless feels that, 'in the great day, wherein the secrets of all hearts shall be laid open, it may be reasonable to think, no one shall be made to answer for what he knows nothing

of' (p.48).

We see therefore that the point made by this objection, i.e. that a person cannot have a persisting identity through lapses of memory, rather than posing a problem for Locke, is a welcome corollary of the memory theory.

A natural reaction here is to claim that if the memory theory leads to such conclusions, then it is not an adequate account of personal identity. For we do seem to allow that persons forget things and yet maintain their identity. There are at least two things that could be said to this. First, I am not convinced that we do allow this in every case. In situations in which someone has total or near total amnesia, i.e. forgetting one's past, not recognizing one's husband or parents, and so on; especially where the condition seems irreparable, I think many people would agree that the person after the memory loss is numerically a different person than the one before. And if it is allowed that there is a critical amount of memory loss that carries with it a loss of personal identity, then it is allowed that memory plays some form of constitutional role in personal identity. Of course, there are others who would refuse to acknowledge a change of identity, whatever the extent of memory loss. If I am right about the discrepancy of agreement here, then this too would offer support for my point made earlier; that there is no unified

common sense view about the nature of personal identity.

Secondly, even if there were a general consensus that personal identity continued despite the loss of memory, it is not clear what this would show. It is true that some would view such a common agreement as an argument against the memory theory, but by the same rule we should allow the popular belief in life after death as an argument against mortality. It seems wiser, therefore, to not put too much stake in conclusions drawn from any putative consensus of belief.

There is however a place where Locke's theory does come upon some obstacles. And the root of this problem, as we shall see, lies simply with his failure to make clear the complexities of the theory and follow through their implications. The objection with which we are concerned was first proposed by Berkeley in Alciphon(2), and later refashioned by Reid into the following example. Suppose, says Reid, that a brave officer who once captured an enemy standard, was both flogged as a school boy for raiding an orchard, and in later life made a general. Further, suppose that while the general remembers capturing the standard as an officer, and while the officer remembers being flogged as a boy, the general has absolutely no recollections of the youthful flogging. On Locke's theory, says Reid, the brave officer is identical with the boy who raided the orchard, and the general is identical

with the brave officer who captured the standard. Whereby it should follow, by the law of transitivity, that the general is identical with the boy who raided the orchard. But as the general has no recollection of the boy's orchard escapade and subsequent flogging, he cannot be the same person as the boy. Therefore, concludes Reid, 'the general is, and at the same time is not, the same person with him who was flogged at school' (p.115). And so it seems that Locke's theory is inconsistent.

One solution to this paradox is proposed by Anthony Quinton. Although it is Quinton's intention to defend the memory theory, it should be noted that the account of the theory which he presents is not strictly Lockean. For Quinton suggests that it is not memory alone which determines personal identity, but rather that it is determined by what he more generally calls mental states, i.e. memory plus character. This view, which is more broadly called the psychological theory, is not so much a different theory from the memory theory as it is an expanded and more thoroughgoing version of it. For both memories and character traits are the same sort of thing insofar as they both might be considered somewhat persistent aspects of an individual's psychology.

Quinton starts by introducing the empirical concept of the soul. Though the word 'soul' is normally used to refer a

substantival and therefore non-empirical entity, Quinton uses it here to refer to a temporal series of related mental states. He seems therefore to be using 'soul' in much the same way as Locke uses the word 'person'. Each moment in the life of a soul will be made up of what Quinton calls a soul-phase; that is, 'a set of contemporaneous mental states belonging to the same momentary consciousness'.(3)

Two soul-phases belong to the same soul if they are connected by a continuous path of character and memory. This continuity between different soul-phases can either be direct or indirect. Two soul-phases are directly continuous if, being temporally contiguous, the character revealed by the constituents of the latter is closely similar to the character revealed by those of the former, and if the latter contains some memories of elements of the former. Indirect continuity occurs, however, when two soul-phases are connected through a series of other soul-phases such that each of these intermediate soul-phases is directly continuous with the soul-phase immediately preceding it, and the original soul-phases constitute the two end-points in the chain.

The way out of the brave officer paradox, therefore, would be to hold that the general's inability to recall the school boy's flogging does not permit the conclusion that he is not identical with the school boy. For although there is no

recollection of the elements of the school boy's soul-phase within the soul-phase of the general, both of these soul-phases are nonetheless connected by the relation of indirect continuity. In other words, between the soul-phases of the school boy and the general, there stretches a series of intermediate soul-phases (one of which is that of the brave officer) each bearing the relation of direct continuity to the one immediately preceding it, In this case the school boy and the general share in the life of the same soul, and so are identical with one another.

But all this is to no purpose. For quite plainly our psychology does not work in the way here described. It is not true that each instance of consciousness necessarily or even usually contains recollections of elements of the previous instances. For this to be the case consciousness would have to be continuously infiltrated by states of memory which would inevitably hinder it from performing its multitude of other functions. Of course in those rare instances when we are engaged in some sustained goal-directed activity which depends on our being aware of the immediate past, then something like a relation of indirect continuity between two soul-phases may occur. Examples of such activities might include solving a mathematical equation or conducting a search. Fortunately, however, most of our mental life is not composed of these sorts of structured activities. And even on those occasions where we

do engage in them, our engagement never lasts very long, if for no other reason than we finally fall asleep. It is thus false that the soul-phase of the school boy and that of the general are related by Quinton's notion of indirect continuity.

Maybe, however, this is being too pedantic. For the nature of the memory relation here under review seems to suggest some form of indirect continuity, even if it is not exactly of the Quintonian sort.

The difficulty with this manoeuvre is that the memory relation in the case of both the general and the officer, and the officer and the boy, differs from Quinton's account in an important way; for here the relations lack, or at least could lack, the requirement of the temporal juxtaposition or continuity. The whole point behind Quinton's appeal to the idea of indirect continuity is to enable him to discover a 'continuous character and memory path' which can be traced through a series of temporally contiguous mental states. But nothing in the brave officer paradox suggests the existence of temporally contiguous mental states. For all we know the general may remember nothing else but the taking of the enemy standard. Furthermore, this may be the only time since the standard was captured that he has been able to recall the event. And the same could be said mutatis mutandis for the officer's recollection of the school boy's escapade. In this

case it would become pointless to insist on the existence of a continuous path which ran throughout the general's life. One might feel, however, that where the relation of memory breaks down at least there would be a relation of character. But it is doubtful whether character by itself would be enough to confer personal identity. For character cannot serve to individuate in the way that memory can. Terence Penelhum makes this point succinctly: 'I can have my father's character without being my father; whereas I cannot have his memories, for to do so would be to be my father'.(4) Character therefore can only be a supplementary consideration, serving to test or confirm the ascription of identity where the veridicality of someone's memories are in question.

Further, even if one could discover such a memory path, it would still need to be explained just how this establishes a relationship of personal identity, especially since memory is supposed to account for the sense of identity. The initial reason, it seems, why one might be tempted to invoke memory to explain identity is because memory is a process which ties us experientially to the past, i.e. it makes us aware of the past. That the tie is an experiential one is important; for there are numerous 'ties' to the past - such as digesting yesterday's food, circulating previously oxygenated blood, conditioned reflexes, etc. - none of which, in themselves, seem to give us a sense of identity. But if this is the importance of memory,

then the construction of a memory path of indirect continuity is a pointless endeavour. For the ineluctable fact remains that the general has no remembrance of the school boy's flogging. And the existence of someone temporally between the two, someone who remembers one and is remembered by the other, does not fundamentally alter this problem.

We are thus forced to consider other possible escape routes from the paradox. One way which naturally suggests itself here is to accept that because the general has no memory of the school boy's experience, he is not the same person as the school boy. This, evidently, is the route Locke would want to take. And yet the paradox is constructed in such a way to eliminate this means of escape. For if the school boy is identical with the officer who is himself identical with the general, then it follows by transitivity that the school boy and general are likewise identical.

The interesting thing, however, about this purported identity between the boy and the general, is that it can only be deduced because Locke has allowed his theory to resemble the strict theory of personal identity. To see how this has come about, let us return to the example mentioned in the last chapter.

There we cited Locke's assertion that if I who am now

writing remember both seeing the Thames overflow last winter and remember seeing Noah's flood, then I who am now writing must be identical with both the person who saw the Thames overflow and with the person who witnessed Noah's flood. But why, we might want to know, is my identity with a person of thousands of years ago established just because I who am now writing have a memory of a single experience had by that person? Certainly the person who witnessed the Biblical flood - let us call him Shem - also underwent a host of other experiences during his life; say, the celebration of his tenth birthday, the first meeting with his wife, or the felling of cypress trees with which to build the ark. But apparently, so far as Locke's theory goes, it is not necessary that I remember all or even most of these experiences. All that is necessary is that I remember just one of Shem's experiences. This will suffice to determine my identity with Shem. For, says Locke, 'let him once find himself conscious of any of the actions of Nestor, he then finds himself the same person with Nestor' (p.44). It seems, therefore, that Locke is expounding a view which renders personal identity an all or nothing affair, in other words a strict theory. Either I have a memory of someone who witnessed Noah's flood or I do not. In the former case I am identical with that observer of the flood, in the latter I am not. This means that any memories of Shem's life which I possess over and above the observation of the flood are superfluous as far as defining my identity goes - not because

there is anything particularly definitive about seeing the flood, but only because all I have to do to be identical with Shem is to have a memory of one of his experiences.

But why would anyone who thinks that personal identity is constructed out of memory hold such a view? If it is my memories that bestow my identity upon me, then it would seem that the more memories I acquire, the more of an identity I acquire, and equally, that the more memories I lose, the more of an identity I lose. Consequently, personal identity, according to its explanation in terms of memory, is not rightly considered an all or nothing affair, rather, it is something which proceeds by a matter of degrees. Thus having merely one of Shem's memories would not entitle me to claim full identity with him. But if, on the other hand, I could remember large sequences of Shem's life, then the identity between him and me would be of a greater degree.

Locke's failure to see this point - that under the memory theory, personal identity is not an all or nothing affair - would seem to be the result of an unintentional commitment to the notion of an enduring self - the very notion he is attempting to overcome. For the only way that my having a single memory from Shem's memory life could render me identical with Shem, as if memory were taken, as Reid has proposed, as evidence for personal identity rather than as the stuff of

which personal identity is made. On this story, if I have a memory of some earlier person's experience, then I am identical with that person; not because personal identity is defined in terms of memory, but just because memory is always memory belonging to an enduring self; a self which has memories, but is not itself memories. And if I have the memory of an earlier self, it can only be because the enduring self which is me is one and the same with that earlier self. But if we adopt a view of personal identity which explains the identity of the self as constructed out of memories, then a memory cannot be constructed as evidence for a self; at the most it can only be part of a self. Therefore, if I have a memory of only one of Shem's experiences, then the most I can hope for is but a very partial identity with Shem.

We can now see the resolution of the brave officer paradox. For the officer to be fully identical with the school boy, he must be able to remember all of the school boy's experiences. Plainly, however, such a feat will be next to impossible. Thus although the officer recalls being flogged as a school boy for raiding the orchard he will not remember all the vicissitudes of the boy's life. And so his identity with the school boy will only be a partial identity. The same holds true for the general's identity with the brave officer. For although the general remembers capturing the enemy standard, he will not have a recollection of all the officer's experiences.

Thus the identity which holds between the old general and the brave officer must also be a partial identity. But if the brave officer shares only a partial identity with the school boy, and the old general shares only as partial identity with the brave officer, it does not follow that the old general shares any identity with the school boy. This is because the relationship of partial identity is non-transitive. This is evident from the fact that even if A is approximately equal to B and B is approximately equal to C, A might still be totally unequal to C. And once we have pulled the carpet of transitivity out from under Reid, he has nothing left on which to erect his paradox. The paradox that the general is and at the same time is not identical with the person who was flogged at school, only gets going if we allow that the general is in fact identical with the person who was flogged at school. But since the general has no recollection of the school boy's flogging, his identity with the school boy must be established by appealing to his identity with the officer, and then by pointing to the officer's identity with the school boy. That is, by making an illegitimate appeal to the law of transitivity.(5)

There is, however, an objection to this method of escape from the paradox. This is the objection to the notion of partial personal identity. Reid himself tells us that 'The identity of a person is a perfect identity, wherever it is

real, it admits of no degrees; and it is impossible that a person should be in part the same, and in part different' (p.111). Reid's reason for thinking this is his belief in indivisibility of the person:

A part of a person is a manifest absurdity. When a man loses his estate, his health, his strength, he is still the same person, and has lost nothing of his personality. If he has a leg or an arm cut off, he is the same person he was before. The amputated member is not part of his person, otherwise it would have a right to part of his estate, and be liable for a part of his engagements. It would be entitled to a share of his merit and demerit, which is manifestly absurd. (p.109)

Although it is always good to come across instances of light humour in the often weighty writings on metaphysics, in this case it seems that Reid's irony has led him astray. As far as I can see, Locke's theory nowhere suggests that a person is composed of his estate, health, strength, or even of his various limbs. It is true that Locke remarks that in the event of one's little finger being amputated, were consciousness or memory to leave the rest of the body and 'go along with' the little finger, then it would be the little finger that would maintain the personal identity, not the unconscious body. Despite this apparent equation of a person with a body-part, it is evident that what Locke means to say is that as it is memory alone that bestows personal identity, it matters not where the memory or personal identity is lodged, be it the body or the little finger.

This is analogous to the contention of Locke's described earlier; namely, that the relationship between a person's identity and an immaterial soul could only be a contingent one. The same holds true for the relationship between a person's identity and the body or body-parts.

So Locke, of course, will agree with Reid that the loss of such trivial items as one's health or limbs entails nothing about the loss of one's personal identity. However, were Reid to have said the same about the loss of one's memory, then Locke would have a call for remonstrance. For the loss of memory is just the loss of personal identity. The question that comes to mind then is 'Why does Reid, who is after all attacking the memory theory, point to the unimportance of losing one's estate and limbs, but fail to say the same of memory?'. A possible answer here is that Reid could foresee a difficulty in persuading his readers were he to openly make similar claims about the loss of memory. The claim that even after a person has suffered a complete loss of memory he is still the same person he was before, does not strike one as indubitable. It is therefore safer to stick with remarks where one is on surer ground - remarks about personal identity persisting through the loss of one's estate and limbs - and hope that people will think this implies a similar truth about the loss of memory.

Maybe the reason why Reid does not produce a serious argument for his view, is because he thinks the absurdity of 'part of a person' is so manifest that one is not needed. I, however, see nothing obvious in our notion of a person that implies there can be no splitting into parts. We do allow, for instance, that certain individuals lack qualities that are considered central to the notion of being a person. Thus we allow that someone who is criminally insane is not responsible for his actions. Yet responsibility for one's actions, i.e. agency, is central to our idea of being a person. It therefore seems reasonable to say of one who is criminally insane that he is missing part of what it means to be a person.

Further, it will be recalled that Reid says a person is not thought, feeling, or action, but rather is something which thinks, feels and acts. But what, for Reid, is the nature of the relation between thought, feeling, and action, and the person who does them? Is it a contingent relation or a necessary one? If it is a contingent relation, then it should be possible for there to exist a person who lacks the ability to think, feel, and act. But what sort of person could this be? Such a person could have no sense of being a person, and therefore no conviction of his identity: to do so he would have to be able to think or at least to feel. Reid would no doubt be the first to admit the manifest absurdity of such an idea.

If, on the other hand, it is a necessary relation then the ability to think, feel, and act become definitive marks of personhood. That is, it will be essential for anything to be a person, that it must at least be able to think, feel, and act. But were Reid to accept this, he could no longer maintain that a person is not to be identified with his thoughts, feelings, and action. For by allowing that certain abilities are essential aspects of a person he has thereby allowed that a person is not something which can exist apart from those abilities. A person would therefore consist of at least thinking, feeling, and, action; all of which would consequently be parts of a person.

One reply here could be to claim even if a person were not able to exist without certain abilities, it does not follow that a person is therefore to be identified with these abilities even though they are necessary elements of personhood. But just what sort of necessity would, in this case, bind a person to the abilities of thinking, feeling, and action, I cannot begin to imagine. For if a person is something other than thinking, feeling, and action, then it seems we should be able to conceive of a person without these abilities. And if we can form such a conception, then the relation between a person and these abilities is one of contingency, not necessity.

Another way that we can arrive at the idea of a partial person is by noticing that the characteristics of a person are not themselves an all or nothing affair. Thus when Madell tells us that a person has no essential qualities except consciousness, already implicit within this remark is the notion of a partial person. For consciousness is something which comes in various degrees. At one end of the spectrum we have the person who is fully alert, attending completely to the flow of experience, while at the other we have someone who is barely awake, bordering perhaps on the edge of a coma. In between these extremities there stretches a gradient of varying intensities of consciousness. So if consciousness is an essential quality of persons, the less of consciousness there is, the less of a person there is. And if someone gradually loses consciousness over a period of time, then it follows from what has been said that at the end of his decline he is only minimally identical, if that, with the person he was before the decline.

One might be tempted here to object that no matter how little consciousness a person has, as long as he still has some consciousness he is still a person. And, therefore, that although consciousness may be something which is more or less, being a person is still something which is all or nothing; for either one has consciousness or one has not. The difficulty in sustaining this objection is that it ignores the fact that

persons are not the only things that are conscious. And further, that one of the features that is meant to distinguish persons from other animals, especially the lower animals, is their degree of consciousness. Thus it seems reasonable to expect that when a person has lost a specific degree of consciousness, there would then be little to distinguish the form his consciousness has taken from certain forms of non-human animal consciousness. One could take exception to this on the grounds that human consciousness is of a fundamentally different sort than is non-human consciousness, but I do not see any evidence to sustain such an objection. We interact with animals and have little difficulty discerning their desires, fears, intentions, and other conscious states. The reason for this would seem to be that we recognize in them the same sort of conscious states, and hence the same sort of consciousness, that we experience in ourselves.

It is worthwhile to note that Reid (in agreement with Butler) is quite happy to permit the notion of partial identity with reference to physical objects. The identity, says Reid, 'which we ascribe to bodies, whether natural or artificial, is not a perfect identity; it is rather something which for conveniency of speech, we call identity' (p.112). So Reid's objection is not with the concept of partial identity, that is, with an imperfect identity which is non-strict and holds only to a greater or lesser degree. Rather, it is with the

application of this concept to persons. However, there is no apparent reason why the idea of partial identity cannot be applied to persons.

In light of this it does not seem that the brave officer paradox presents an insurmountable difficulty for the memory theory. There is, however, another objection to this theory, or rather to the more expanded psychological theory, and here an appeal to the concept of partial identity will be of no avail.

This objection is proposed by Bernard Williams in 'Personal Identity and Individuation'.⁽⁵⁾ Williams' attack on the psychological theory starts with his imagining a situation in which someone called Charles, wakes up one morning with a totally different set of character traits and memories from those he had before going to sleep. Previous to retiring for the night Charles had been quiet, deferential, church-going, and home-loving. Now he has suddenly become and continues to be loud mouthed, blasphemous, and bullying. In addition to this he can no longer recall any of the life he led before that fateful night, and now claims to remember witnessing events and performing acts that are totally at odds with what he earlier claimed to remember. Upon investigation it turns out that all of Charles new memory-claims point overwhelmingly to the life of a particular historical person; namely, Guy Fawkes. (Presumably, the newly acquired character traits also suggest

those of Guy Fawkes, though Williams does not say). According to the psychological theory we would be obliged to conclude that Charles now has the identity of Guy Fawkes.

But this conclusion, says Williams, is premature. For if it is possible that Charles should undergo this drastic change of character and memories, it is equally possible that someone else - Williams suggests Charles' brother Robert - should undergo an exactly similar change. But what should we say if this were to happen? They both cannot be Guy Fawkes, for then Guy Fawkes would be in two places at one time. Further, if they were both Guy Fawkes they then would both be identical with one another. And both these implications, states Williams, are absurd. One could try to elude this difficulty by saying that one is Guy Fawkes while the other is exactly like him. But this move, we are told, would be vacuous because there would be no way of deciding which one is to be Guy Fawkes and which one is to be exactly like him. All we could say therefore would be something like both Charles and Robert have mysteriously become like Guy Fawkes, or as Williams also puts it, clairvoyantly knew all about him. And if that is all we could say about both of them, then it seems that is all we could say about one of them.

The reason for the psychological theory's inability to establish an identity between Charles and Guy Fawkes, which

Williams spells out in his further paper, 'Body Continuity and Personal Identity', (7) is because identity is a one-one relation, while the tendency to make sincere memory-claims is a one-many relation. What Williams is saying here seems to be this: in deciding whether any object of one particular time T1 is identical with an object of another particular time T2, whatever criterion is used, it cannot be such that it allows there to be more than one object at T2 which has the relation of identity to the object at T1. It is the breach of just this requirement that is the downfall of the psychological theory. For although only one person at T2 can be identical with a particular person at T1, and only one person at T2 can correctly remember being a particular person at T1, it is always possible that many persons at T2 can sincerely claim to remember events which exactly match the life of the person at T1.

And this difficulty persists inspite of the fact that, for the psychological theory, personal identity over time will be a matter of degree. For it is perfectly possible that Guy Fawkes, Charles, and Robert all have exactly similar sets of memories. This, as we have seen, was different in the brave officer paradox. There the school boy did not have the same memories as the officer since he had not yet, among other things, captured an enemy standard. Also, the general did not have the same memories as the officer since he lacked, at

least, the memory of stealing fruit as a boy. And it was these discrepancies which allowed us to argue that the general could only share a partial identity with the school boy. However, in Williams' example there is nothing which guarantees any discrepancies between the memory sets of the three characters. Even if it turns out that Charles is only partially identical with Guy Fawkes, i.e. he only possesses some of Fawkes' memories, this will have no bearing on the possibility of Charles and Robert having exactly similar memories. If this much is true, then the psychological theorist will be obliged to conclude that Charles and Robert are fully identical with one another. And this would be a reductio ad absurdum of his position.

What are we to say to this? Certainly we must agree that identity is a one-one relation, and also that the tendency to make sincere memory-claims is not. But need we agree that only one person can correctly remember being a particular person in the past? This issue is not so easily resolved; for it depends how we understand the phrase 'remember being a particular person'. If it is taken to mean 'remember being identical with a particular person', then because identity is one-one we will agree with Williams. But if it is to mean that only one person can 'remember the experiences of a particular earlier person', i.e. be psychologically continuous with a past person, then there is good cause to reject Williams' proposal. For

there is reason to believe that psychological continuity need not be a one-one relation; that is, that it is logically possible for two persons, at the same time, to correctly remember having the same experiences of one earlier person. One of the reasons for this belief comes from the findings of research on brain bisection.(8)

It has been known for a while that the human brain, partially composed as it is of two near-equivalent cerebral hemispheres, can be surgically divided with only minor impairment of its normal functioning. The original purpose for this sectioning of the brain was to alleviate severe cases of epilepsy by limiting the cause of the seizure to one half of the brain. In their normal everyday life patients having undergone this operation seem to have a full recovery, showing nothing peculiar or disintegrated in their mental abilities or behaviour. However, in experimental conditions it proved possible to isolate the two halves of the bisected brain and so induce each hemisphere to function independently of, and even at variance with the other hemisphere. Thus, for example, when a smell is presented exclusively to the right nostril - which is connected to the right hemisphere - the split brain patient will insist verbally (since language is under the domain of the left hemisphere) that he smells nothing. Nevertheless, his left hand, which like the right nostril is connected to the right hemisphere, will at the same time obey the experimenter's

instructions and pick out the object that is causing the smell. Thus what we have seems to be two separate streams of consciousness within the same person.

Not only does this state of affairs undermine the doctrine of the unity of consciousness (two streams of consciousness are not a unity), but, as David Wiggins realized, it opens a Pandora's box for the problem of personal identity.

In his booklet on Identity and Spatio-Temporal Continuity Wiggins refers to an operation imagined by Sydney Shoemaker in which the brain of an individual called Brown is transplanted into the body of someone called Robinson (this of course is just the Lockean tale of the prince and the cobbler in modern guise). Although the resulting individual - Shoemaker calls him Brownson - would have Robinson's body, he would nonetheless have Brown's personality and memories. Shoemaker believes this would dispose us to identify Brownson with Brown and not with Robinson. Shoemaker admits that Brownson would have part of Brown's body, i.e. his brain, but nevertheless sees it as an absurdity to equate personal identity with brain identity.

Wiggins expands on this story and by drawing attention to the research on brain bisection. Because of this research, says Wiggins, rather than transplanting Brown's entire brain

into one body, we could easily imagine splitting Brown's brain and placing the subsequent halves into two separate bodies. The intelligibility of this suggestion is further enhanced by the fact that one cerebral hemisphere seems all that is necessary to carry a person through life. For human beings who have suffered destruction of one of their hemispheres have been known to recover without major disturbances to their personality or mental functions.

It now becomes clear why 'being psychologically continuous with a past person' need not be a one-one relation. For both recipients of Brown's cerebral hemispheres will be psychologically continuous with Brown: both will display his personality and have his memories. Williams' objection that two individuals cannot both be identical with an earlier person, though true, must now lose some of its force. For we cannot here, in Williamsian fashion, simply banish the memory-claims of Brown's cerebral inheritors to the realms of clairvoyance and mystery. For as Wiggins puts it, 'we understand far too well why they have these memories'.(9)

Still, there is a major difficulty here; for if the two recipients of Brown's cerebral halves are not identical with Brown, we are bound to ask what has become of the person who is identical with Brown. Has he ceased to exist? And if he has ceased to exist, then it seems odd to say that the two

persons with his hemispheres in their heads are, in any important sense, Brown's psychological survivors.

This problem has been explored in some depth by Derek Parfit.⁽¹⁰⁾ The work Parfit undertakes in this area, though largely a bringing together of earlier ideas, is worth detailed attention because it represents an interesting attempt to save the psychological theory while at the same time allowing that it is not fully equivalent to the idea of personal identity.

Parfit approaches this issue by pointing out that for the person about to undergo cerebral fission and transplant, there are only four possibilities: 1) I do not survive; 2) I survive as one of the two people; 3) I survive as the other; 4) I survive as both.

Unfortunately, says Parfit, each of these possibilities is implausible. The first possibility is implausible because, as Shoemaker has argued, I would survive if my entire brain were successfully transplanted. Because of this, and because I can survive with only half my brain it seems to follow that I would survive the transplantation of only one of my hemispheres. But if my survival is possible here, then why should I not survive the instance in which both my hemispheres are separately transplanted? In Parfit's words 'How can a double success be a failure?' (p.256).

The reason for the implausibility of 2) and 3) is that since both of my hemispheres are exactly similar, there is nothing which could make me one person rather than the other. Waking up after the operation with only one hemisphere, I might well believe that I was the sole survivor of the fission and transplant. But the person with the other hemisphere might equally believe this of himself. Since both of us would be recipients of exactly similar brain halves, there would be no way to adjudicate between our rival claims.

The fourth possibility is rejected by Parfit because of Williams' point; namely, that two individuals cannot be one and the same person.

We thus seem to be left with a question which admits of no answer. This puzzle, Parfit claims, arises because the question of what happens to me after my fission is being asked from the non-reductionist point of view. For only if we think that the self is something which exists in addition to psychological or bodily continuity will we think that the four possibilities represent genuinely distinct options. However, when we ask this question from a reductionist perspective the problem disappears. 'On this view', says Parfit, 'the claims I have discussed do not describe different possibilities, any of which might be true, and one of which must be true. These claims are merely different descriptions of the same outcome'

(p.259). The outcome is that after the operation there will be two persons each of whom will be psychologically continuous with myself before the operation. This is all there is to know. For the person who persists in asking 'But what happens to me after my fission?' Parfit replies merely that this is an empty question. It is not empty however, in that it admits of no answer, but rather that it simply gives different descriptions of the same outcome. Nevertheless, Parfit feels that description 1) I do not survive, is the best of the competing descriptions. Parfit's reason for choosing 1) as the best description is that 'I cannot be identical with two different people, and it would be arbitrary to call one of these people me' (p.262). And yet, because the four descriptions do not represent distinct possibilities, the important question is not 'Which is the best description?' but is rather 'What ought to matter to me? How ought I to regard the prospect of division? Should I regard it as like death, or as like survival?' (p.260). Further, once it has been decided what matters, says Parfit, then we will be able to decide whether 'I do not survive' is the best description.

For Parfit, what matters is relation R; that is, the relation of psychological connectedness and/or psychological continuity, with the right kind of cause. Psychological connectedness refers to instances of particular direct psychological connections, while psychological continuity is

just the occurrence of overlapping chains of strong connectedness. It is important that the connectedness of psychological continuity is strong connectedness, because connectedness is something which can hold by varying degrees. I may have several direct psychological connections with myself of yesterday, or I may have only one. In the latter case Parfit feels the ascription of identity between myself of yesterday and myself of today would not be justified. Strong connectedness by itself, however, is not enough to justify the ascription of personal identity; for identity is transitive, strong connectedness is not. A, for example, can have strong connectedness with B, who can in turn have strong connectedness with C. And still A may have no connectedness at all with C. This is where the relation of psychological continuity comes in. For if A is related by psychological continuity to B (i.e. by overlapping chains of strong connectedness), and B is similarly related to C, then A is likewise related by psychological continuity to C. In other words, psychological continuity is transitive.

The last requirement of relation R is that psychological connectedness and/or continuity have the right kind of cause. On a narrow interpretation of the psychological theory of identity, the right kind of cause refers to a 'normal' cause. The purpose of stipulation that the right kind of cause be a 'normal' cause is to avoid the possibility of connections of a

delusive sort entering into the chain. I may, for example, seem to remember as a child turning a garden snake loose in the house; and it may be true that as a child I did such a thing. However, my memory experience might be caused by my mother's continuous reference to the incident, rather than by my actual experience of the incident. That is, it would not be produced by the right kind of cause. This, it might be recalled, is the same sort of principle that we invoked in the second chapter to obviate Butler's complaint concerning the circularity of the memory theory. It is a way of demonstrating that reference to a past person is not necessary to substantiate a present person's memory-claims.

Parfit, however, rejects this narrow interpretation in favour of a wider one in which the right kind of cause could be any reliable cause, or even any cause. This is because he wants to avoid an interpretation of the psychological criterion of identity which refuses to acknowledge memories or other mental states acquired in an abnormal way. This is Parfit's way of allowing for the logical possibility of such things as person replication. The example that Parfit has in mind is the concept of 'teleportation' first discussed by Daniel Dennet.⁽¹¹⁾ We can imagine the case, says Parfit, in which a scanner records the exact state of all the cells in my body, destroys my entire body, and beams the information to another planet where my brain and body are reconstituted by a

replicator designed for such purposes. The brain and body of my replica on the other planet would therefore be exactly similar to my brain and body as they were on earth. Consequently, my replica would be psychologically continuous with me. However, since the psychology of this new extraterrestrial person would be caused in an abnormal way, i.e. by teleportation, my identity with this person could only be asserted if we allowed that the cause of psychological connectedness and continuity be any reliable cause.

Now, says Parfit, when a future person is R related to my present self, and there is no other future person who is R related to my present self, then my present self and that future person are one and the same. However, when relation R takes a branching form - as in the case of Wiggins' brain-splitting operation - then, because there will be more than one future person who will be R related to me as I am now, the relation of identity between myself and these future persons cannot hold. Thus personal identity is nothing more than a non-branching instance of relation R.

But why should it matter to me whether my relation R takes a branching or a non-branching form? Why should it be a worry to me that I will be psychologically continuous with two future persons, rather than with just one? Of course, neither of the future persons enjoying the use of my cerebral hemispheres will

103-5, 2
do not follow this,

be me, but this is just something to do with the logic of identity, and so is extraneous to the question of what really matters; namely, the continued existence of relation R. Thus, says Parfit, 'When we see why neither resulting person will be me, I believe that, on reflection, we can also see that this does not matter, or only matters a little' (p.263).

Coming back to the original question of what will become of Brown once he has divided into two separate persons, we can now give Parfit's answer: Brown will die, but since both his psychological continuators will maintain relation R to Brown, 'this way of dying is about as good as ordinary survival' (p.262).

It is important to remember, however, that for Parfit, this answer is merely the 'best' of various descriptions: it does not refer to an outcome distinct from the other answers. This is because when the four possible outcomes are seen from the non-reductionist point of view, i.e. as constituting separate outcomes, each answer seems implausible. Oddly enough, however, three of the possible answers turn out to be just as implausible when seen from Parfit's point of view as 'merely different descriptions'. Thus, the reason why 2) I survive as one, and 3) I survive as the other, are dismissed as implausible outcomes, is because there is no way of choosing between them: both of the post-operative individuals will have

exactly similar character traits and memories. But this is the same reason Parfit adduces for rejecting 2) and 3) as candidates for the best description. Likewise, 4) I survive as both, is said to be implausible as a distinct outcome because of its logical incoherence: I cannot be one and the same person as two different people. And, here again, it is for this same reason that Parfit decides not to choose it as the best description. But if the considerations which count against 2), 3) and 4) being plausible as outcomes also count against their being the best of the possible descriptions, then I fail to see the purpose of calling them 'merely different descriptions' rather than outcomes. In fact, with 4) it is quite clear that it makes no difference whether we call it an outcome or a description; for being logically impossible as an outcome, it will also be logically incoherent as a description. And a logically incoherent description is not merely a different description of a particular outcome, for, being logically incoherent, it describes nothing at all.

It might be thought that in construing the possible answers as descriptions at least Parfit is able to come up with one acceptable answer, i.e. 1) I do not survive, whereas when they are seen as separate outcomes none of them is acceptable. But now we must ask the following. If the reasons for discounting 2), 3) and 4) as implausible outcomes are good enough for discounting each of them as the best description,

why is the reason for discounting 1) as an implausible outcome not good enough for likewise discounting it as the best description? Why Parfit thinks 1) is implausible as an outcome, it will be recalled, is because since my survival is indicated when only half of my brain is transplanted, it seems I should also survive the occasion in which both my brain halves are separately transplanted. And yet, if these are adequate grounds for the dismissal of 1) as an implausible outcome, and if the grounds on which an outcome is rejected are also acceptable grounds on which to reject candidates for the best description, then why does not Parfit also reject 1) as the best description?

Parfit's acceptance of 1) is made all the more baffling by his statement that once we have answered the question of what ought to matter to me, or how ought I to regard the prospect of my fission, then we can judge whether 1) is the best description. For as it turns out, that which matters - relation R - is something which fission cannot destroy. The prospects for personal identity are less fortunate of course, but as personal identity is not what matters, this is of no grave concern. But if what is of most consequence to me passes undisturbed through my fission, then clearly I ought not to regard the prospect of my fission as though it were death; especially since it is 'about as good as ordinary survival'. And if I ought not to regard it as though it were death, then

'I do not survive', hardly seems the 'best' description. And yet because Parfit rejects the possible descriptions in which it is claimed that I do survive, there seems no other description available to him. So Parfit with his descriptions is now in no better a position than the non-reductionist with his outcomes. Both have no answers.

One way of remedying these confusions, it might seem, would be for Parfit to hold to the claim that everything about the outcome of the operation could be known without choosing one of the descriptions. This would seem to free him from the hopeless task of choosing between four unacceptable descriptions.

Unhappily, this is an escape route which Parfit himself has already sealed off. For he expressly says that although the question of what happens to me during my fission is an empty question, it is not empty so far as it admits of no answer. Some empty questions have no answer simply because any answer we decide to give them runs the risk of being arbitrary. Choosing an answer to such a question 'would be pointless and might be misleading' (p.260). But the empty question Parfit is considering is different. It is empty because its possible answers are but different descriptions of the same outcome. Since Parfit does not think it pointless or misleading to choose one of these descriptions over the others,

he must allow that the question of what happens to me during my fission, although empty, does have an answer.

It is natural then to wonder why Parfit insists that this question is the sort of empty question that has an answer; especially since the answer which he chooses, according to his own account of how I ought to view my fission, is at least somewhat misleading. The reply here, I think, is that for Parfit to accept that there is no answer to this question would be, on his own admission, to accept that the choice of any answer would be arbitrary. As we have seen, however, the reasons which lead Parfit to his choice of an answer, e.g. the logical incoherence of a rival answer, are far from arbitrary. But this is just what is wrong with Parfit's whole account of the matter. For if the possible answers to our question are merely different descriptions of the same outcome (as according to Parfit they must be if we are to avoid non-reductionism), then there is a definite sense in which our choice of answer is arbitrary. This follows from the banal truth that different descriptions of the same thing, for all their difference, still describe the same thing. Consider, for example, three different descriptions all of which are answers to the question 'Who is Francis?' These are a) the person who lives above the flower shop in Nicolson Street, b) the woman who works in the reserve section of the library, and c) Mary's best friend. Now, so far as my purpose is to answer the question 'Who is

Francis?', then it is arbitrary which description I decide to use as an answer. For they all refer to Francis. Of course, there may be other factors which guide me in my choice of description. For instance, the choice of c) might not be apposite if the person asking me the question does not know who Mary is. Also, there may be considerations, of completeness or aesthetics which lead me to choose one description over the other. For example, b) might be considered a more complete description than either a) or c) because it at least tells us the sex of the person being referred to. Or again, 'living above a flower shop' may seem a more poetic designation than would 'working in a library'.

The considerations which guide our choice of description a), b) or c), are not, however, of the same sort as those which dictate Parfit's choice between descriptions 1), 2) 3) or 4). For whether we choose a), b) or c), our response to the question 'Who is Francis?' will be correct. We could make the same point by saying that a), b), and c) are merely different descriptions of the same person. Parfit, on the other hand, has no such freedom of choice with answers 1), 2), 3), or 4); for his choice, as we have shown, is bound by logical considerations. It is this point which militates against Parfit's claim that the possible answers to 'What happens to me during my fission?' are merely different descriptions of the same outcome.

Does this mean we are forced back into the non-reductionist's camp with its four implausible outcomes? Not necessarily; for there appears to be another option. This is to reject all the outcomes on the grounds that they depend on an unsupported assumption; namely, the assumption that there is an I which is there to undergo fission in the first place, and of which, consequently, we must be able to give a post-fissional account. If we do not allow this assumption, then the question of what happens to the I does not arise: for the I does not exist. Here then we are not saying that 1)-4) represent genuine outcomes, nor are we saying they are merely different descriptions of the same outcome: we are rejecting them as loaded answers to a loaded question. Parfit would no doubt object here that he is not assuming the existence of an I because on the psychological theory what we call the I is nothing more than relation R. This is presumably what he means when he says he could give a complete description of reality without claiming that persons exist (p.212). But what if relation R turns out to be just as spurious as the persisting I that it was invented to replace? Let us have a closer look.

What strikes us first about this relation is that its foremost components - psychological connectedness and psychological continuity - are not altogether dissimilar from Quinton's direct and indirect continuity. For both

psychological connectedness and direct continuity refer to instances of direct psychological connections, while both psychological continuity and indirect continuity refer to overlapping chains of direct psychological connections. One difference is that while Quinton describes the relations as holding between soul-phases, or 'sets of contemporaneous mental states belonging to the same momentary consciousness', Parfit refers to the relations as holding between persons on different days. However, it is evident that, for Parfit, persons are little more than constructions out of something very similar to what Quinton calls soul-phases. Another difference is that, in Quinton's account, the connections between the directly continuous soul-phases that constitute the relation of indirect continuity need not be strong. All that is required for a relation of indirect continuity to obtain is that there occurs a series of directly continuous soul-phases, each of which contains 'some' connections with elements of the former. The two end-points in this chain will then be indirectly continuous. Parfit, on the other hand, requires that the relation of psychological continuity (Quinton's indirect continuity) be comprised of overlapping chains of strong connectedness; with 'strong' referring to 'at least half the number of direct connections that hold over everyday, in the lives of nearly every actual person' (p.206). Since Quinton does not specify what he means by 'some', when it comes to the amount of connections required we cannot say just how different

his criterion of identity will be from Parfit's. Admittedly, 'some' does have a weaker sound to it than does 'strong', but this might merely be a function of Quinton's discussing connections between sets of contemporaneous mental states of the same consciousness rather than, like Parfit, discussing connections between persons over each day. If it does happen that Parfit's account of continuity requires a greater amount of connections than does Quinton's, this only means that, as far as connectedness goes, it will be more difficult to qualify for the ascription of personal identity on Parfit's account than it will be on Quinton's. The nature of the psychological connections and the role they play in personal identity, however, remain the same on both accounts.

Since this is so, it follows that the criticisms which we levelled against Quinton earlier will also apply to Parfit. For here too we are bound to point out that our mental lives quite clearly do not consist of 'overlapping chains of strong connectedness'. If this was the fate to which we were all condemned, consciousness would be perpetually filled to the brim with memories of where we came from and intentions about where we were going. Life would be like a relentless guided tour from which we could not break free. Within consciousness there would be no room for those events which are experienced as instantaneous and unconnected, as for example when we have a flash of insight, a revelation, a moment of terror, or when we

are merely sitting quietly doing nothing.. Such events, by their very nature, instantaneously clear consciousness of all its connections. When I am in such a state of consciousness I am completely absorbed in the moment. And when I am thus absorbed, there is no place where memories, intentions or desires could exist within my awareness without destroying my absorption. In moments like these, which infiltrate consciousness through and through, there is no chain of connectedness to which I belong.

One could always object here that even if there are unconnected moments of consciousness this need not ultimately interfere with the existence of a chain of connectedness. For even though my consciousness may at that particular moment exist as unconnected, in the future it need not. That is, although I may be absorbed in the present at this particular moment, sometime in the future I might be able to look back upon the unconnected event and so bring it back into connection with the rest of my mental life. This of course is true. The problem, however, is that my identity at the moment in which I am unconnected is then to be decided by a future event, i.e. my eventual recalling of that moment. This means that during the time I am unconnected I have no identity. We thus have the curious situation in which my identity at various instances - which are by no means rare instances - is floating about waiting to be determined, i.e. is waiting for the future

occasion in which I might (or might not) recall my unconnectedness. Quinton, it seems, was aware of this problem and tried to steer clear of it by arguing that each moment of consciousness has a similar character to and contains some memories of the immediately prior moment of consciousness. And this would seem to be Parfit's reason for alleging the existence of overlapping chains of strong connectedness: each moment of consciousness is strongly connected to prior moments through overlapping chains. If this were true, it would avoid the present difficulty. But it is not true.

Parfit tells us that when he used to believe in non-reductionism he felt his life was like 'a glass tunnel, through which I was moving faster every year', however, when he changed his view to reductionism 'the walls of my glass tunnel disappeared' (p.281). But although Parfit may have shed his glass tunnel, it is debatable whether he is any better off; for now he lives his life in chains - overlapping chains of strong connectedness.

Another criticism we had of Quinton was that even if indirect continuity did hold between me and an earlier person, this would still not explain why I should be considered identical with that person. And the same can be said of Parfit's relation of psychological continuity. Of the two relations of psychological connectedness and psychological

continuity, Parfit says that psychological connectedness is the most important. Psychological connectedness, however, is not a transitive relation, and so cannot represent a relation of identity. This is the reason for the introduction of psychological continuity which, like identity, is a transitive relation. But many relations are transitive. Being a relative of someone or being the cause of something are both transitive relations. But that does not make them relations of identity.

Still, we must remember that Parfit is not arguing for the reduction of personal identity to relation R, i.e. psychological connectedness and/or psychological continuity, with the right kind of cause; for the possibility of fission has shown the futility of this project. Rather, he is arguing that relation R is what ought to matter, not personal identity. Looking past the difficulties we have found with relation R, we can now address the question of whether this relation is what ought to matter.

What then are Parfit's reasons for proclaiming the importance of relation R? The reasons for the importance of the first component of the relation, i.e. psychological connectedness, are given when Parfit is arguing against the view that psychological continuity is all that matters. Parfit starts by saying 'we cannot defensibly claim that it would not matter if there were no connectedness' (p.301). And this may

well be so. For it seems that some connectedness in some circumstances is important; as, for example, when we are acting on a certain intention. But this is a far weaker claim than his original one: that what matters is strong connectedness.

There is an important difference here; for it does seem we can defensibly claim that it would not matter if there were no strong connectedness. Of course, in Parfit's view, a person who is not strongly connected with his earlier selves would not be the same person as those earlier selves, but since identity is not what matters this cannot be a relevant consideration.

Parfit's first argument for the importance of connectedness asks us to 'consider the importance of memory. If our lives have been worth living, most of us highly value our ability to remember many of our past experiences' (p.301). But just how highly we are supposed to value such memories is thrown into question when we compare this with Parfit's previous claim that he would rather suffer amnesia than undergo surgery that gave him an obnoxious character (p.208). With this we see that memories are not to be valued, at least by Parfit, more highly than a pleasant character. Further, Parfit's assertion about the value of memory is patently not a consideration of the importance of memory; it is a consideration of the importance of the ability to remember past experiences of a life 'worth living'. And this is a distinction which needs to be drawn; for it is a debatable

point whether many of the people suffering from alcoholism, depression, suicidal thoughts, and chronic boredom - not to mention the poverty-ridden masses of the third world - would say they had lives worth living. (12) Maybe what Parfit ought to have said is that the ability to remember our past experiences is a necessary though not sufficient condition of having a worthwhile life. But if we accept what he says about preferring amnesia to an obnoxious character, then it seems that memory is not even a necessary condition. For the life of an amnesiac is being presented as being at least more worthwhile than that of an obnoxious character.

The second reason given for the importance of psychological connectedness concerns the importance of the continuity of our desires and intentions. Here Parfit remarks that he would greatly regret both the loss of his love towards certain people and the replacement of strong desires which he has to achieve certain aims. This is because 'I must care more now about the achievements of what I now care about. Since I care more about the fulfillment of these present desires, I would regret losing these desires, and acquiring new ones' (p.301). But this sort of clinging to present desires at the expense of different future ones rules out the possibility of growth and evolution. Thus a child may have strong desires to stay with his mother or play with his toys. He would therefore care strongly about the fulfillment of these desires and deeply

regret their loss. Clearly, however, if the child is to grow out of his childhood, he must be able to eventually give up these desires and acquire some new ones. And of course someone need not be a child to have desires or intentions which it would be best for him to eventually replace. Parfit's view on the importance of the continuity of present desires and intentions cannot accommodate the importance of growth or change because he is forced to evaluate his desires and intentions from his present perspective. But, as Williams has asked, 'Why should I hinder my future projects from the perspective of my present values rather than inhibit my present projects from the perspective of my future values?'. (13) It could be replied that I cannot know what my future values will be. But this is not always true: even children seem to know that one day they must give up certain attachments and move on to being a grown-up. Besides, we certainly do not have to know what our future desires or values will be to know that we will no doubt acquire some, and, consequently, that they may well replace our present ones. Knowing this, we would also know the folly of desperately clinging to our present desires or regretting their loss.

Another difficulty arises for Parfit's account because it is always possible that I may care strongly about the fulfilment of my present desires, and yet not regret their loss; indeed I might be overjoyed at finally ridding myself of

them. This is because a person may have a countervailing but less strong desire. Consider the case of someone who is ruled by a passion for power and control over others. Such a person might well be aware of the destructive nature of this desire and long for release from its grasp. Yet, at the same time he may feel himself powerless to change. A person in this predicament is someone who cares strongly about the satisfactions of his present desires - so far as he continues compulsively to pursue them - but would nevertheless rejoice to wake up one morning and find himself free of them.

And a similar situation could be imagined in the case of love. Someone, for example, may be deeply in love, and yet, because of the suffocating nature of the particular relationship crave to be free of its bonds. He may, however, feel too deeply entangled in the situation to be able to give up his love. Still, were he to somehow shake free of his desire it is altogether probable that he would experience great relief. Something like this, I suspect, is the motive behind many crimes of passion. The person in love may feel himself too strongly drawn towards the one he loves to ever envisage the possibility of escaping his own desire. A desperate solution to this dilemma would be to murder his beloved.

The discussion of desire is an opportune place to question Parfit's self-proclaimed affinities with Buddhism. First,

although he tells us that Buddha would have agreed with reductionism (p.273), only a superficial reading of the Buddhist texts could support such an interpretation. If we attend closely to the passages where the self is discussed (which include the passages from which Parfit has drawn his quotations), it is plain that Buddhism rejects not only the non-reductionist view of the self but also the reductionist view. This is because Buddhism rejects the notion of a self altogether. We shall come back to this idea in our final chapter. But even waiving this point, when it comes to what Parfit sees as the central issue; i.e. what ought to matter to me, it is evident that the Buddha would have firmly disagreed. For the Buddha, desire is the germ of all suffering; and only through the cessation of desire could we ever achieve what is really important; that is, freedom from suffering. From a Buddhist perspective, the importance that Parfit invests in the continuance of his desires is little more than the deluded attachment of someone who has not fully accepted the doctrine of anatman or no-self. This is because desire, which is always desire of something that I believe I lack, is based on the delusive distinction between self and not-self. In the words of the Buddhist philosopher Nagarjuna, 'When desire becomes related to "the one who desires," then desire comes into existence. If there is no one who desires, how then will desire come into being?'.(14)

Unfortunately, Parfit's third reason for the importance of psychological connectedness, i.e. the importance of the continuity of character, fares no better. Parfit allows that continuity is here maintained if character changes in a 'natural' way, but claims that most of us value some aspects of our character and would not want these to change.

The first point to note here is that since character is, among other things, a constellation of desires and intentions, our above criticism will apply here with equal force.

Secondly, we cannot help but wonder what counts as a 'natural way' for character to change. If one's character changes because of a near-death experience, a mid-life crisis, or falling deeply in love, has one's character changed in a natural way? It would be peculiar to insist that the character changes in these circumstances were unnatural; for it is just experiences like these that do bring about character changes. But if we were to allow that these changes are natural, then, according to Parfit, we would have to maintain that in such cases continuity is preserved. Obviously, however, it need not be: people can come out of these experiences with fundamental character changes. According to Erich Fromm, 'In very drastic cases of character change one can speak even of genuine conversions, which means a complete change in values, expectations, and attitudes because something entirely new

has occurred in the life of the converted person' (my emphasis).(15)

A further problem with Parfit's claims about the importance of character continuity is evident in his assertion that most of us value some aspects of our character. For although this may be true, valuing some aspects of our character is not equivalent to valuing our character itself. For, unlike memories, desires and intentions, character is a constellation of interrelated psychological and behavioural phenomena. Or, in the words of Fromm, it is an organized system of passions. And valuing some aspects of a system is wholly consistent with deprecating the system itself. What Parfit needs to do in order to demonstrate that most of us value the continuity of character is to show, not that there are some bits or pieces of our character that we do not want to change, but rather that we want our character itself not to change. And this, I suggest, he cannot do.

What we have learned from these arguments is not why psychological connectedness matters, but only why it matters to Parfit. Consequently, his arguments turn out to be little more than an espousal of conservative preferences. Parfit does not want things to change. He wants his memories, desires, and intentions to persist and would deeply regret their giving way to other ones. If his character is to change he would prefer

it did so in a natural way (whatever that means). This 'personal preference' nature of his arguments comes out clearly in his assertion that 'More generally, I want my life to have certain kinds of overall unity. I do not want it to be very episodic, with continual fluctuations in my desires and concerns' (p.301).

But what if we are not conservatives? What if we prefer a life that is full of episodes and shifting desires, a life where memories and the past are of little concern to us? Or what if we simply do not care one way or the other? To such people Parfit has nothing to say. One might feel the urge to complain here that anyone leading this sort of life is plainly maladjusted or is at least living a life that no-one would seriously choose to follow. But this would be little more than an ethnocentric judgement of Western middle-class culture. In Hindu thought the life of the wandering yogin - someone who has discarded his attachments to the past and is unconcerned about the future - is put forward as the ideal. Also, in Buddhist writings we are told that 'the noble one is he for whom past, future, and present are as nothing; he has nothing and grasps after nothing'.(16)

We can now move on to the second component of relation R: psychological continuity. As we saw above, the purpose of introducing psychological continuity into relation R is to

ensure that R, like identity, will be transitive. And it is because of its ability to secure transitivity for relation R that psychological continuity ought to matter to us. But this is a sad confusion. For the whole point of Parfit's work on personal identity is to convince us that personal identity is not what matters. If identity really does not matter, why bother introducing the relation of psychological continuity whose only discoverable virtue is that, being transitive, it supposedly represents the relation of identity? To do so is paramount to asserting that what matters about x is that it has the same property as something which does not matter. And yet, were Parfit not to appeal to the transitivity of psychological continuity he would have nothing left with which to recommend it. For there is no other apparent reason why the existence of someone who is psychologically continuous with me ought to matter more to me than the existence of someone who is not.

The last component of relation R is the causal requirement, i.e. \Leftarrow psychological connected^{ness} and/or psychological continuity have the right kind of cause. The importance of this requirement, as we have noted, is that it guards against the occurrence of such things as delusive memories in the chain of psychological connectedness. Problems appear, however, when Parfit asserts that the right kind of cause could be any cause. Parfit's reason for proposing that the right kind of cause could be any cause, as we have stated,

is to permit the existence of relation R even in those cases where connectedness and continuity are acquired in abnormal ways, as for example in person replication. But in permitting this he has re-introduced the possibility of delusive memories finding their way into the chain of connectedness. Consider again my apparent memory of turning a snake loose in the house. If the cause of this ^{apparent} memory was not the actual incident of my releasing a snake, but rather my mother's mentioning of the incident, then my ^{apparent} memory is delusive. It is delusive because it lacks the right kind of cause. If we allow, however, that the right kind of cause could be any cause, then we must likewise allow that my mother's reminiscences are the right kind of cause; for they are, at least, any cause. And if we grant this, then it seems we are barred forever from being able to distinguish delusive memories from veridical ones.

Our conclusion here can only be that relation R does not, after all, really matter. Having arrived at this conclusion we now have a clearer picture of where Parfit went wrong. Parfit's error, it seems, was that in trying to argue for how I ought to view my fission, or what ought to matter to me, he failed to base his view on the nature of human experience. He looked rather to the old Butlerian model of strict and philosophical identity and, by appealing to a psychological imitation of this view, tried to force experience into its mould: he tried to pour new wine into old bottles; needless to say the wine went

sour.

The psychological theory ultimately fails, it seems, because it cannot explain what it was invented to explain; namely the sense of identity. Without thinking too deeply into the matter we might feel the psychological theory is our best choice if we want to explain what is at the heart of personal identity: the sense that we are supposed to have of being continuing and distinct persons. But if we reflect a bit further we see that the very nature and functioning of memory, which is the central element in the theory, bars it as a plausible candidate for this office. This is because the sense of identity is supposed to be something of which we are intimately and continuously aware. But memory - especially memory of the self - is a state we only lapse into on discreet occasions. If we were forever full of memories, we would never have the chance to experience the events to which our memories would later refer. In other words we would not have anything to remember. Further, a memory is not a homogeneous moment which reveals itself all at once to consciousness. Rather it is an intermittent unravelling of past experiences which breaks and stops, is infiltrated by present concerns, future projects, and the stillness of the moment. In such a coming and going where could there abide a continuous sense of self? At last giving up the attempt to construe our existence as continuous character and memory paths or overlapping chains of strong connectedness,

we can begin to see in ourselves something akin to those Buddhist sages mentioned in the Dhammapada: 'They live in emptiness, in the unmarked, in freedom; difficult it is to trace their course, like that of birds in the air'.(17)

NOTES TO CHAPTER THREE

1. 'Of Identity and Diversity' in Personal Identity, edited by John Perry (Berkeley, California, 1975), p.36. Further references to this edition are included in the text.

2. The Works of George Berkeley, Bishop of Cloyne, edited by A.A. Luce and T.E. Jessop, 12 Vols (London, 1950), 3, pp.298-299.

3. 'The Soul' in Personal Identity, edited by John Perry (Berkeley, California, 1975).

4. Terence Penelhum, Survival and Disembodied Existence (London, 1970), pp.60- 61. As will be seen below even the possession of various memories are not sufficient grounds to individuate someone.

5. Strictly speaking, the general's having no memory of being flogged at school does not thereby leave him totally non-identical with the school-boy; for it is possible that the general has other closely associated memories of the boy's experience; say, kicking the headmaster in the shins during the flogging.

6. Problems of the Self (Cambridge, 1973), Chapter 1.

7. Problems of the Self (Cambridge, 1973), Chapter 2.

8. For a bibliography of this research see Thomas Nagel, 'Brain Bisection and the Unity of Consciousness', in Personal Identity, edited by John Perry (Berkeley, California, 1975).

9. Identity and Spatio-Temporal Continuity (Oxford, 1967) p.53.

10. Reasons and Persons (Oxford, 1984), part III. Further references to this edition are included in the text.

11. The Mind's I: Fantasies and Reflection on Self and Soul, arranged and composed by Douglas R. Holstadter and Daniel C. Dennet (Brighton, 1981), pp. 3-7.

12. As is well known, one reason behind the addiction to alcohol, along with other depressants such as tranquilizers, is its ability to help the addict to forget; i.e. to get rid of psychological connectedness.

13. 'Persons, Character and Morality', in The Identities of Persons, edited by Amelie Oksenberg Rorty, (Berkeley, California, 1976), p.206.

14. 'Fundamentals of the Middle Way', translated by Frederick J. Streng in his Emptiness: A study of Religious Meaning (Nashville, 1967), p.189.

15. Erich Fromm, Greatness and Limitations of Freud's Thought (London, 1982), pp.66-67.

16. The Dhammapada, translated by Narada Thera (London, 1972), stanza 421.

17. Stanza 92.

CHAPTER FOUR

THE BODILY THEORY AND THE THEORY OF THE BODY

Because of the problems encountered with both the strict theory and the psychological theory, one may feel inclined to view the human body as constituting the essence of personal identity. For if there is no indefinable self to be the carrier of our experiences, and if there is no principle of identity to be gleaned from the nature of our psychology, all is not lost; for at least we have a body. But what is it about having a body that might dispose us to think it a plausible candidate for the basis of personal identity? The answer seems plain: the body is a physical object which, as long as it exists, is spatio-temporally continuous throughout the different moments of its existence. In consequence, myself of today can be said to be the same person as myself of twelve years ago so far as my body of today is spatio-temporally continuous with my body of twelve years ago. Exponents of this view are not, of course, denying

that over time a person's body will or may undergo various changes; rather they are claiming that so long as these changes occur within a body which maintains a spatio-temporal continuity, then the identity of the person whose body it is will be ensured.

The bodily theory of personal identity therefore equates personal identity with what Locke calls the identity of man. And Locke, as we have seen, considers this equation to be a confusion. Locke's distinction between the two types of identity is brought out in his story of the prince and the cobbler. Here, it will be recalled, Locke imagines the consciousness of a prince being transferred to the body of a cobbler. Were this to happen, Locke feels that the identity of the subsequent person would still be that of the prince even though the identity of the man or body would remain that of the cobbler. Hence, personal identity cannot be the same thing as bodily identity.

This example, as would be expected, has met with numerous protests. Williams, for instance, tells us we should not take the possibility of bodily transfer for granted and that there are even 'logical limits' concerning what we would be willing to say about these cases. Considering the suggestion that an emperor's personality could be transferred to the body of a peasant while the peasant's personality could be transferred to

the body of the emperor, Williams points to the difficulties in understanding what this could mean. For instance, what has become of the bodily function of the voice: 'How would the peasant's gruff blasphemies be uttered in the emperor's cultivated tones, or the emperor's witticisms in the peasant's growl?' Or again, maybe the emperor's face is just not the sort of face that could form peasant-like grimaces; and equally, maybe the peasant's face lacks the lineaments necessary to produce the emperor's royal smile. And these inabilities, Williams feels, are not just empirical; 'such expressions on these features could be unthinkable'.(1) Williams sees these considerations as bearing on the issue of bodily transfer in two ways. First, it shows that although it may seem that the idea of bodily transfer makes sense, there are many cases in which it is not conceivable. Secondly, it shows that when we think of distinguishing an individual's personality from his body, we are not sure what we are meant to be distinguishing from what.

This attempt of Williams' to discredit the notion of bodily transfer does not seem very convincing. It may seem too obvious of a reply to remind Williams of George Bernard Shaw's Pygmalion where a peasant does come to speak in cultivated tones, but I cannot think of what other sort of reply would be necessary here. For Williams' argument seems to consist of little more than the flat denial of the possibility of speech

therapy. But if speech therapy is a possibility, then we must admit that even a peasant could acquire cultivated tones. And if a peasant could do this, then certainly he could also project blasphemies in these newly acquired tones. Likewise I find nothing inconceivable about royal witticisms being emitted in a peasant's growl (Do peasants really growl?). Wit after all is the ability to manipulate language in an unexpected and humorous way, not the ability to elocute in certain tones. And since we have no difficulty imagining that because the emperor has contracted a throat problem he must now produce his witticisms gruffly, we likewise should have no difficulty conceiving of the emperor's witticisms being uttered in the rough voice of the peasant. Concerning his remarks about facial expressions, I agree with Williams that certain expressions on certain features could be unthinkable. But this does not prove what he thinks it does. For it is not unthinkable that by suffering a stroke, surgery, or what have you, the emperor could come to lose his ability to produce his old royal smile. And if it would not be said on this account that the emperor has lost his identity, then it cannot be held that the peasant's inability to smile royally has much bearing on whether or not his body has taken on the emperor's personality. And this applies mutatis mutandis to the problem of the emperor's face being unable to make peasant-like expressions.

Our reply to Williams then must be that he has not shown

that bodily transfer is in many cases inconceivable. There may yet remain some cases in which the limits of our conceptions are strained, but it seems plain that whatever the cause of our doubts in these instances, it could not be for considerations of voice production and smiling.

As for Williams' further conclusion that in attempting to distinguish between an individual's body and personality we are not sure what we are to distinguish from what, this too seems quite unfounded. When the emperor, who was at one time a generous and out-going man, eventually becomes selfish and withdrawn, most people, it seems, would be willing to say that his personality has changed. Few, however, would want to say that because of his newly gained selfishness and introversion his body has changed. Of course, his selfishness will be expressed and made known to us through the actions of his body and the content of his speech, but it seems hardly correct to describe the appearance of these new behaviours by saying his body has changed. When we say that someone's body has changed what we normally mean is that some physical feature of his body such as posture, weight, shape, or colour has altered. These physical changes may be the result or even the cause of certain personality changes, but they are not to be equated with personality changes.

A more cautious and sophisticated approach to the question

of bodily transfer is offered by Penelhum in Survival and Disembodied Existence. (2) Penelhum suggests that since we have probably never come across a situation of putative bodily transfer, just what we are to say in such circumstances does not involve making explicit a decision we already hold implicitly; it rather involves making a new decision. The problem however is that in making this new decision we are drawn in two different directions. For were we to come across a cobbler who claimed to remember doing the things that the prince had done, and a prince who claimed to remember doing what the cobbler had done, our instincts, says Penelhum, would incline us to say bodily transfer had taken place. And yet the normal convention of requiring that a person's body be present at the event he claims to remember indicates otherwise. And this problem cannot be solved by merely overriding the bodily requirement in favour of the memory criterion. For according to Penelhum, although memory is a criterion of identity - so far as having X's memories deductively implies being X - the memory criterion is parasitic on the bodily criterion. For in order to decide whether someone's memory is real rather than delusive we have to be able to apply certain physical tests. We have to be able, for instance, to show whether someone's body was present at the event he claims to remember. And if these tests turn up negative, says Penelhum, we would be entitled to reject his memory-claims.

Penelhum, however, accepts that in the story of the prince and the cobbler there is at least some body present at the remembered events even if it is the wrong body. When this is taken together with the fact that the memory-claims being made are so systematic and consistent, Penelhum feels it is quite permissible to say that bodily transfer has occurred. However, in saying this we would have allowed for a weakening of the priority of the bodily criterion over the memory criterion. The two criteria would then be on par.

But we are not forced to accept this account of the story; for there is another interpretation which enables us to retain the primacy of the bodily criterion. This interpretation would involve the introduction of a new concept which would allow us to speak about one person being able to remember the past of another, as for example, with someone who physically appears to be the prince and yet claims to remember the various happenings of the cobbler's past. The reason why we need a new concept, says Penelhum, is because our present concept of remembering does not extend to instances of one person's remembering another person's past. Of course we allow that one person can remember another's past so far as he was able to observe that person in the past and thence recall it; but we do not allow that one person can remember doing another's actions or having someone else's experiences. To allow this, Penelhum tells us, would be to 'pass the bounds of possible linguistic

legislation'; for it would permit the making of statements which would commit us to 'the possibility of some past action or experience belonging uniquely to two different people, and this is self-contradictory'.(3)

Penelhum therefore proposes the concept of retrocognition to refer to those instances which, because of their accuracy, first-person past-tense format, etc., appear to be memories, but because of their connection to the wrong body cannot be memories. The introduction of this concept, says Penelhum, warrants us to maintain the primacy of the bodily criterion of identity even in the presence of purported cases of bodily transfer. Thus when faced with the tale of the prince and the cobbler we are not driven to conclude that bodily transfer has actually taken place, a conclusion which would force us into the camp of the psychological theorists. We may choose instead to retain the primacy of the bodily criterion and say that both the prince and the cobbler retrocognize one another's pasts.

The introduction of the concept of retrocognition is an attempt to save the primacy of the bodily criterion. But is the bodily criterion really primary? The reason given in support of this contention is that memories cannot be self-authenticating. If a person's memories are to be genuine, then it must be that the person's body was physically present at the event he claims to remember. The concept of memory is therefore dependent on

the concept of a physical body. It should not take us long to see that this is essentially a bodily theorist's version of Butler's critique of the memory theory. Butler, it will be recalled, says that personal identity cannot be defined in terms of memory, because memory presupposes personal identity. Penelhum is saying the same thing. The difference is that for Butler personal identity refers to the identity of an indefinable self, while for Penelhum it refers to the body. This is a major difference, but the principle of both arguments is the same. Both claim that to distinguish a veridical memory from a delusive one, we must be able to refer to persons independently of memory. This, as was shown in the second chapter, is false. There it was argued that the way to demarcate a veridical memory is not by reference to a self which was present at the remembered event, but rather by reference to the causal history of the memory. For a veridical memory can be explained in terms of a memory experience which is causally connected to a veridical perception in accordance with the principles of the psychology of memory. And just as there is no need for a self to enter into this explanation, so there is no need for the mention of a physical body.

At this point someone might want to object that although the notion of the causal history of a memory may allow us to skirt the obligation of making reference to a self, it does not allow us to do the same with the body. This, it could be

argued, is because one of the elements in the causal chain which leads from perception to memory is the brain; and the brain is at least part of the body.

The problem here is that it is not at all clear that the causal history of a memory must make reference to the brain. In the psychology of memory the mechanisms of memory transfer from the sensory-buffer to short-term and long-term memory, the duration of retention and rates of decay, the effects of rehearsal on recall, and so on, are all given within a purely psychological framework.

Secondly, although brains are indeed parts of bodies, brain identity is not the same thing as body identity. This can be seen by considering Shoemaker's example of two people who have their brains transplanted into each other's bodies. If Brown, who witnessed and remembered event E, later has his brain removed and placed in the body of Robinson, which was not present at event E, then were we to ask this resulting person (Brownson) whether he remembered E, the answer most likely would be Yes he did remember E. But were we to apply Penelhum's test of checking whether the body of the individual who claims to remember E was present when E transpired, we would have to admit the results were negative; for the body of Robinson (which now houses Brown's brain) was not present at E. Of course the brain which has been newly installed in

Robinson's body was present at E, but this can only be of significance to the bodily theorist if he is really a brain theorist; that is, if he is willing to hold that the identity of a person goes with the brain, not with the body.

But why would anyone want to hold it is the brain which governs the identity of the person? The answer, it seems, is because the brain is the vehicle of memory and psychological continuity. And yet to hold the brain is important on this account is, in the end, to admit the primacy of the psychological criterion.

We should note here that it is the possibility of brain transplantation which has led Williams to rethink his views on the fission of the self. In his early papers 'Personal Identity and Individuation' and 'Bodily Continuity and Personal Identity' Williams attacks the psychological theory on the grounds that psychological continuity is not, like identity, a one-one relation. In his example, both Charles and Robert could be said to be psychologically continuous with Guy Fawkes; they could not, however, both be identical with him. For in that case they would be identical with each other; and two separate persons cannot be identical. In the latter of these two papers Williams admits that physical bodies are also vulnerable to the possibility of reduplication. That is, just as we can imagine the memories and character traits of one past person

simultaneously being displayed by two present persons, so we can also imagine a person 'splitting, amoeba-like, into two simulacra of himself'.(4) Although some philosophers, such as Quinton,(5) see this re-appearance of the possibility of fission on the bodily plane as indicating a weakness with Williams' argument against the psychological theory, Williams himself claims there is a critical difference between psychological and bodily fission. This is because although an enquiry into the history of a person's psychological continuity need not reveal that fission had taken place, i.e. that another person with an exactly similar psychological continuity had branched off from the person now under scrutiny, a similar enquiry into the spatio-temporal continuity of a person's body would be bound to reveal any bodily splitting and reduplication. In this latter case, says Williams, the spatio-temporal continuity of the body would have been interfered with by the process of fission. It therefore would be wrong to identify either of the resulting bodies with the original pre-fission body. Consequently, the bodily theorist is not forced to the absurd conclusion, as is the psychological theorist, that the post-fission individuals are identical with each other.

In his further paper, 'Are Persons Bodies?', Williams admits that Shoemaker's operation provides a persuasive counter-example to the view that bodily identity is a necessary

condition of personal identity. 'If judgements of personal identity in such a case', says Williams, 'went the way of the character- and memory-traits, as it seems they reasonably might, we would have here a divergence from bodily identity; since clearly it would be absurd to suggest that what governs the identity of the body is the identity of the brain' (my emphasis).(6) One of the virtues of Shoemaker's example, we are told, is that it avoids the problem of reduplication. For since it cannot be that two separate individuals will at the same time end up having Brown's brain and so be psychologically continuous with Brown, so it cannot be that two separate individuals will at the same time satisfy the conditions for being identical with Brown. It would seem that the possibility of brain bisection presents a problem here; but Williams replies merely that the bisection of the brain is in principle no different from the splitting of the body. And if a splitting of the body constitutes a loss of bodily identity, then likewise will the bisection of the brain amount to a loss of brain identity. Consequently, even if the recipients of Brown's hemispheres both display the memories and character-traits of Brown, neither of them will be identical with Brown since neither of them has precisely that object that was Brown's brain.

We thus see a fundamental change in Williams' views on personal identity. In his earlier papers all that could be said

about someone who acquired the memories and character traits of an earlier person was that he clairvoyantly or mysteriously knew all about the earlier person. Consequently, personal identity could not be ascribed on the basis of memories and character-traits. In 'Are Persons Bodies?', however, it is allowed that ascriptions of personal identity might reasonably go the way of memory and character traits. This change of views would seem to be because of Williams' belief that the Shoemaker case - which is said to be one persuasive counter-example to the bodily theory - avoids the problem of reduplication.

But it now becomes apparent that Williams has fallen into confusion. For if the ascription of personal identity goes the way of memory and character-traits in the case of a straightforward brain transplant, why cannot the ascription go the same way in the case of a dual transplant of a bisected brain? To answer that the hemispheres of the post-operative individuals have suffered fission and so are no longer identical with the pre-operative brain is a non sequitur. For it has already been granted that personal identity is to be ascribed according to memory and character-traits. So whether the memory and character traits are produced by a fully intact brain or spatio-temporally ^{whose halves are in informationally equivalent states,} discontinuous brain halves ^{is} really beside the point. And since it is likely that the two persons each fitted with a half of Brown's brain would both make the same memory-claims and show the same character traits, it must

be concluded that the Shoemaker case does not, as Williams thinks it does, avoid the problem of reduplication.

It is curious that Williams resists the possibility of fission in the case of brain transplants, and yet goes on to allow it in what he says is a 'very natural extension' of Shoemaker's example. The extension Williams has in mind is the transfer of information between brains. Thus instead of imagining the physical transfer of brains we might equally imagine extracting the total information content from a brain, encoding the information in a storage device, and later returning it to the same or another brain. If this procedure were carried out on someone, say for the purposes of brain surgery, and the person were to recover all the memories he had before the information removal, Williams feels 'we should not dream of saying that he did not, at a later stage, really remember' (p.79). He admits that, normally, causal claims which go outside the body do not constitute genuine memories. Were I to rely completely upon my diary to re-learn the experiences I had some years ago, then just because of this reliance, I could not be said to really remember those experiences. But in the case of brain-state transfer Williams thinks things are different. For the replacement of information, we are told, is not 'as such, 'learning again''.

And with this the road is open to the problem of

reduplication. For rather than returning the information to only one brain, says Williams, we could just as easily feed the information into numerous brains. If this were to happen, there would then exist many individuals all of whom would have the same memories and character traits, and consequently, all of whom would be identical, both with the earlier person and with each other.

Now, for the author of 'Personal Identity and Individuation', and 'Personal Identity and Bodily Continuity' this conclusion would have been enough to dismiss the claim that personal identity could be preserved through brain-state transfer. For in those papers it is insisted that identity is a one-one relation. But in 'Are Persons Bodies?' we are told there is a sense in which all the recipients of one person's brain-state could be both identical with that person and with each other. These persons would all be the same person in a type sense of 'same person': 'If the prototype person was a Smith, all the resultants would be, significantly, Smiths; and 'person' like other type-token words, will have a plural differently applicable under two different principles of counting - a room containing two Smiths and three Robinsons will contain, in one sense, five, in another, two, persons' (p.80).

On this account there would be two senses of 'person':

type- persons and instances of type-persons; that is, token- persons. A type-person states Williams, would be something like a class of bodies all of which exhibit the same total brain-state, i.e. have the same memories, character-traits, and so on. When using the type sense of 'person' it would make perfect sense to say that today you had met the same person you met yesterday even if the person you met yesterday has a numerically distinct body and brain from the person you met today. As long as the two bodies you had encountered had the same total brain-state, then they would be the same type-person. It is thus clear that, in the type sense, persons are not bodies; for the identity of a type- person is not tied to the identity of any particular body. However, Williams does not feel that this necessarily means that token-persons are not bodies. For any statements about personal identity which were not reducible to statements about bodily identity, says Williams, 'could be adequately represented in terms of bodies belonging to the same or different person-types' (p.80).

But what raises our suspicions about this conclusion is that if the identity of person-types diverges from bodily identity, then any statement about personal identity which makes use of the notion of person-types ipso facto cannot be equivalent to statements about bodily identity. Of course, what Williams says is that such non- reducible statements could be represented in terms of bodies belonging to certain person

types. But we should not be fooled by the presence of the word 'bodies'; for we still have to decide what person-type each body will belong to; and here again we are forced to leave the realm of bodily identity.

We ought now to ask whether the notion of a type-person makes sense. This can best be done by returning to Williams' assertion that if someone were to have brain information removed and then later replaced, we should not dream of saying afterwards that he did not really remember. The reason given for this assertion is that although the causal route has gone outside the person's body, and although this would normally count against the memories being genuine, in the case of brain-state transfer, the acquisition of memories is not 'as such, 'learning again'', as it would be in the case of, say, reading a diary. It is this feature which allows memories acquired through brain-state transfer to be genuine.

• 4

It is probably significant that in describing this redeeming feature of brain-state transfer Williams qualifies himself twice: first by the use of 'as such' and secondly by putting quotation marks around 'learning again'. Why does he not simply say that the replacement of information is, straightforwardly, not learning again? The answer, it seems, is because Williams himself is aware that the distinction between brain-state transfer and learning is not all that clear.

Consider his example of learning anew about one's past by reading what one had written earlier in a diary. A person in this situation is re-acquiring information that had at an earlier time been placed in storage; that is, in writing. Thus there is a sense in which a diary is also a brain-state transfer device: it serves to record and transfer the brain-state of the writer to that of the reader. Of course, the transfer of information will not be as thoroughgoing as it would with Williams' imagined device which records the entire state of the brain, but I doubt that this is a philosophically significant difference. What is important is that the device succeed in having the person come to have the same memory experiences prior to their having been placed in storage. There is no reason in principle why a diary cannot fulfil this function as well as Williams' machine. It is true that someone who had come to know about his past by reading a diary might be aware that he had gained his memory experiences in this abnormal fashion and so remain somewhat sceptical or oddly disposed towards them. But this is equally true of the brain-state transfer patient.

Maybe Williams thinks there is something involved in the act of reading the diary which is lacking in a case of brain-state transfer; something which brings reading, but not brain-state transfer, under the description of learning. What could

this be? If it is that the information acquired by reading is acquired by the subject's purposely engaging in an activity, while that acquired by brain-state transfer need not involve any effort or intention on behalf of the subject, then we can alter the example suitably. We can imagine that rather than reading the diary himself someone else reads the diary to the subject. Or again, he might simply be made to listen to a tape-recorded reading of the diary. Or, better still, the contents of the diary could be encoded into a machine which feeds directly into the visual or auditory centres of the brain, putting it into exactly the same state as it would be if the subject had read or been read the diary. At what point do we say that the subject is no longer 'learning again' but is now actually remembering? This problem is one which Williams does not bring himself to face.

• 4

Just why Williams introduces the notion of a type-person is further obscured by the fact that in the closing paragraphs of 'Are Persons Bodies?' he himself offers something of an argument against the idea. Here it is suggested that if someone loved a token-person just because she was a certain type-person, say, because she was a Mary Smith (i.e. one of the possibly numerous brain-state transfer replicas of Mary Smith), then it would be unclear that what he really loved was the token-person. What he loves is the type Mary Smith. 'We can see dimly what this would be like' says Williams:

It would be like loving a work of art in some reproducible medium. One might start comparing, as it were, performances of the type; and wanting to be near the person one loved would be like wanting very much to hear some performance, even an indifferent one, of Figaro - just as one will go to the scratch provincial performance of Figaro rather than hear no Figaro at all, so one would see the very run-down Mary Smith who was in the locality, rather than see no Mary Smith at all.

Much of what we call loving a person would begin to crack under this, and reflection on it may encourage us not to undervalue the deeply body-based situation we actually have. While in the present situation of things to love a person is not exactly the same as to love a body, perhaps to say that they are basically the same is more grotesquely misleading than it is a deep metaphysical error; and if it does not sound very high-minded, the alternatives that so briskly grow out of suspending the present situation do not sound too spiritual, either. (p.81)

Because of a (seemingly purposive) lack of decisive statement, it is difficult to put our finger on what exactly Williams is arguing for here. What he seems to be saying, however, is that although it is not inconceivable that someone could love a type-person (we can see dimly what it would be like), such a love would not fit in with our normal concept of love. This is because love is basically the love of a body, i.e. a token-person.

If this is true, then it would appear to constitute another argument for the bodily theory of identity. For if to love a person is just to love a body, then it seems a person

is just a body.

We must therefore turn our attention to the question of what is involved in the notion of loving a person. But before we set out on this exploration we first ought to clarify some of the peculiarities surrounding the use of the word 'love'.

It is a cause for some wonder that in English the word which we use to denote one of our deepest and most meaningful experiences can also be used to describe our relation to various activities or objects to which we are merely positively disposed. Thus I may say both that I love someone and that I love playing cards or love a certain vase. Despite the use of the same word in each of these instances there is, however, an important difference which separates the first use of the word 'love' from the other two. For although in each example 'love' functions as a transitive verb, only in the first example can it also function as an intransitive verb. Only when I love someone can I be said to be in love. It is true we can take the intransitive expression 'in love' and, making it transitive, say such things as 'I am in love with this vase', but in these cases - where what we claim to be in love with is not a person - it is evident that 'in love' is being used metaphorically as a convenient hyperbole. The proof of this is that when it is left in the intransitive form, as when I declare simply 'I am in love', no one doubts that what I am in

love with is a person.

Why then must it be a person with whom I am in love? Why cannot I be, strictly speaking, in love with a vase? The most obvious answer to this question is because only a person is capable of loving me back. But why should it matter to me that that with which I am in love is capable of loving me back? Plainly, because in being in love with another I want that person to love me back. This is a point well made by Jean-Paul Sartre when he says 'love is the demand to be loved'.(7) We have to, however, be careful here. For being in love is not just the demand to be loved: I may demand to be loved and yet not love the person on whom I place this demand. I may, for example, want to be loved merely because it feeds my narcissism. But in such an instance I am clearly not in love with the person whom I want to love me. Further, although being in love may be to want the one I love to love me back, this still does not explain what love itself is. To understand what being in love is we must therefore go straight to the question of what love is.

This is naturally a momentous question, one which I cannot hope to answer fully here. I think, however, we shall be on the right track if we see that at the very core of love lies a complex of desires involving the desire to be vulnerable before another person in order that I may be nurtured or cared

for by that person, and, at the same time, the desire to have the other person vulnerable before me in order that I may nurture or care for that person.

This account would be supported by Freud's observations that the desires of the adult are to a large extent built up from earlier infantile impulses.(8) For the desire to be vulnerable before another in order that I might be cared for would seem to be a desire to recreate the infant's state of complete vulnerability before the care-giver. The reciprocal desire to care for another would probably be a later development coming from an identification with the care-giver.(9)

It might, however, be felt that invoking desires for vulnerability is out of order here. Could not love be simply the reciprocal desires to care and be cared for? The problem with this is that once we remove the desires to care and be cared for from their corresponding desires for vulnerability, we shall be unable to explain the importance that we place on intimacy and trust; both of which seem central to our idea of love. For I can easily want to care and be cared for without wanting to become intimate. I might, for example, want to care for Mary because I feel sorry for her, and want her to care for me because I feel sorry for myself. To be intimate with each other, however, we must place our inner-most workings in each

other's hands; that is, we must disclose ourselves to each other and enter into a state of mutual emotional, psychological, and physical vulnerability. Here too we see the importance of trust. To engage in such a relationship we must feel we are able to trust each other. While trust may bear some relation to the notion of caring, it does not seem nearly as central as it is to the notion of vulnerability. For while it is understandable that I might want to care for someone, even if I do not trust the person, it is not quite as understandable that I would want to be vulnerable to someone I do not trust.

The analysis of love into a complex of simpler desires is useful because it puts us in a position to explain various other interpersonal orientations that seem intimately affiliated with love. This can be achieved by eliminating or altering different parts of the complex.

Thus dependency would be constituted by only half of the complex of desires that make up love. What a dependent person wants is to be vulnerable before someone who will care. He has little desire to have the other person vulnerable before him. And further, he does not want to care for the one who will care for him. This lack of desire for a reciprocal relationship also appears in the compliment of dependency; that is, in paternalism. The paternalistic person is both dominant and nurturing. He is dominant because he refuses to enter into a

state of vulnerability or to be cared for, and he is nurturing because he desires the other person's vulnerability in order that he may show care towards that person.

Care, however, is only one way of responding to vulnerability; one can also respond with hostility. Thus someone might desire to have another person vulnerable before him in order to harm or control that person. When this occurs together with a lack of desire for any reciprocal relations, then we have sadism. Because sadism is often thought to be essentially the desire to inflict pain or harm on another person, it might appear that the desire for the other person's vulnerability is not basic to sadism. But it should be evident that with sadism the desire to inflict pain is intimately wedded with the desire for the other person to be vulnerable. This would explain the importance the sadist places on controlling another person; for to control a person is to have that person vulnerable before the one who is in control. There are, of course, those cases in which the desire to cause harm appears unaccompanied by the corresponding desire for the vulnerability of the other person. But when this takes place we seem to have a case of necrophilia rather than of sadism. As has been suggested by Fromm, the distinction between sadism and necrophilia is the distinction between the desire to control and the desire to destroy.(10) The necrophiliac's interests, then, skirt round the issue of vulnerability. His concern,

which may manifest itself as sexual desire for a corpse, lies rather with injury and destruction. When seen from our theory then, necrophilia turns out to be a less complex phenomenon than does sadism.

In the same way that dependency is the complement of paternalism, so is masochism the complement of sadism. For, masochism, like sadism, is a mobilization of only part of the complex of desires that make up love. But whereas the sadist desires the vulnerability of another person, what the masochist wants is to be vulnerable himself before another person. Further, like the sadist, the masochist is one who rejects the notion of care. He is not interested in being vulnerable in order that he might be cared for; his vulnerability is engaged in order that he may be harmed or controlled.

It must be stressed that these interpersonal orientations, including that of love, are ideals or prototypes which different persons will normally approach only more or less. It will more often be the case that someone may love another and yet lean slightly more towards, say, a dependency orientation, i.e. he will desire to be vulnerable in order that he may be cared for more than he desires the other person to be vulnerable or more than he desires to care for that person. Or, again, although someone may display a predominantly paternalistic orientation, his paternalism may nonetheless be

coloured with shades of sadism. Just what it is that leads a person to adopt more of one orientation rather than another would then be an area of enquiry for the psychology of individual differences.

We are now able to see what is wrong with Williams' claim that to love a person is basically the same as to love a body. For to love a person is to engage in a network of desires which necessarily involves reference to more than just the person's body. Firstly it involves reference to the person who is doing the loving: one of the things that makes up my love for another is, as we have seen, the desire to be vulnerable before that person in order that I may be cared for. Secondly, since to love is to want to be loved, then another component of my loving someone is my wanting the person I love to have the same desires towards me that I have towards her.

It is this reciprocal feature of love, this wanting to be vulnerable myself, wanting to be loved back, etc., which allows us to draw a fundamental distinction between loving a person and loving an activity or object. For in loving a vase, for example, I have no desires concerning the vase's desires; for I know that vases do not have desires. And herein lies the error of Williams' claim. To say that loving a person is the same, even basically the same as loving a body is to put the love for a person on par with the love for an object; for bodies, like

vases, are physical objects. What makes loving a person radically different from loving an object is that in the former case I am 'in love', in the latter I am not.

None of this is to say that the body does not enter into the concept of being in love; for an important dimension of this form of love is sexual desire, and this is a desire intimately concerned with the body. This, however, should not present a problem for the theory of love just given. For sexual desire, it seems, can be easily accommodated within the theory as but one more way of mobilizing a complex of reciprocal desires for vulnerability and care. As Maurice Friedman says in his book on The Hidden Human Image, 'In attitude as well as in physical fact sex means that posture of the ultimate baring of one's self', or as he also says, a 'mutual baring'. (13) Thus in sexually desiring someone one of the things a person wants is to bare his body, especially the erogenous areas of his body, to another person, while at the same time having the other person's body similarly bared to him. But why does someone desire this mutual baring of bodies? In order, it would seem, to both caress and be caressed. This would explain why sexual intercourse is often thought to be the epitome of the sex act. For in sexual intercourse one of the things that transpires is a mutual baring of and caressing of and by the most sensitive and hence vulnerable of the erogenous areas.

It seems reasonable to argue that in desiring to bare one's body in this way before another person one is, at the same time, desiring to be vulnerable before that person. A person in this state of total physical bareness is vulnerable simply because he has nothing left to hide. He can no longer pretend: his body is presented to the other person for all that it is: nothing more, nothing less. And the same could be said about the reciprocal desire for the baring of the other person's body.

Likewise, just as the desire for the mutual baring of bodies is a form of desire for mutual vulnerability, so is the desire for mutual caressing a form of desire for mutual care. For in desiring to care for another person one of the things someone wants is to satisfy that person. And this is also true for the desire to caress someone. One can of course attempt to caress another in order to satisfy oneself rather than the person one is attempting to caress, even as one can attempt to care for another purely for one's own sake. But just as care thus imposed does not amount to real care, so does the caress thus attempted fall short of a real caress. To really caress, rather than merely attempting to caress, the caress must be done in response to the other person's receptivity, i.e. desire for vulnerability. Only in this way is the attempted caress a genuine caress rather than an intrusion or an assault. All of this would apply mutatis mutandis to the reciprocal desire to

be cared for.

The idea that sexuality is connected with the notion of vulnerability finds expression, I think, in what anthropologists call myths of vagina dentata and penis aculeatus; that is, myths of the toothed vagina and the stinging or stabbing penis. Thus in Hindu mythology we find Siva, the god of destruction (and also the ithyphallic god), often engaged in destroying demons with his phallus. Similarly, other mythical beings use their vaginas in an attempt to kill, swallow-up, or castrate their enemies. In one myth, for example, a demon sneaks into Siva's abode: 'Once inside, he took the form of Parvati [Siva's consort] in order to deceive Siva, and he placed teeth as sharp as thunderbolts inside her vagina, for he was determined to kill Siva. When Siva saw him he embraced him, thinking him to be Parvati'. Eventually, however, Siva 'recognized the magic form of the demon, and he placed a thunderbolt in his own phallus and wounded the demon with it, killing him'. (12) Surprisingly enough, a version of the vagina dentata myth, namely, that some women can trap a man's penis within their vagina, persists quite openly within our own culture. (13)

That sexual desire plays an important role in our idea of being in love should not however, lead us to infer that one need be in love in order to have sexual desire. As is well

known, sexual desire operates quite happily outside the domain of love. The reason for this seems best explained by the fact that since sexual desire is primarily concerned with the body, it need not make reference (as love does) to another person's desires. So although love is the demand to be loved, sexual desire is not the demand to be sexually desired. I may want to be sexually desired (especially if my sexual desire occurs within the context of being in love), and being sexually desired might even heighten my own sexual desires, but this would not be basic to the structure of sexual desire in the way that the desire to be loved is basic to the structure of love. What this means is that to sexually desire another person it is enough to desire to be bare for another person's caress and to desire that person to be bare for my caress. That sexual desire works in this unilateral fashion is well-supported by the existence of sexual desires which, far from being directed towards another person's desires, are not even directed towards a person; as, for example, in the case of fetishes and zoophilia.

At the beginning of this chapter it was suggested that the reason why the body might seem a plausible candidate for the basis of personal identity is because the body has spatio-temporal continuity. On this account the identity of the person would be fundamentally no different than that of a physical object. Thus to ask whether a person of now is the same as a

certain person of twelve years ago will, in principle, be no different from asking whether a particular painting now hanging above the mantel is one and the same as one that was placed there twelve years ago. What determines whether the two paintings are the same is if there exists a continuous path in space and time which connects the earlier painting with the later one. What determines whether the two persons are the same is if there likewise exists a path in space and time which connects the earlier person's body with the later person's body.

There is, however, an important disanalogy here which bodily theorists seem to have overlooked. For although there is no particular painting which I experience in an essentially different way from other paintings, there is one body which I experience quite differently from other bodies. This is the body I call mine. Because of this it will turn out that although I might be able to apply the criterion of spatio-temporal continuity to the bodies of other persons (for it was only with a view to the bodies of other's that this account was invented), when it comes to my own case no such criterion can apply. As a consequence, the bodily theory of identity is one which I can never use to give an account of my own identity. And this, it seems, is a good enough reason for us to reject it.

To introduce the idea of a basic distinction between my body and the bodies of others, let us examine a list given by John Hospers. The features which distinguish my body from others, says Hospers, are at least these:

(1) It is the only body I cannot get away from - it is always there whenever I am conscious, and I can never see it walking away in the distance. (2) I can see it, unlike other bodies, only from certain perspectives: I cannot see its face except in the mirror, or the back of its head except in two or more mirrors, and its chest and shoulders (for example) always appear in about the same place in my visual field, at about the same apparent distance. (3) It is the only body of which I have kinesthetic and other somatic sense experiences; of other bodies I can have visual and tactile experiences, but not kinesthetic. (4) Most important, it is the only body I can directly control. I can decide to raise its arm (the arm of the body I call mine), and the arm raises; the act follows upon the decision. I cannot control any other body in this way, but only indirectly via command or physical force. I can also alter the positions of things in the physical world, as in moving a chess piece from this square to that, but I can move these other things only by moving this body: my influence on the outside world is always by means of this body. (14)

Hospers is not, of course, saying that only in his case will all of this be true; as if there were something singularly unusual about his own body. Rather he is saying that these propositions will be true for each person in the case of each person's own body. Thus while it is true for me that I can never see my body walking away in the distance, it is also true for you that you can never see your body walking away in the distance, and it is equally true for that person

over there that he can never see his own body walking away in the distance, and so on. A corollary of this is that what is here true for each person of his own body will not be true of his body when it is experienced by someone else. Thus although I cannot see my body walking away in the distance other persons can have this experience of my body.

Unfortunately, Hospers does not attempt to integrate these observations into a coherent theory of the body. He merely lists them as peculiar but apparently unconnected features of the body I call mine. It is understandable then that certain difficulties and omissions occur within his otherwise intriguing list.

We can start by noting something of an inconsistency between (1) and (2). For (1) tells us that my body is the only body I cannot get away from, and it is always there whenever I am conscious. But (2) says that I can never see my face or head, except with the use of mirrors. It seems, therefore, that if (2) is true, then (1) cannot be completely true. For according to (2) there are parts of my body that are, in one sense, never there whenever I am conscious: my face and head are never there for me as a direct visual experience.

Further, although Hospers has realized that my body is the only one of which I can have kinaesthetic and other somatic

experiences, and also that it is the only body through which I can influence the world, he seems to have overlooked the fact that it is the only body through which I can experience the world. This is an important omission; for it is this fact that helps to explain (2). The reason why I can see my body only from certain perspectives is just because it is through that same body that I must do the seeing. This is also why I cannot see my face; for it is from my face that my organs of vision scan the world. We should note, however, that although I cannot see my entire face, I can, in certain instances, see parts of it e.g., my nose, upper lip, and eyebrows. However, try as I may, I cannot see those parts of my face which do the seeing; that is, I cannot see my own eyes.

And this does not just hold for vision and the eyes, but also for other sense modalities and their related parts of the body. Thus what I perceive with my hand is not my hand itself but rather the environs in which it is placed. When I submerge my hand in a basin of water or pick up some sand, what I feel is just the fluidity of the water or the abrasiveness of the sand. Likewise, when I listen to a symphony what I hear is music, not my eardrums. And the same is true of smell and taste: when I smell and taste a glass of wine, what I smell and taste is wine, not my olfactory receptors and tastebuds.

This does not mean that I cannot experience my sense-

organs. I can, after all, use one sense-organ to explore another. I can touch my eyes or look at my hands. And I can even touch one part of my hand with another part of the same hand. But it is important to see that even on these occasions the sense-organ that is doing the sensing is not sensing itself. So if I touch the tip of my left hand's little finger with the tip of the same hand's thumb, what the tip of my little finger is feeling is the tip of my thumb, and, conversely, what the tip of my thumb is feeling is the tip of my little finger. In neither case are the fingertips sensing themselves. Similarly, even though it is conceivable that someone could have both his eyes situated in such a way that they might see each other, still neither eye would be seeing itself.

Although this all seems to have a bit of mystery to it - How can I have eyes and yet never see them? - when we consider the nature of perception, it is just what we should expect. For it is this 'non-reflexivity' of the sense-organs which allows us to perceive the world. Were my eyes to somehow see themselves they would be 'clogged' so to speak, with their own image and nothing else could 'get in'. Yet this very fact requires us to acknowledge a phenomenology of the body from two different perspectives: the body as it is from a first-person perspective and the body as it is from a third-person perspective. It must be emphasized, however, that this

dichotomy of perspectives is not the same dichotomy which sets apart 'the body as it is for the person to whom it belongs' from 'the body as it is for those to whom it does not belong'. For, as we have seen, I can in some ways take a third-person perspective on my own body: I can look at my hand or touch various parts of my body in the same way that I might look at someone else's hand or touch someone else's body. In this sense, the third-person perspective cuts across the distinction between my body as I experience it and my experience of other person's bodies.

With the first-person perspective, however, things are different. From this perspective I cannot perceive the body of another. For here the body is experienced as a point of view on the world, not as a spatio-temporal object in the world. This is why the hand which I look at, be it mine or someone else's, is not a hand from the first-person perspective. It is merely something on which I am taking a point of view. However, the hand which allows me to feel the fluidity of the water, the warmth of the air, or the texture of my other hand, is not a hand that enters into my perception of the world; it is my perception of the world. More precisely, it is my manual-tactile perception of the world. Thus, from the first-person perspective my body is a maze of centres of perception which constitute my phenomenological point of view on the world.

There is a possible objection here. For what are we to make of the somatic sensations in my hand? What of the kinaesthetic sensations of my fingers moving in relation to one another? As Hospers points out it is of my body alone that I receive kinaesthetic and other somatic experiences. And these experiences are not, like other experiences of my hand, available to others.

Moreover, although we have shown how I can perceive my body and yet do so from a third-person perspective, here the situation seems somewhat different. When, for example, I taste my lips or smell my wrist, what I am doing is using one of my sense-organs to explore another part of my body. But in kinaesthetic and somatic sense experience I am not directing any particular sense-organ upon any particular part of my body. I seem merely to be receiving sensations from my body. I can of course direct my attention to a certain area of my body.; but even here there seems to be no clearly demarcated sense-organ involved. Nor are the sensations thus attended to, except perhaps in cases of illness, that clear and distinct.

All of this may tempt us to conclude that such experiences of the body could not be from a third-person perspective. And if this is so then it appears we have a counter-example to the claim that the body of the first-person perspective is a point of view on something other than itself. For in the case of

kinaesthetic and somatic sense experience my body seems to be a point of view on just itself.

We should start our reply by pointing out that although kinaesthetic and somatic experiences are often classed together, they are in fact fundamentally different sorts of experiences. The criticism to which we are replying therefore demands different responses in each case.

In the case of somatic experience even though it may be true that I can have them of my body only, this does not mean they are experiences of my body from the first-person perspective. For as we have said it is not whose body it is that determines from which perspective it will be viewed; it is the manner in which it is perceived that is important. And if we attend to our somatic sense experiences we will see that, despite their being generally indistinct and not perceived by any specific sense-organ, the perceptual story is not altogether different from those instances in which we use one of our sense-organs to observe another part of our body. In both cases the object of our awareness is something we contemplate from a distance, so to speak; it is something upon which we take a point of view. Thus the fluttering in my stomach or the throbbing in my leg is just as much something upon which I can take a point of view as is my abdomen which I touch or my leg which I see. Of course the somatic experiences

I have of my body are quite different from the visual experiences I have of my body, but the same is true for those visual experiences when compared with the other sense experiences I might have of my body.

Turning to kinaesthetic experience we discover a rather different type of awareness. Although the term 'kinaesthesia' literally means 'sensation of movement' there is a certain type of awareness that we have when we are not moving that is much the same as kinaesthetic awareness. What I am referring to is the awareness of where my body-parts lie in relation to one another, i.e. the spatial distribution of my body. When, for example, I am seated in a chair I am aware that my hips are on about the same level as my knees, that my body is folded in the middle bringing my thighs into a certain angle with my chest, that my head is thus and so a distance from my feet, and so forth. What constitutes ^{sort of proprioceptive} this awareness of my spatial distribution is just an awareness that my body, being a system of sense-organs, is a collection of points of view upon the world. I know, for instance, that the point of view that is my hand is above the point of view that is my foot. That is, I am aware that the arm of the chair which my hand senses is above the floor which my foot senses. I have all this awareness without the sensation of movement. Were I, however, to get up from the chair my sensation of movement would then be engaged. But what could this sensation of movement be if not an

awareness of the shifting of the spatial relations that my body-parts bear to one another. When I rise from the chair my hips gradually become higher than my knees, the angle between my thighs and chest increases as does the distance between my head and feet. There is nothing new that arrives in my consciousness just in the instance that I start to move. In this sense, then, kinaesthesia is only an awareness of the shifting spatial distribution of the points of view that make up my body from the first-person perspective.

It is also from this point of view that my body is a point of action; that is, a point from which I can, as Hospers says, alter positions of things in the physical world. But this is not just one more peculiar fact about my body (and here we see the relation between (2) and (4)). The reason why it is the only body through which I can act is because it is the only body which I can perceive in certain ways, i.e. from the first-person perspective. When I throw a stone at a piece of wood floating down-stream, I do not perceive my arm as I would perceive your arm doing the same action. What I do is sense the weight of the stone, note the speed and angle of the driftwood's movement, and, keeping my eye on the target, let fly the stone. That is, I must take a point of view on the physical world in which I am acting. Were I to try to perform this action while observing my arm or body from the third-person perspective, viewing it as a physical object, the stone

might well hit my foot before it hits the piece of wood.

What we are left with then is a body that cannot be identified with any of the objects of perception. This, as we have seen, is because the body from this perspective is the point of view from which objects are perceived. We should not, however, be led into thinking that a point of view has some sort of existence which is independent of the objects upon which it is a point of view. For a point of view refers to nothing more than a certain perspectival ordering of things. It is as if the physical world were a pool and my body were a vortex in the middle of the pool. Because a vortex is literally a hole in the water, the existence of the vortex is logically dependent upon the existence of the water. The vortex is just a certain ordering or configuration of the water. Thus if we consider what my eyes are for my own vision we see they are merely the place around which all lines of perspective diverge; they are the perspectival opposite of that point in the distance to which things converge, i.e. the vanishing point.

Caution, however, is required here. For it is easy to confuse the notion of a point of view as we have just given it with the notion of something which, though not experienced as an object for perception, is nonetheless experienced 'implicitly': as a mediator or something through which objects appear to consciousness. This seems to be the error into which

Sartre has inadvertantly fallen in his discussion of the body. I say 'inadvertantly' because it is clear from Sartre's overall phenomenological position that the body from the first-person perspective cannot count even as an implicit object of perception.

According to Sartre, although my body is a point of view on which no point of view can be taken, it is nevertheless 'a conscious structure of my consciousness' (p.329). Thus, the body-for-me, as he calls it, belongs to the realms of unreflective consciousness; that is, it is something of which I am conscious without having to reflect upon it. Indeed, were I to reflect upon it and so make it into an object upon which I were to direct my attention, I would no longer experience it as the body-for-me, but rather as the body-for-others, i.e. as the body from the third-person perspective.

Sartre's idea that the body-for-me is something which is grasped unreflectively means that it is perceived in a manner not completely dissimilar to the gestalt notion of a ground. In Sartre's exposition of the gestalt theory the field of perception is organized in terms of figure and ground. The figure is that part of my perceptual field upon which I direct my attention while the ground is the undifferentiated mass of experience upon which the figure appears. Thus the ground is 'that which is seen only in addition, that which is the object

of a purely marginal attention' (p.10). This does not mean that my consciousness of that which is the ground is a weaker consciousness than my consciousness of the figure upon which I direct my attention, but only that the ground of my perceptual field is experienced as undifferentiated. Now the relation of figure and ground is such that a new figure is forever rising up from the ground to displace the present figure; in other words the direction of our attention is always shifting. An example of this process is given by Sartre in his account of looking for his friend in the cafe. In surveying the cafe for Pierre, says Sartre, 'each element of the setting, a person, a table, a chair, attempts to isolate itself, to lift itself upon the ground constituted by the totality of the other objects, only to fall back once more into the undifferentiation of this ground; it melts into the ground' (p.10).

Throughout this process of perception my body, like the ground, is also grasped unreflectively. It is not, however, grasped in the same unreflective way as the ground, i.e. as a totality of undifferentiation. For the body-for-me, says Sartre, is given only implicitly. It is what the world as ground indicates. But in what way can I experience my body if it is only indicated by something else? Sartre's answer is that I experience it as that which is surpassed. In an analogy given here he says that consciousness of the body is comparable to consciousness of a sign: 'Now the consciousness of a sign

exists, for otherwise we should not be able to understand its meaning. But the sign is that which is surpassed toward meaning, that which is never apprehended for itself, that beyond which the look is perpetually directed' (p.330). In this way then, though I experience the world through my body, in the process of doing so I nonetheless (unreflectively) experience my body. I experience it as something I neglect in order to experience the world.

But where is the evidence for this supposed experience of a body which is grasped unreflectively as surpassed? In the process of perception the existence of our unreflective consciousness of a ground is well supported. We need only turn to our experience of the world to see that, as Sartre says, a figure asserts itself on a ground of undifferentiation. But with the body things are quite different. When my sense-organs experience the world it is plain that they neither reflectively nor unreflectively experience themselves. It is true that, in a sense, the ground of my different perceptual fields 'indicate' my body so far as they indicate a point from which the world is being viewed. But this point, as we have seen, refers to nothing other than a certain perspectival ordering of the objects of my perception. It is not something of which I am conscious in addition to the ground. The sign analogy is, unfortunately, of no help here. For although I may not reflectively apprehend the sign when I am engaged in viewing it

to understand its meaning, I can easily change this situation and turn my gaze upon the sign qua physical object. That is, rather than looking at the sign to realize its meaning I can look at it in order to explore its physical features, say its fount or its calligraphy. And this shift of attention is carried out, significantly, within the same sense modality. But there is no analogous situation with body-consciousness. When I am visually apprehending the letter-opener in front of me I cannot suddenly shift my attention to the visual apprehension of my eyes; for there is nothing there that can be visually apprehended. It is precisely because there is nothing there for me to be conscious of that Sartre seems driven to argue, in Kantian fashion, that I must have this unreflective experience because something else is the case. Thus, after agreeing that the senses cannot perceive themselves reflectively he goes on to say 'Nevertheless, the senses are there. There is sight, touch, hearing' (p.316). But this does not show, as Sartre seems to think it does, that there must be some unreflective way in which the senses are there. The fact that the sense of sight is there for me is nothing over and above the fact that objects present themselves visually to me. And there is no reason why my senses must unreflectively experience themselves for this to be the case.

Having thus arrived at a theory of the body it now becomes clear why the bodily theory of identity had to fail. For the

bodily theory construes the body as a physical object that persists through time and space; that is, as a body from the third-person perspective. But the body from this perspective is not the body 'in which I live', it is rather the body I examine from the point of view of the body in which I live. But to live in a body, or as we might more correctly say, to live a body, is just to undergo the world from a particular point of view. And here the notion of physical object identity can get no foothold; for a point of view is not a physical object. It is true that the point of view which is my body is defined in terms of the physical objects on which it is a point of view. In this case we could say with Sartre that 'my body is co-extensive with the world, spread out across all things' (p.318). But such a thesis seems to go beyond what the bodily theorists originally had in mind.

NOTES TO CHAPTER FOUR

1. 'Personal Identity', p.12.
2. Chapter 7.
3. P.84.
4. P.23.
5. The Nature of Things (London, 1973), p.95.
6. Problems of the Self (Cambridge, 1973), Chapter 5, p.77. Further references to this edition are included in the text.
7. Being and Nothingness: An Essay on Phenomenological Ontology, translated by Hazel Barnes (London, 1957), p.375. Further references to this edition are included in the text.
8. The Standard Edition of the Complete Psychological Works, Volume 7, Three Essays on the Theory of Sexuality, edited by J. Strachey (London, 1964), p.233.
9. By as early as eighteen months children can be seen to imitate the care-giving behaviour of those about them, e.g. feeding a doll or combing its hair. See M. Lowe, "Trends in the Development of Representational Play in Infants from One to Three", Journal of Child Psychology and Psychiatry, 16 (1975), 33-47.
10. The Anatomy of Human Destructiveness, (New York, 1973), Chapter 12.
11. The Hidden Human Image (New York, 1974), p.176.
12. Quoted in Wendy Doniger O'Flaherty, Asceticism and Eroticism in the Mythology of Siva (London, 1973), p.188.
13. The gynaecologist Derek Llewellyn-Jones cites this as one of the myths about women that many people still believe. See his A-Z of Women's Health (Oxford, 1983), p.158. I myself have heard this myth earnestly related on more than one occasion.
14. An Introduction to Philosophical Analysis, second edition (London, 1967), pp.405-406.

CHAPTER FIVE

NO SELF TO BE FOUND

Having safely navigated our way between the Scylla of the strict view and the Charybdis of the reductionist views, we must now set out to explore the no-self theory of personal identity. This final leg of our journey, however, will not be without its own hazards; for other problems still lie in our way. One such obstacle is the failure of many philosophers to distinguish between reductionism and the no-self view. The reasons for this error are perhaps understandable. For there is a sense in which the two theories are in agreement. Both theories, for example, reject the notion of a substantial self which somehow exists beyond the bounds of experience. The difference, however, is that while the reductionist accounts then go on to resurrect the self and, consequently, its identity, in terms of putative psychological relations or confused theories of the body, the no-self theory lets the self

lie where it has fallen. This is because the no-self theory is not a theory about the self at all. It is rather a rejection of all such theories as inherently untenable. And since reductionism is just one more theory about the self, it too must be untenable. In explaining the distinction between these two theories it is instructive to turn to the philosophy of mind where we find a similar distinction being employed. Here a distinction is often drawn between reductive materialism and eliminative materialism. Both these theories are in agreement so far as they reject the existence of mental phenomena. But while the former attempts to do so by showing how the notion of the mental can be reduced to the physical - i.e. those things called minds are really just brains - the latter rejects the notion of the mental as fundamentally confused, and so hopes to eliminate the idea of the mental altogether. The eliminative materialist would argue that although the reductive materialist is right to reject the idea of the mental, he is wrong to think he can reconstruct the mental in terms of the physical. This is because, it is argued, discourse about things like intentionality and awareness is simply irreducible to discourse about things like neurological states. Importing this distinction in the discussion of personal identity, we could then say that the no-self theory is an eliminative rather than a reductive theory of personal identity. The no-self critique of the reductive theory of identity would then be that the concept of self and personal identity cannot be reduced to our

ideas about psychology or bodies and, therefore, that such concepts have to be eliminated. We must, however, not push the analogy too far. For although some eliminative materialists see the elimination of the theory of the mental as implying an elimination of mental language, an analogous implication does not hold for the no-self theory. That is, within the no-self theory we can, as I shall argue presently, eliminate the notion of the self and its identity, and yet on pragmatic grounds continue to permit the use of the language of personal identity.

A further but related point that divides reductionism from the no-self theory is that since reductionism seeks to give an account of personal identity - a notion which has its roots deep in the soil of the strict theory - then it has already accepted a certain view into which it must now force the structure of human existence. This, as we have seen, cannot but lead to distorted accounts of how we undergo our consciousness and our bodies. The no-self theory, on the other hand, has no such prior commitments. And in this sense it is more phenomenologically based than are the reductionist theories. To borrow Husserl's phrase, it goes back to the things themselves. That is, it starts with an examination of experience rather than with an attachment to the project of how to account for personal identity. This does not mean of course that the no-self theory need not face the issue of why someone

might come to believe in his own identity. For if there is no such thing as personal identity then it is essential that we can offer some other account of why someone might be led to think there is.

It is just this sort of approach to the problem of personal identity that is found in the writings of the most important no-self theorist, David Hume. As we pointed out at the beginning of this study, Hume was the first Western philosopher to unmask the confusions attending our idea of personal identity and subsequently to reject the idea as a fiction. It will be worth our while therefore to conduct a detailed examination of his position. The problem, however, is there is much disagreement about how we are to interpret Hume on this point. On the one hand there are some commentators who, while agreeing that Hume does reject the idea of personal identity, go on to proclaim that his account is deeply confused and inconsistent with what he says elsewhere. On the other hand there are those who argue that his account is quite consistent, but only because he did not really reject the notion of personal identity but only a particular version of that notion. It shall be my contention in what follows that both these accounts are wrong. The latter is wrong because it fails to accept what Hume clearly says, namely, that identity of the self is a fiction. The former is wrong because it finds confusion and inconsistency where, if we are fair to Hume,

there is none. This is not to say that Hume is completely clear and exhaustive in his treatment of personal identity, but only that within his writings we have the makings of a solution to the problem of personal identity.

What then is Hume's position? If we attend to the section 'Of personal identity' in Book I of A Treatise of Human Nature (1740) an answer is immediately forthcoming. Hume starts by pointing out that although some philosophers believe we are continuously aware of something we call the self, when we look to our experience there is nothing to substantiate this belief. We are never, says Hume, aware of any constant invariable impression that could answer to the name of self. What we experience rather is a continuous flow of perceptions that replace one another in rapid succession. 'When I enter most intimately into what I call myself', says Hume, 'I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception, and never can observe any thing but the perception'.(1) Within the mind, he continues, these perceptions 'successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations' (p. 253). And there is nothing to the mind but these perceptions. There is consequently never any simplicity within the mind at one time nor identity at two different times. Nor, says, Hume, do we

have any idea of a self; for every real idea must be derived from some one impression: 'but self or person is not any one impression, but that to which our several impressions and ideas are supposed to have reference' (p. 251).

With this much said against the notion of personal identity, Hume turns to the question of why we have such a proclivity to ascribe identity to our successive perceptions. Prefacing his reply to this question, Hume points out that a distinction must first be made between personal identity as it regards our thought or imagination, and as it regards our passions or the concern we take in ourselves' (p. 253). His concern here, he says, is with the former. He then starts his answer by distinguishing between the ideas of identity and diversity. In the former case we have the idea of an object that persists invariable and uninterrupted through a particular span of time. It is this that comprises our idea of identity. In the case of diversity we have the idea of several different objects existing in succession, and connected together by a close relation' (p. 253). Now although these two ideas are plainly distinct, it is certain, says Hume, that in our 'common way of thinking' we generally confound them. That is, we often claim that an object at one time is identical with an object at another time, when in fact the two are little more than a succession of different objects connected by a close relation. To justify these absurd ascriptions of identity we either come

up with the notion of a substantial self by feigning the continued existence of our perceptions, or we imagine the existence of something mysterious which binds our many perceptions together. Even where we do not do this we at least have a propensity to do so. We can see then, says Hume, that because we often assert the existence of such fictions, the problem of personal identity is not merely a verbal dispute. It is natural to ask therefore what it is that induces us to mistakenly attribute identity to something while, being a succession of objects, is really an instance of diversity. Hume feels the answer must lie in the workings of the imagination. Consider for instance a mass of matter whose parts are contiguous and connected. This mass will maintain its identity provided it suffers no alteration. If, however, it does suffer such an alteration, and we continue to ascribe identity to it, then it is evident that we are only imagining it to maintain its identity. The reasons why we do this, says Hume, is because 'the passage of thought from the object before the change to the object after it, is so smooth and easy, that we scarce perceive the transition, and are apt to imagine, that 'tis nothing but a continu'd survey of the same object' (p. 256). This is especially true when the alteration is either relatively small or when it occurs gradually and insensibly.

From here Hume goes on to consider the various other ways in which we ascribe identity to objects which are 'variable and

interrupted. Thus, another 'artifice' which induces the imagination to allow the ascription of identity to a changing object, comes from our observing how its parts refer to each other, and, in combination, to some common purpose. Consequently, 'a ship, of which a considerable part has been chang'd by frequent reparations, is still consider'd as the same; nor does the difference of the materials hinder us from ascribing an identity to it' (p. 257). This is because the common purpose of the changing parts allows the imagination to glide easily over the succession of related objects.

Our propensity to ascribe identity in such instances is rendered even stronger when we add a 'sympathy of parts' to the common end of the parts; that is when the changing parts of an object are not only related by the purpose which they share in common, but also by the reciprocal causal relations which they bear to one another. This is true, says Hume, of all plants and animals. The effect of this relation is such that although it is agreed that in the course of a few years all organisms will suffer a total change of their substances, 'yet we still attribute identity to them'. Thus, 'an oak, that grows from a small plant to a large tree, is still the same oak; tho' there be not one particle of matter, or figure of its parts the same. An infant becomes a man, and is sometimes fat, sometimes lean, without any change in his identity' (p.257).

In addition to this it is often the case that we confuse the notions of specific and numerical identity. Hume gives the example of a man who claims to hear the same noise at different times despite the fact that it is continually interrupted. Because the noise is thus interrupted there can be nothing about it, save its cause, which is numerically the same. What the man hears is a noise whose occurrences have only a specific identity. Similarly, says Hume, 'it may be said without breach of propriety of language, that such a church which was formerly of brick, fell to ruin, and that the parish rebuilt the same church of free-stone, and according to modern architecture'. In this situation there is nothing the same about the two churches other than their relation to the parish, 'yet this alone is sufficient to make us denominate them the same' (p. 258).

A further influence on the imagination is our expectations about a particular object. Thus although we normally allow a changing object to maintain its identity only if it changes gradually, we do make exceptions where our expectations are otherwise. As a consequence, since we expect a river to be continually undergoing a change of its parts, we do not allow this to affect its identity.

Hume has now laid the ground for his account of personal identity. For the identity of the mind, like

that of plants, animals, repaired ships, and rebuilt churches, says Hume, is only a fictitious identity. It must therefore be similarly explicable in terms of the workings of the imagination.

It might be objected here that the analogy does not hold because the parts of the mind are connected in a way that is not true of the parts of plants and animals. However, Hume points out that whether contemporary or successive, all our perceptions are distinct existences. That is, none of our perceptions is bound to another by the necessary relation of identity: they are all distinct and different. And, as Hume has argued in an earlier section, there is no necessary connection between distinct existences. From here it follows that 'identity is nothing really belonging to these different perceptions and uniting them together; but is a mere quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect upon them' (p. 260). And the only qualities that are capable of uniting perceptions in the imagination are the relations of resemblance, causation, and contiguity. When we consider which of these relations is responsible for allowing our imagination to unify the successive perceptions of a person's mind, we must, however, dispense with the relation of contiguity which, in this case, has little effect. We are thus left with resemblance and causation, and it is to these relations that Hume now

turns his attention.

Dealing first with resemblance, Hume points out that memory is the natural vehicle for bestowing this relation upon our perceptions. For memory is no other than the ability to make images of past perceptions reappear. 'And as an image necessarily resembles its object, must not the frequent placing of these resembling perceptions in the chain of thought, convey the imagination more easily from one link to another, and make the whole seem like the continuance of one object?' (pp. 260-261). Thus, says Hume, since memory is responsible for creating the relations of resemblance between our perceptions, it not only discovers but also produces our identity. The opposite, however, is true when we consider the role memory plays in causation. For if we consider that the mind is a system of different perceptions all of which are joined together by causal relations, then it is evident that, from this point of view, memory is responsible more for the discovery of identity than for its production. For although memory enables us to acquire the notion of cause and effect, the relation itself is capable of existing beyond the operations of our memory. We can easily see, therefore, how the chain of cause and effect, and thus our personal identity, can be extended to times which we no longer remember. To someone who claims that memory must be the basis of identity, Hume asks whether he can recall the events of certain days long past: 'Or will he affirm, because

he has entirely forgot the incidents of these days, that the present self is not the same person with the self of that time; and by that means overturn all the most established notions of personal identity?' (p. 262).

It is within the discussion of causation that Hume draws his analogy between the self and a republic. For a republic consists of persons who both are united by various legal ties and produce other persons who, in turn, continue the existence of the same republic, 'and as the same individual republic may not only change its members, but also its laws and constitutions; in a like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity. Whatever changes he endures, his several parts are still connected by the relation of causation' (p261).

Hume concludes his account with the important remark that all 'nice and subtile' questions concerning personal identity are best considered as grammatical rather than philosophical difficulties. Thus, except where the notion of a fictional entity or principle is involved, all disputes about personal identity are merely verbal disputes and can never be possibly decided.

A hurried reading of this section of the Treatise might

well lead one to conclude that Hume is confused; for there are some apparent inconsistencies. These difficulties, however, are resolved by keeping in mind the scope of Hume's project and not allowing his claims to be taken out of context. Let us start our appraisal of Hume by dealing with a criticism raised by one of his commentators. In a paper called 'Hume on Personal Identity' Penelhum argues that Hume's account is an immense blunder, 'an excellent example of how complex and far-reaching the consequences of a mistake in linguistic analysis or conceptual investigation can be'.(2) According to Penelhum, Hume's fatal mistake is to think that one object cannot have many parts. This is false, claims Penelhum, because whether or not an object has parts will depend upon what sort of object it is. A melody, for example, is still one melody even though it consists of a succession of notes. A related mistake, we are told, is Hume's muddling of the distinction between the numerical and specific senses of identity. To remain the same through a period of time is to remain the same in a specific sense: to be exactly the same as one was at an earlier time. For an object to change, however, it must remain numerically the same; for it must be one and the same object which is doing the changing.

Unfortunately for Penelhum, Hume makes it quite plain that he does allow for one object to have many parts, and he even allows that the parts may undergo certain changes without

affecting the identity of the whole:

Suppose any mass of matter, of which the parts are contiguous and connected, to be plac'd before us; 'tis plain we must attribute a perfect identity to this mass, provided all the parts continue uninterruptedly and invariably the same, whatever the motion or change of place we may observe either in the whole or in any of the parts. (p.255)

What Hume does not allow is that an object can maintain its identity with a coming and going of its parts. It might seem, however, that this is all Penelhum needs to have his point go through. For what is a melody but one thing which is a succession of other things? And if this is so then Hume would appear to be wrong in holding that a succession of objects cannot be one object.

This, however, will not sustain Penelhum's objection. For if a melody is an object, it is a different sort of object than are the objects that Hume is concerned with. A melody, by definition, is a temporal sequence of musical notes: it is something whose existence is necessarily spread out over time. It is therefore logically incapable of existing instantaneously in the specious present. But there are no such logical restrictions on the existence of trees, ships, churches, or persons. We can easily imagine a tree, for instance, which suddenly pops into existence only to immediately disappear. And we can also imagine, it would seem, a person who exists but

for a second. It is true that trees and persons as a matter of fact tend to exist for varying durations of time, but I see nothing logically incoherent about their existing merely in the specious present. Indeed it is just because our idea of a person does allow for the momentary existence of a person that we are able to ask questions about personal identity in the first place. For only once we have allowed this can we raise the question of whether a person existing at this particular instance is the same as an earlier person existing at an earlier instance. We cannot, however, ask the same question about a melody; for melodies do not exist at particular instances, only their notes do.

So Penelhum's alleged counter-example of one melody that contains several parts is quite acceptable to Hume. It is acceptable, first, because Hume never denied that one object cannot consist of several parts, and secondly, because even though a melody is a succession of parts, this is a necessary feature of a melody in a way that it is not a necessary feature of a person.

What about Hume's purported muddling of the numerical and specific senses of identity? This would be a strange error for Hume to make since, as Penelhum is aware, Hume himself draws this very distinction in the course of his argument. But, according to Penelhum, Hume has made this mistake just so far

as he thinks that for something to remain the same it must not change. For an object to remain unchanged is for it to remain the same in the specific sense. But for an object to change through time it must remain the same in the numerical sense. Of course, says Penelhum, an object can lose its identity by changing, but only if the object is by definition an unchanging thing.

It should not be too difficult to see what is wrong here. Consider Hume's example of a ship that has its parts gradually replaced. It is possible that eventually none of the parts of the original ship remain. And yet the ship of today might be exactly similar to the earlier ship. We can express this relationship by saying the two ships have a specific identity; that is, they exactly resemble each other. It would be false, however, to say that the two ships have a numerical identity; for there is nothing about them that is numerically the same. In contrast to this, Penelhum argues that the only reason for saying that something has lost its identity (i.e. has become another thing) is if it is by definition an unchanging thing. Since nothing in the definition of a ship seems to rule out the possibility of the ship's repairing its parts, then even though the ship of today shares not one plank or bolt in common with the earlier ship, the two are nonetheless numerically identical, i.e. one and the same ship. But were we to accept this scenario we would immediately be

faced with a difficulty. For how could we distinguish between the situation where a ship has persisted without changing any of its parts, and the very different situation where it has changed all its parts? The natural way for us to mark this distinction would be to say that in the former case the ship of today is exactly one and the same as the earlier ship, while in the latter case it only resembles the earlier ship. That is, in the first case the two ships are numerically identical, but in the second case they are only specifically identical. But on Penelhum's account we cannot draw this distinction because there is no distinction to be drawn. Both cases have equal claim to being instances of numerical identity. And yet it is obvious that there is an important distinction to be made here. In the case of the unaltered ship what we have is a ship that is identical in the strongest sense with the earlier ship. In the case where the ship suffers a total change of its parts it can be the same as the earlier ship only in a weaker sense. And this distinction, which is natural to make, is merely the distinction between numerical and specific identity. Hume's account of identity allows us to make this distinction, Penelhum's does not. It seems therefore that it is Penelhum who has muddled things.

We should not, however, be led to conclude that the distinction between numerical and specific identity is central to Hume's discussion. For Hume is not really interested in

specific identity and only introduces the distinction by way of showing how we might come to ascribe (numerical) identity where there is none. One such way - and there are other ways - is to confound the ideas of numerical and specific identity and to claim that two successive objects are numerically identical when they are only specifically identical.

Despite the peripheral significance of the distinction, some people see this distinction as lying at the heart of Hume's thesis. Thus, in an attempt to save Hume from an 'outrageously paradoxical' conclusion (i.e. that the belief in one's identity is absurd) James Noxon (3) argues that Hume's real purpose is to show how identity terms can be meaningfully applied to persons once the numerical/specific distinction is made clear. The reason why there might be a paradox in Hume, we are told, is because Hume appears to be saying two incompatible things at once. To support this claim Noxon quotes Hume's statement that 'the identity, which we ascribe to the mind of man, is only a fictitious one, and of a like kind with that which we ascribe to vegetables and animal bodies' (p.259). Noxon then contrasts this with other of Hume's statements which, we are told, clearly imply that he accepts judgements of identity in such cases. One example cited is Hume's assertion that 'An oak, that grows from a small plant to a large tree, is still the same oak; tho' there be not one particle of matter, or figure of its parts the

same. An infant becomes a man, and is sometimes fat, sometimes lean, without any change in his identity' (p. 257). Another such example is Hume's remarks about someone who supports the memory theory: 'Or will he affirm because he has entirely forgot the incidents of these days, that the present self is not the same person with the self of that time; and by that means, overturn all the most establish'd notions of personal identity?' (p. 262).

This inconsistency is resolved, says Noxon, by noting the different senses of identity that Hume is using. When Hume denies that persons, plants, or animals have identity through time, all he is denying is that they have a numerical identity, he is not denying that they may have a specific identity. This is why he can quite consistently go on to claim that a seedling which becomes an oak is still the same plant, and an infant which becomes a man does not thereby lose his identity; for here the words 'same' and 'identity' are only being used in their specific sense. Noxon is quite convinced that Hume finds no absurdity in identifying a present object, whose nature it is to change, with an earlier object. Such identifications, claims Noxon, are based on 'points of resemblance evaluated in light of general knowledge of the changes which things of a certain sort undergo during a certain period of time'.(4) What Hume is attacking, we are told, is only the view that the identifications which apply to a succession of related objects

depend on the persistence of a single unaltered feature. To hold this view is to confuse the conditions that allow for the ascription of specific identity with those that allow for the ascription of numerical identity. Consequently, Hume's attack on the identity of the self is only to be seen as an attack on its claim to numerical identity; that is, on the self's claim to being a persisting entity that remains unaltered throughout the duration of its existence. And this is quite in accord with Hume's main purpose which is to show that the self does maintain its specific identity through time; i.e. that the present self is exactly similar with the earlier self.

Although this interpretation of Hume may help to avoid the conclusion that Noxon finds so unacceptable, it does not actually fit with what Hume says. It does not take seriously Hume's explicit assertion that the identity of persons, like that of other successions of related objects, is a fictitious one. What Noxon would have us believe is that Hume means only that the numerical identity of such things is fictitious. Hume's tacit claim, according to Noxon, is that what persons, plants, animals, and rebuilt churches really have is a specific identity. But this is an unfounded reading of Hume. For not only does Hume refer to the notion of specific identity only in passing, but nowhere does he mention or even imply that the real identity of these variable and interrupted things is a specific identity. If we turn to the quotations

cited as support for this view, we can see the error immediately. When Hume says that a seedling which becomes a large tree is still the same oak, and that an infant which becomes a man is still the same individual he cannot be referring to specific identity; for it is evidently false to say that a seedling is exactly similar (i.e. specifically identical) to a hundred-foot tree, and just as false to say that an infant is exactly similar to an adult. There may be some vague similarities in either case, but not enough to justify a claim to specific identity. Noxon seems aware of this problem and tries to avoid it by asserting that specific identifications are based on 'points of resemblance evaluated in light of general knowledge of the changes which things of a certain sort undergo'. But I am not even sure if I know what this means. How is the knowledge that a seedling will change into a large oak supposed to aid us in evaluating points of resemblance between the two? For regardless of our knowledge of arboreal ontogeny the fact remains that a sprouting acorn bears little or no resemblance to a massive oak.

Further, it seems clear that Hume, in his examples, is intent on denying similarity rather than affirming it. Not only does the tree have no particles of matter in common with the seedling, but even the 'figure of its parts' is different. Likewise, the infant not only grows into an adult, but is 'sometimes fat, sometimes lean'. Noxon could always

protest here that the tree remains exactly similar to the seedling so far as it is still an oak. But Hume says more than the tree is still an oak, he says it is still the same oak. And this strongly suggests that he is not here using 'same' in its specific sense. If he only meant that the seedling is as much an oak as the tree (a pointless thing to say), there would have been no need for the insertion of the words 'the same' before 'oak'.

But does this not leave us with a problem? For if Hume is not here referring to the specific identity of the oak, then, we might think, he must be referring to its numerical identity; but as we have just noted the tree is said to have not one particle in common with the seedling which means, for Hume, the two cannot be numerically identical. The solution to this problem is to see that Hume is not referring to the tree's identity in any sense: he is referring to our attribution of identity to the tree (that is, our attribution of numerical identity). This is evident from the line which immediately precedes his example of the tree. Hume remarks here 'tho' everyone must allow, that in a very few years both vegetables and animals endure a total change, yet we still attribute identity to them, while their form, size and substance are entirely alter'd (p. 257). The reference to the growing oak and the growing man are merely illustrative examples of a changing plant and a changing animal to which we attribute

identity. Since, however, there can be no identity where the form, size and substance of a thing has entirely altered, the identity which we attribute in such cases can only be a fictitious one; that is, it is the work of the imagination, not a property belonging to the object, or better, to the succession of objects to which we attribute it. And this is not just true for plants and animals but for all things which are variable and interrupted, e.g. repaired ships, rebuilt churches, rivers, republics, and persons. So when Hume discusses 'the identity' of such things, he is only discussing how we come to attribute identity to them, not their actual identity.

The only reason we might think that Hume is making positive statements about the actual identity of things is if we fail to notice the structure of his overall argument and to focus only on specific remarks. If we view the section 'Of personal identity' as a whole we will see that in the first four paragraphs Hume discusses the actual identity of the self and categorically rejects the notion as untenable. He does, however, feel that we have a natural propensity to ascribe identity to ourselves, and it is to an explanation of this propensity that Hume devotes the rest of his discussion. This division in the text is also attended, for the most part, by a division in the type of language that Hume uses to discuss identity. In the first part Hume speaks in a categorical way.

Thus it is claimed, for example, that 'there is no impression constant and invariable' (p. 251), 'there is no such idea [of the self]' (p. 252). 'I never catch myself at any time without a perception, and never can observe anything but the perception' (p. 252), 'I always stumble on some particular perception or other' (p. 252), and finally, 'there is properly no simplicity in [the mind] at one time, nor identity in different' (p. 253). However, when we come to the second part of the discussion, the language becomes more psychological than categorical. That is, Hume's concern here is more with how we imagine, suppose, ascribe, or attribute, identity rather than with the actual identity of things. And so the second part begins by asking 'What then gives us so great a propensity to ascribe an identity to these successive perceptions, and to suppose ourselves possess of an invariable and uninterrupted existence thro' the whole course of our lives?' (p. 253; my emphasis)'. This psychological language continues, for the most part, throughout this latter half: the repaired ship 'is still consider'd as the same; nor does the difference of the materials hinder us from ascribing an identity to it' (p. 257); although plants and animals undergo a total change 'yet we still attribute identity to them' (p. 257); a man who hears an intermittent noise 'says, it is still the same noise' (p. 258); and because an earlier church is demolished before its successor appears we do not think of them as being different 'and for that reason are less scrupulous in

calling them the same' (p. 258; my emphasis throughout the last four quotations). Likewise, when Hume says we can extend our identity beyond our memories his discussion makes it plain that he is only talking about 'the most established notions of personal identity' (p. 262; my emphasis), that is, what is commonly believed to be true about personal identity, and not about personal identity itself.

This use of psychological language to discuss the supposed identity of interrupted and variable objects should keep us alert to the fact that Hume is here discussing only the origin of our belief in such identity, and not the actual identity of what is really an instance of diversity. Once we see this we should not be worried by Hume's occasional use of categorical statements in what is overtly a discussion of psychology. Thus when, in the second part of the text, Hume states that a seedling which becomes a tree 'is still the same oak', all that is before this makes it plain that he can only mean that it is still called the same oak, or it is still supposed to be the same oak, or some such thing, not that it is actually still the same oak. Likewise, when he says that the infant becomes a man 'without any change in his identity', all he can mean is 'without our attributing any change to his identity'.

There will be those, no doubt, who will charge that I am riding roughshod over what Hume says and am twisting the text

to suit my own purposes. But if this is true, then my critics must explain why Hume says that although the addition or removal of even an inconsiderable amount of a mass of matter 'absolutely destroys the identity of the whole, strictly speaking; yet as we seldom think so accurately, we scruple not to pronounce a mass of matter the same, where we find so trivial an alteration' (p. 256). For in saying this Hume has not only given the reasons why growing plants and humans cannot maintain their identity over time (they are continually losing and gaining amounts of matter), but he also explains why we scruple not to pronounce them the same (we find their moment to moment alterations too trivial to note). Of course over many years the alterations that a seedling has undergone will no longer be trivial, ' but where the change is produc'd gradually and insensibly we are less apt to ascribe to it the same effect [i.e. the loss of its identity]' (p. 256).

This is also how we should understand Hume's famous comparison between the mind and a republic. Much commotion has arisen over the fact that Hume here makes an apparently categorical statement. A republic or commonwealth, states Hume, consists of both legal ties and the coming and going of persons who keep alive the same republic. And just as the same republic continues despite a change in its laws and persons, 'in a like manner the same person may vary his character and disposition, as well as his impressions and ideas, without

losing his identity' (p. 261).

Penelhum remarks that this passage 'does not, taken out of context, sound like the account of an alleged mistake at all, but it is quite clear from everything that has led up to it that it is'.(5) His stand on this point, however, is challenged by Lawrence Ashley and Michael Stack in their paper on 'Hume's Theory of the Self and its Identity'.(6) Like Noxon, Ashley and Stack believe that Hume does want to endorse a particular notion of identity. They argue for this reading of Hume not, however, by distinguishing between numerical and specific identity but by alleging that Hume is employing the two concepts of perfect and imperfect identity. Consequently, they say, when Hume states that persons, like republics, can vary their parts without losing their identity, he is in fact making a positive claim about the identity of a variable and interrupted thing. It is just that the identity Hume is here referring to is an imperfect identity, an identity that is only more or less and so is compatible with an object's undergoing alteration. The identity of variable and interrupted objects that Hume dismisses as fictitious is perfect identity, which is only applicable in cases of the invariable and uninterrupted.

But what is the evidence for Hume's drawing a distinction between perfect and imperfect identity? In support of their

claims, Ashley and Stack point to the following passage:

'Since this interruption [in the case of some small change] makes an object cease to appear the same, it must be the uninterrupted progress of the thought which constitutes the imperfect identity(p. 256). They then say 'suppose we take this phrase "imperfect identity" at its face value' by which they mean 'imperfect identity is correctly attributable in cases where the progress of thought is uninterrupted when we reflect on a succession of objects related in certain ways. (This statement does not imply that there is no identity for objects never perceived.)'.(7)

But this is more an instance of sleight-of-hand than it is of taking something at face value. Where does Hume say in the above passage (or anywhere else for that matter) that imperfect identity is correctly attributable when the progress of thought is uninterrupted? What Hume says is that imperfect 'identity is constituted by the uninterrupted progress of thought. And this shows that the identity being referred to, whether it is modified by the word 'imperfect' or not, is a product of the mind, something we imagine, suppose, or ascribe, and not something that is an actual property of successive objects. By substituting the phrase 'correctly attributable' for 'constituted' Ashley and Stack have changed what was meant to be a psychological statement into a categorical one. Their apprehension that Hume's statement is a

psychological one, and their desire that it should not be, is made evident in their parenthetical remark: 'this statement does not imply that there is no identity for objects never perceived'. For although their rewording of Hume may not carry this implication, Hume's own wording clearly does. To say, for example, that it is the trees which constitute the forest is to imply that were there no trees then there would be no forest. And likewise, to say that it is the uninterrupted progress of thought that constitutes the identity of successive objects is to imply that were there no uninterrupted progress of thought then there would be no identity of successive objects.

Further, we cannot help but notice that the term 'imperfect identity' occurs only once in the section on personal identity, and this alone would suggest that Hume is using it stylistically as a form of elegant variation for 'imagined identity' or 'supposed identity', rather than to mark a real distinction between two forms of identity. By the same token, although Hume uses 'perfect' and 'perfectly' to occasionally modify 'identity' and 'identical' respectively, it is apparent that they are not being employed metaphysically, but merely as a form of emphasis. This is supported by the fact that he uses the same modifiers to indicate emphasis in other contexts. He says, for instance, that should both his perceptions and body disappear he would be a 'perfect non-

entity' (p. 252), or 'to explain it perfectly' personal identity must be compared to the identity of plants and animals (p. 253), or again, that he finds an argument 'perfectly decisive' (p. 259).

Coming back to Hume's comparison between the identity of a republic and the identity of a person, we see then that he cannot be referring to something called imperfect identity. And since such things as republics and persons are nothing but successions of related objects, and since successions of related objects are instances of diversity not identity, the identity that Hume is here referring to is only an imagined identity which is constituted by the uninterrupted progress of thought.

It is important to be aware of the place in Hume's account where this comparison appears. Towards the end of the psychological part of the discussion, after concluding that identity does not really belong to our different perceptions 'but is merely a quality, which we attribute to them, because of the union of their ideas [i.e. copies of perceptions] in the imagination, when we reflect upon them' (p. 260), Hume goes on to say that the qualities responsible for such a union must be resemblance and causation. He then discusses both these relations in turn, and it is during his review of the latter relation that he presents his analogy between a republic and a

person. For just as a republic is a series of persons who are united by the relation of legality, so is a person a series of impressions and ideas linked together by the relation of causation. And although Hume here makes the apparently categorical statement that a person may vary his parts without losing his identity, he immediately qualifies this with an important explanatory remark: 'whatever changes he endures, his several parts are still connected by the relation of causation' (p. 262). That is, his parts are still connected by one of the relations responsible for our imagining the existence of identity where there is none.

Does this mean that Penelhum is right in asserting that for Hume it is a mistake to attribute identity to such things as republics and persons? Yes, Hume is quite clear about this. He says in several places that the attribution of identity to instances of diversity is both a mistake and an absurdity (see especially pp. 254-255). But we must be careful here; for its being a mistake does not imply what Penelhum thinks it does.

After noting Hume's claim that we are making a mistake in referring to a person over time as the same person, Penelhum remarks that 'a little effort of imagination is enough to indicate just how much chaos would result from adopting Hume's diagnosis as the source of a prescription and using a

different proper name whenever we noticed the slightest change, even in ourselves (or rather in the separate people that we would be from minute to minute)'. And if this is a mistake, continues Penelhum, it is one whose correction 'would require a complete overhaul of the concepts and syntax of our language'. (8)

Yet there is no reason to think that the correct prescription for Hume's diagnosis need be using a different proper name every time we notice a change in someone or something. Indeed, Hume even says that a rebuilt church, which we only imagine to be the same as an earlier one, can still be called the same as its predecessor 'without breach of the propriety of language' (p. 258). He does not, unfortunately, elaborate on this. However at the end of the section of personal identity something is said that suggests how it might be permissible to call two things the same which in fact are only imagined to be the same. Here Hume states that all disputes about the identity of successive objects 'are merely verbal, except so far as the relation of parts give rise to some fiction or imaginary [i.e. mysterious] principle of union' (p. 262). Thus if two persons are arguing about whether or not an earlier church is the same as its rebuilt predecessor, and neither of them is asserting the existence of a fictional entity or principle of union which somehow unites the two churches, then their dispute will be merely about how

the word 'same' is to be used in these circumstances. That is, they will not be disputing about an actual identity but only about the linguistic conventions surrounding our use of identity terms. If we imagine that our disputants finally agree that our linguistic conventions permit us to call the two churches the same (a conclusion to which Hume would give his assent), then we can see why it is permissible to call two things the same which are in fact different. This is because there are two levels at which the notion of identity can be employed: one which deals with questions about identity at the metaphysical or ultimate level, and one which deals with them at the verbal or conventional level.

It is disappointing that Hume does not have more to say about the two levels of 'disputes'. For it is with such an account that we are able to explain why it is acceptable for us to continue to talk in terms of selves and personal identity despite the fact that there are no such things. We can, however, arrive at a fuller understanding of what the two-level account involves by turning to another version of the no-self theory.

At certain points in our inquiry we have had the chance to refer to various Buddhist pronouncements on the problem of personal identity. And here too it seems Buddhist theory can offer some insights. For at the very heart of this theory lies

the doctrine of the two levels of truth. Although the different schools of Buddhist thought disagree on the exact nature of the distinction to be drawn between the two truths, there are enough similarities - at least in the early Hinayana Schools - for us to give a general account. This will be useful to our project because it will allow us both to see what Hume might have been getting at, and also to acquire more munitions with which to fend off this attack on the no-self theory. I hasten to add, however, that I am not here attempting a scholarly exposition of Buddhist thought; for my present interests are confined to an exposition of the no-self theory of personal identity. It is just that Buddhism has some valuable contributions to make here. There are of course problems involved in the cross-cultural discussion of ideas; Hume and the Buddha, after all, lived their lives in very different social and historical contexts. And yet I do not think these difficulties need detain us; for when we go to the texts where Buddhist thinkers are grappling with the problem of personal identity we find their concerns are essentially the same as Hume's.

In the earliest texts of Buddhism, the Pali Canon (about 500 B.C.), we come across a distinction drawn between two types of discourse: those of direct meaning and those of indirect meaning. The former type of discourse is said to be one whose meaning is plain while the latter type needs to have its

meaning inferred with reference to the former. In the discourses of indirect meaning words are used which apparently refer to persisting entities such as a self or an I which, according to the Buddha, are merely 'expressions, turns of speech, designations which the Tathagata [i.e. the Buddha] makes use of without being led astray by them'.(9) That is, although we may use words like 'self' and 'I', we should not be led into thinking that they actually refer to something, for they are but grammatical devices. This non-denoting aspect of these expressions is something which must be inferred in light of the discourses of direct meaning. In this latter type of discourse the non-existence of anything permanent or enduring, such as the self or I, is asserted and the misleading features of language - those features which lead us astray into the belief in an I - are made explicit. Here there is no need for inference since the meaning of such discourse is plain.

As it happens, however, we are apt to confuse the two types of discourse: 'there are these two who misrepresent the Tathagata. Which two? He who represents a discourse of indirect meaning as a discourse of direct meaning and he who represents a discourse of direct meaning as a discourse of indirect meaning'.(10) Although the Pali Canon does not elaborate here we can easily see what sorts of errors are being referred to. On the one hand we might think that someone who is using the words 'self', 'I', or 'Buddha' (which are mere

turns of speech) is in fact denoting a particular entity. Or, on the other hand, we might think that someone who is denying the existence of the self cannot really mean what he is saying and so we might be tempted to infer a further meaning which would still allow the existence of the self. We might, for instance, think that the person making this claim is only denying the existence of a certain type of self.

The discussion of the two types of discourse is continued in the various Buddhist commentaries on the Pali Canon, and here we are introduced to the related ideas of two levels of truth. In one commentary it is stated that all 'Buddhas [i.e. enlightened beings] have two types of speech; conventional and ultimate. Thus 'being', 'man', 'person', [the proper names] 'Tissa', 'Naga' are used as conventional speech. 'Categories', 'elements', 'sense-bases' are used as ultimate speech'. Because of this division in speech we are told that the Buddha 'declared two truths; the conventional and ultimate, there is no third. Words (used by) mutual agreement are true because of worldly convention; words of ultimate meaning are true because of the existence of elements'.(11) Although the various elements are said to be the constituents of which everything else, including what we call the self, is made, it is not because the elements are more basic than the self that the self is said not to ultimately exist. It is simply because there is nothing in the world, not even an assemblage of the elements,

that can be identified with the self. Although the Buddha cites various characteristics that something must have if it is to be considered a self, the most important is that of permanence or identity over time. But when we look to our experience there is nothing but impermanence: our bodies, feelings, and thoughts are forever coming and going. In this sense the Buddha is in complete agreement with Hume: where there is diversity there can be no identity. None of this, however, implies that statements which make use of words like 'self', 'I', 'you', 'Tisa', or 'Buddha', are false or nonsensical at every level of discourse. For they can be true at the conventional level; which means they can be true because of their being used in accordance with mutual agreement, i.e. linguistic convention.

A good illustration of how this distinction is to be drawn is given in a well-known passage from the The Questions of King Menander (about 100 A.D.).¹² In this dialogue the Indo-Greek king Menander puts various questions about the nature of the self to the Buddhist monk Nagasena. At the opening of the dialogue Menander asks 'How is your reverence known, and what is your name?' The somewhat provocative answer given to the King is 'I'm known as Nagasena, your Majesty, that's what my fellow monks call me. But though my parents may have given me such a name . . . it's only a generally understood term, a practical designation. There is no question of a

permanent individual implied in the use of the word'. Menander is quite astonished by this reply and eventually asks 'if your fellow monks call you Nagasena, what then is Nagasena?' He asks whether Nagasena is any part of the body or the mind, or whether he is all of these things taken together or whether he is anything apart from them. To all of this Nagasena replies merely 'No, your Majesty'. Menander then exclaims triumphantly 'then for all my asking I find no Nagasena. Nagasena is a mere sound! Surely what your Reverence has said is false!' But Nagasena is not to be so swiftly dealt with and, in good Socratic fashion, replies by himself asking a question: 'Your Majesty, how did you come here - on foot, or in a vehicle?' 'In a chariot', says Menander. Nagasena then asks what the chariot is, whether it is the pole, axle, wheels, frame, reins, or yoke, or whether it is all these taken together or again whether it is something other than the separate parts. Menander replies here in the negative. With this Nagasena fires back at the King his own reasoning: 'Then for all my asking, your Majesty, I can find no chariot. The chariot is merely a sound. What then is the chariot? Surely what your Majesty has said is false! There is no chariot!' Menander protests that he has not said anything false; 'it's on account of all these various components the pole, axle, wheels, and so on, that the vehicle is called a chariot. It's just a generally understood term, a practical designation'. Nagasena's rejoinder is to praise Menander for this remark and to point out the same holds

true of himself. For it is because of his various components that he is known by the practical designation 'Nagasena'. However, he adds, in the ultimate sense there is no person to whom the name refers.

*But why
is this the
right answer?*

It is easy to mistake this passage, as some people have done, for a statement of reductionism. But the text clearly disallows this interpretation. The reductionist view would be that the person of Nagasena can be reduced without remainder to his various impersonal constituents. But when the King asks if Nagasena is all of the parts of his body and mind taken together, Nagasena answers 'No'. This is because Nagasena rejects any notion of a person that exists in the ultimate sense: a person is not ultimately something other than his parts (the strict theory), nor is a person ultimately the sum of his parts (the reductionist theory). This does not mean however that the word 'Nagasena' is a mere sound; for it is more than that: it is a generally understood term whose proper use is determined by mutual agreements concerning how, when, and where it is to be used. Or, as Nagasena says, it is because of his various components that he is known as 'Nagasena', even though 'Nagasena' does not refer to anything.

It is crucial to see, however, that in relegating such terms to the realms of conventional discourse, Buddhism is not proposing a conventionalist theory of personal identity. That

is, for Buddhism, a person's identity is not something to be ultimately decided by convention. For to hold this position would be to commit the very error against which the Pali Canon warns us, i.e. to represent a discourse of indirect meaning as a discourse of direct meaning. It would be to conflate the levels of truth and think that questions at the ultimate level can be answered by agreements at the conventional level.

We are now able to see why the no-self theory does not imply that our language is in need of an overhaul. For it is quite consistent with the non-existence of the self or I that we continue to employ the words 'self' and 'I' in their practical everyday usage, provided we do not mistake them for denoting some particular entity at the ultimate level, or, as Hume would say, feign the existence of a fiction. This is why, contrary to what many of Hume's critics think, Hume's own use of the first-person pronoun does not undermine his theory. In Hume's statement 'when I enter most intimately into what I call myself, I always stumble on some particular perception . . . and never can observe anything but the perception' (p. 252), the word 'I' is being used at the conventional level: it is merely a generally understood term whose proper use is determined by mutual agreements. We should not therefore think that in using the first-person pronoun Hume has committed himself to the existence of a self at the ultimate level.

Some will no doubt find it paradoxical that we can use personal language correctly when there is nothing to which these terms ultimately refer. It was reasoning akin to this, it seems, that led Descartes to his famous proclamation 'I think, therefore I am'. I must exist, reasoned Descartes, because even when I doubt that I exist there is still an I that is doing the doubting. But Descartes has become led astray by his own language; for there is no need for the 'I' in 'I think' or 'I doubt' to refer to anything. What Descartes was aware of, as both Hume and the Buddha would agree, was just thinking, not an I that was doing the thinking. Consequently Descartes might have just as well said (and should have said if his concern was with ultimate rather than conventional truth) 'there is thinking, therefore there are thoughts'. And such a deduction, if we may call it that, does not suffice to prove the existence of an I.

A possible response here would be to say that although there need be no reference to an I when we use the nounal sense of 'thinking' or 'thoughts', when the verbal sense 'I think' is employed then plainly there must be some reference to a subject; for what is it that thinks. To this it can be replied that although the term 'think' does require a subject this is little more than a grammatical requirement. And so we might just as well employ a non-referring grammatical subject rather than the misleading term 'I'. This is a point that is

recognized by George Christoph Lichtenberg who says about Descartes' dictum, 'We should say, "It thinks", just as we say "It thunders". Even to say cogito is too much if we translate it with "I think". To assume the "I" to postulate it, is a practical need'.(13) Thus since the use of the verbal sense 'thunders' also requires the introduction of a subject, we bring in the word 'it' and say 'it thunders'. But this does not mean that the grammatical subject 'it' here refers to anything' All we are saying when we say 'it thunders' is 'there is thunder'. Consequently, since the requirement that the word 'thinks' has a subject is also a convention of grammar, or as we might say with Lichtenberg, a practical need, we could likewise employ 'it' to serve this purpose. We could say 'it thinks, therefore there are thoughts' and the appearance of 'it' here would no more imply a reference to an actual subject than would 'it' in 'it thunders'.(14)

There remains, however, a further problem which needs our attention. It was mentioned earlier that one of the criticisms levelled at Hume is that his account of personal identity is inconsistent with what he says elsewhere. The passages which are supposed to contradict the view of the section on personal identity come from Book II of the Treatise wherein Hume discusses the nature of the passions or emotions.(15) Because Hume refers here to the importance of the role played by the self as the object of such passions as pride and

humility, some writers have been quick to accuse him of contradicting his earlier claims about the self. Thus Norman Kemp Smith cites the following passages from Book II as being incompatible with Hume's claims from Book I:

'Tis evident, that the idea, or rather impression of ourselves is always intimately present with us, and that our consciousness gives us so lively a conception of our own person, that 'tis not possible to imagine, that any thing can in this particular go beyond it. (p.317)

The stronger the relation is betwixt ourselves and any object, the more easily does the imagination make the transition, and convey to the related idea, the vivacity of conception, with which we always form the idea of our own person.(p.318) (16)

Kemp Smith suggests that the reason why Hume so freely makes use of the idea of the self in Book II when he has just dismissed the notion in Book I is because Book II was written before Book I, and, consequently, by the time Hume had got around to writing Book I he had forgotten what he had said in Book II.(17) Before we accept this picture of a bizzarely absent-minded Hume it seems we should look at what is actually being said about the self in Book II. And what we find in the opening pages of Book II is this: ''Tis evident, that pride and humility, tho' directly contrary, have yet the same object. This object is the self, or that succession of related ideas and impressions, of which we have an intimate memory and consciousness'. And a few lines later we are told about 'that

connected succession of perceptions, which we call self' (p. 277). These hardly seem like the words of someone who is totally incognizant of the view expressed in the section on personal identity. Indeed by taking the time to insert these descriptions of the self in Book II it seems that Hume is anxious to remind us of the conclusions reached in Book I. And so, when Hume says in Book II that an awareness of ourselves is always intimately present to us we should understand this in terms of what was said in Book I; namely, that to enter intimately into what is called the self is just to encounter various impressions. That is, what is always intimately present to us is just those particular perceptions whose succession we call the self (the fact that Hume uses the word 'intimately' in both the passages from Book I and Book II also suggests that he is intent on discussing the same notion of the self in both places).

Still there is an apparent problem here. And this is highlighted by the fact that Hume often speaks in Book II of our idea or impression of the self as though it were something which occurs instantaneously within our awareness, a singular perception which we experience as the object of pride and humility. But as Hume has told us there is no impression or idea of the self; there is only the smooth and uninterrupted progress of thought within our imagination. But how is it that a smooth and uninterrupted progress of thought could appear

instantaneously within awareness? The notion of an uninterrupted progress is a temporal notion; it is something that occurs over an extended period of time. Accordingly, it is difficult to see how such a train of perceptions could be an object to which we could, in an instance, direct our emotions.

I doubt, however, that this is a major difficulty. Hume himself is aware that a different account is needed to deal with those occasions, such as in certain emotional states, where we suddenly come across the object of what we call self-awareness. He says, we have noted, that in order to discover what is responsible for the tendency to believe in personal identity a distinction must be drawn between 'personal identity, as it regards our thought or imagination, and as it regards our passions or the concern we take in ourselves' (p. 253). Personal identity of the first sort is his professed concern in Book I while in Book II it is the latter sort. The problem is that Hume never explains just how we are to understand the relation between the two sorts of personal identity. He obviously wants the personal identity that concerns the passions to be basically the same as that which the imagination constructs. This is evident from the descriptions of the self given in his discussions of the passions. And yet his theory seems to require that the object of the passions be capable of appearing to awareness in a way that does not render it a succession.

The way to deal with this problem is to see that when we enter a psychological or emotional state that seems to depend on an instantaneous awareness of self, what we are doing is not reflecting on a succession of related ideas (which could not be done at any one instance), rather we are latching on to a particular collection of some of these ideas which, by virtue of their being related, can instantaneously present themselves in a condensed form to our awareness. It is with this discovery of a constructed or condensed self-image, as we shall call it, that we are now brought back to a point set aside earlier.

In the second chapter we saw that although McTaggart claimed we know the self by direct acquaintance, he was nevertheless alive to the possibility that the establishment of self-perception would not be enough to prove that a self existed, but only that something was perceived as being a self. McTaggart's suspicion that we might perceive something which we mistakenly believe to be ourself is borne out by an examination of our states of self-awareness. We can start by noting what was said in Chapter One, namely, that although we do experience occasions of self-awareness there are numerous instances in which it is plain that we have no awareness of anything that can be considered a self. The examples cited, such as reading and day-dreaming, are cases where I become immersed in the

activity to the point that there is no room in consciousness for the awareness of an I. Here there is just the experience of the activity: the gradual unfolding of the novel or the spinning of the day-dream. It is of course possible for me to shift my awareness to a point where I am aware that I am reading or day-dreaming, but then I am no longer immersed in the activity. It is important to see, however, that when I am thus not immersed then my experience of the activity is drastically altered. This is because in entering a state of self-awareness I undergo a 'giving-up' or a 'forsaking' of the activity in which I was previously engaged. My reading gets left behind as I conjure up the I to which I now direct my attention. Here the act of reading loses its sense of spontaneity; it becomes something I must struggle with, something impeded by the new awareness that it is I who am doing the reading. The act of self-awareness is thus a reflective and complicated act which involves both a stepping back from the flow of experience and the introduction of a further element (or rather collation of elements) into that experience. It is in this sense then that self-awareness can be called a secondary phenomenon; for the object of self-awareness is not part of the basic fabric of experience, rather it is something which experience itself fabricates and then takes as its object. It is consequently understandable that self-awareness, with its complicated and reflective qualities, is but an infrequent visitor to consciousness.

This view of the secondariness of self-awareness is supported, I think, by what we know about the ontogeny of consciousness. The developmental psychologist Jerome Kagan, for example, delineates a sequence of four psychological functions which lead up to but do not imply the emergence of self-awareness. According to Kagan, the first function to appear which comes close to implying the existence of consciousness is recognition memory. This makes its appearance in the first eight months of life and is displayed in the ability to discriminate between familiar and novel stimuli. A few months later there appears the functions of retrieval and inference, by which Kagan means 'the retrieval of schemata without any cues in the immediate field and the generation of inferences following the relating of those schemata to present experience'.(18) These functions are demonstrated by a child who looks persistently for a toy he just saw hidden a moment earlier. Here the child maintains a schemata of the hiding of the toy and infers that the toy must still exist somewhere. The fourth function, which emerges at about sixteen months, constitutes the child's ability to monitor, select, and control the other functions. This function is the awareness of one's potentiality for action and, for Kagan, comes closest to our notion of consciousness.

The difficulty, however, appears with the emergence of the

fifth function, namely, the awareness of self as an entity with characteristics. This is a difficulty because although the function of self-awareness seems to depend on the appearance of the earlier functions, nothing in the earlier functions implies that self-awareness need emerge: 'To explain why a 30-month-old declares, laughingly, "I'm a baby", as she crawls on the floor and sucks on a bottle, requires more than an awareness of one's ability to act. This behaviour presupposes not only a schema for one's actions but also a schema for the self as an object with variations in attribute' (19). Consequently, says Kagan, some new processes must be introduced into the developmental sequence if the self-awareness function is to make its entrance. But just what this must involve Kagan cannot say; it remains, we are told, part of an ancient enigma.

However, it is not too difficult to see that an answer to this 'enigma' is already contained within the account. We need only observe that Kagan's schema for the self as an object with variations is just what we have called the constructed self-image. That is, it is a collation of earlier experiences which presents itself in a condensed form to our awareness. Once we spell out the schema for the self in this way we can see that the process needed to generate it is not altogether dissimilar from the process which generates the fourth function, i.e. the awareness of the potentiality for action. The child arrives at his action-schema by directing his awareness on his own

abilities, that is, by monitoring, selecting, and controlling the more primitive functions of recognition memory, retrieval, and inference. Once the child has acquired the capacity thus to direct his awareness, it is but a small step for him to then turn his awareness on some of the perceptions which constitute the succession of related ideas and so arrive at a schema for the self. However, although the self-schema will contain components not to be found in other schemata, the components of other schemata will nevertheless be discoverable within the self-schema. This is what Kagan means, I take it, by saying that the function of self-awareness depends on the earlier functions. Thus the infant's awareness of being a baby will require more than an awareness of the ability to act, but will still involve such an awareness. This is because to have the awareness 'I'm a baby' is, among other things, to be aware that one acts like a baby; and even a 30-month old will know that crawling on the floor and sucking on a bottle are things that are (normally) only done by babies. The reason, however, why self-awareness will include more than the child's awareness of his ability to act is because in addition to acting, the child also is acted upon by others, has an idea of his own appearance, experiences emotions, and so on; all of which are condensed into the object of self-awareness. Thus self-awareness is a secondary phenomenon on the ontogenic theory because it does not involve the acquisition of a new ability: it is merely the deployment of an earlier function upon a wider

range of objects.

Having established that self-awareness is both a complicated and rare occurrence, we can now move on to the phenomenology of the constructed self-image. For in laying bare its structure we will see that although it is the object of what we call self-awareness, it is constituted by nothing more than a collection of transient images.

One of the central features of the constructed self-image is that it is a condensation of related experiences. It is this feature that helps it to masquerade as being a self; for being a condensation of our experience it appears both as a singular thing (though as we shall see this need not be so) and, at the same time, as something which contains our experiences. Since the experiences of which it is composed may be taken from the different times of our life, we can also see how the constructed self-image might, in an instance, convey something of the notion of identity over time; that is, present itself as something that has persisted throughout our lives.

The psychological process of condensing several experiences into a single image or idea is not unique to the structure of self-awareness. On the contrary, it is a process commonplace in much of our psychology. Thus when I call up an image of my friend Mary I am not presented with a single

portrait, as it were, of Mary. What I find is several different images of Mary which have been collated and superimposed upon one another. I may, for example, have an image of her hair in the wind, which at the same time incorporates the movement of her dress. This composition may be further blended with an image of her face which is at once a frontal view and a profile which in turn may allow the blueness of her eyes to merge with the turquoise of her ear-rings. In the same instance this will all be bathed in emotional qualities which will reveal the feelings I have for Mary. Such images are of course transient with certain of their parts being introduced, expelled, and replaced in an instance. However, as long as I dwell on the image of Mary there will continue to be an indistinct network of constituent images, the older of which are forever dying away while new relations take their place. A similar process of condensation also seems to take place in the construction of dream images; and here, because of what Freud has called primary process or the unrestrained character of hypnagogic cognition, the resultant complex may incorporate numerous diverse elements that are held together by only the remotest of connections.(20) We may, for example, encounter someone in our dreams who displays the characteristics of several different people at once: in one way he is our childhood playmate, in another way he is our old school-teacher, while in yet a third way he is someone who we saw only yesterday. It is as if we were looking at one of those anatomy

text-books which depicts the different organ-systems on a series of overlapping transparent pages. On the bottom page is a skeletal-system, on the next page is the nervous-system, then comes circulatory-system, and so on. The end result is a single image of the human body in which the different organ-systems are nevertheless discernable.

The constructed self-image is put together in much the same way as other condensed images. When I enter a state of self-awareness, the I that is summoned before my consciousness is not a simple entity that infixes itself changelessly in my mind. It is rather a composite of various fading images which will have some reference to how I see and feel about myself. I may, for example, have an image of my face as it appeared to me in the mirror this morning which is nevertheless infused with features of previous images of my face. Thus although my eyes and lips might appear to me as they did today, my cheeks and the shape of my face might seem more like those of myself of twelve years ago. Or again it might include features of how I would imagine myself to look in twenty years. This composite image of my face might itself be superimposed on some familiar scene, say the beach where I often go for walks. Here the sand-dunes might be incorporated into the cheeks and the rising of the waves into the forming of a smile. And all of this will be presented in a suffusion of affective tones which will exhibit the emotional evaluations I have of myself. Like

other of our condensed images, the self-image will have but an ephemeral existence; the constituent images continuously dissolving as new associations make their way into the complex. Just what are the constituents of the constructed self-image will naturally be different for each person, since each person will see himself in a different way. While for one person it may consist mainly of idealized images of his physical appearance, for another it may be a mixture of certain sensations or emotions, while for a third it may be images of how others respond to him. To verify that this is so one need only ask different persons to describe what it is they are aware of when they are aware of themselves. It will then be seen that such images vary quite markedly. Further, it also seems to be that although some people have a relatively consistent image of themselves over time (they have always seen themselves as a whirling complex of emotions), others do not (they now see themselves as more of a physical entity than they did five years ago). This diversity of self-images would help to explain both why there is no unified common-sense notion of the self and why there is no commonplace conviction about one's identity over time.

I just said that the constructed self-image need not exist as a singular thing; that is, at any one moment it is possible for a person to have more than one constructed self-image. And this much seems to follow both from what has been said about

the possible inconsistencies in the self-image over time, and from the nature of the psychological mechanism of condensation. For in the first place, if one can construct two or more different images of oneself at different times, then the stage is set for the possibility of constructing as many different self-images at the same time. It might be objected that consciousness could not operate in this divided way, but such a complaint must depend on the doctrine of the unity of consciousness, a doctrine whose truth we have had grounds to question throughout this study.

Further, because the constructed self-image is a condensation of related ideas, it seems likely that certain of our ideas which normally appear when we think of ourselves are such that they bear little relation to other ideas that also appear in the same instance. Consequently, any attempt at condensing these ideas into a composite idea will tend to fail. Now in most cases the solution to this dilemma will be to somehow repress one of the inconsistent experiences or at least deny it entry into the constructed self-image; that is, not allow it to be an element in the state of self-awareness. The problem, however, is that some of these offending experiences may present themselves as too salient or meaningful to accept banishment from the self-image. What we have then is an experience or group of experiences that is at once pulled into and yet expelled from the self-image. Under these

conditions it is quite conceivable that the exiled experiences might set up their own camp in which they too could fly the flag of the self-image. There would then exist two or more groupings of experience, each of which would present itself as the appropriate object of self-awareness. To enter a state of self-awareness in these circumstances would be to have one's awareness divide itself between two objects competing for the same phenomenological status. In understanding how this could occur in consciousness we must not think of the constructed self-image as an object to which I direct my attention in the same way that I might direct a beam of light upon a plane in the sky. For if we use this analogy we will be tempted to see the constructed self-image and my consciousness of it as if they were two separate entities, the latter of which somehow engulfs the former. We will then imagine that to have two or more simultaneous self-images will involve little more than having one instance of consciousness which simultaneously engulfs more than one thing; much like one beam of light might illuminate two planes. But in reality the constructed self-image just is an instance of consciousness. For when I enter into a state of self-awareness what happens is this: my consciousness throws itself into a certain configuration that is structured in the form of a condensed self-image which permeates or diffuses over my world at that instance. In this situation we can begin to see how parts of my experience, which are neither assimilable nor repressible, might fissure

themselves off to become a separate sphere of consciousness.

A person in this state might then begin to experience himself as two distinct persons. Something like this could well play a role in the genesis of the type of dissociative condition known as multiple personality. Here the problem of having more than one self-image would be dealt with by producing a schism in awareness which would subsequently be attended by the appearance of two or more personalities, each of which would be matched to the appropriate self-image. I am not saying of course that the constructed self-image is to be identified with the personality, but only that whatever sort of image we have of ourselves will tend to relate to both the way we behave and to the way we think, and, as a consequence, that having two or more radically distinct self-images will tend to correlate with having two or more radically distinct personalities.

We now have to ask what it is that leads someone into perceiving the constructed self-image as being a self. And here again the work of Buddhist philosophers is most enlightening. According to Buddhist theory what we call a person is really just an aggregation of the five khandhas or elements. These are physical form, perceptions, feelings, motives, and consciousness. But none of these elements whether considered separately or in combination can rightly be

identified with the self, for they lack the various qualities which we attribute to the self.(21) This, however, does not stop someone from mistakenly identifying himself with one or other of the elements, and indeed this is a ubiquitous confusion from which Buddhism hopes to set us free. But what is it that leads a person to this mistaken identification? To answer this we need to refer back to our previous discussion of the conventional and ultimate levels of truth. There we saw that although personal names and personal pronouns do not at the ultimate level refer to anything, at the conventional level it is quite acceptable to use such expressions for pragmatic reasons. Thus the Buddha uses the language of the self as convenient designations without being led astray by them. The problem is that, unlike the Buddha, many of us do get led astray by the expressions we use; that is, in failing to notice we are using language at the level of convention, we end up thinking there must be something to which the words 'I' or 'self' refer. And so we turn our gaze inward (because this is where the self is supposed to exist) and coming upon one or other of the elements, or a collection of the elements, hasten to identify it with our self. Buddhism underlines the importance that language plays here by making a didactic use of the Pali word ahamkara which can mean both 'the utterance of "I"' and 'I-maker'.(21) That the same word has both meanings helps to suggest a connection between the two meanings; not only that the language of the self leads to the fabrication of

a self but also that a fabricated self leads to a misconstrual of the language of the self. Here then we see the cyclical nature of the trap in which the straying language-user is caught. In uttering 'I' he is led to misidentify an element in his experiences as the self. Having affected this delusory identification he then goes on to make similar utterances firm in the belief that these utterances ultimately refer to his self. For Buddhism, it is thus we come to weave the first threads of an ever-expanding veil of delusion. For in believing that we have a self we are easily led to other experiences which depend on this illusory self. Pride and humility, for example, can only get their foothold if there is something which we perceive to be the self and to which we can relate the causes of our pride or humility. I cannot be proud of my successes or humiliated by my failures unless I believe there is an I to whom these successes and failures belong. This is why on Buddhist theory, the giving up of the belief in the self - which we must do if we are to be free from delusion - is also attended by the cessation of pride, humility, embarrassment, envy, and other self-orientated (and hence delusive) emotions.

The Buddhist method for over-coming the delusion of self is to engage in sustained meditation; a practise which may be described as pure self-analysis or intensive inward-gazing. Through this technique the supposed internal self comes to be

seen for what it is: a mere collection of transient elements. This realization loosens the grip of the belief in self and so dissolves the constructed self-image back into the elements from which it came. Because we can find ourselves so firmly held by the delusion of self, the point at which we are able to break free from its grasp is often described in Buddhist literature as a mystical sense of liberation; a state in which the boundaries between self and not-self crumble before our eyes. A striking account of this experience is given by the Japanese Zen Master Sakei-on Sasaki:

One day I wiped out all the notions from my mind. I gave up all desire. I discarded all the words with which I thought and stayed in quietude. I felt a little queer - as if I were being carried into something, or as if I were touching some power unknown to me . . . and Ztt! I entered. I lost the boundary of my physical body. I had my skin, of course, but I felt like I was standing in the centre of the cosmos. I spoke, but my words lost their meaning. I saw people coming towards me, but all were the same man. All were myself! I had never known this world. I had believed that I was created, but now I must change my opinion: I was never created; I was the cosmos; no individual Mr Sasaki existed.(23)

We are at last in a position to see how Hume's idea of personal identity as it regards the imagination is related to the notion of the constructed self-image, or personal identity as it regards the passions. When I come to believe that I am the same person I was twelve years ago it is because of the smooth and uninterrupted progress of thought that is produced in my imagination when I reflect on the succession of related

ideas. This is the imaginary self which is contemplated over time and whose identity is fictitious. When, on the other hand, I believe that I am perceiving my self in an instance of self-awareness, such as when I experience a self-orientated emotion, what is happening is my awareness is being directed upon an object which is merely a condensed version of extracts from the succession of related ideas. This is the constructed self-image which is a rare and secondary aspect of consciousness. In either case what I come upon is merely a collection of experiences: in either case there is no self to be found.

This brings us to the end of our enquiry. Before we conclude, however, let us pause to survey the distance over which we have come. At the opening of this study I suggested that the problem of personal identity, which is a problem of identity over time, is essentially an empirical problem and thus is to be approached through an examination of experience. Holding to this principle we have witnessed a common theme re-emerging throughout these pages. This is the theme of the refusal of experience to allow itself to be swept up in the net of personal identity. Being aware of this, the strict theorists were forced to postulate a self whose identity somehow existed beyond the realms of experience. This was what Butler was trying to do when he claimed that consciousness of personal identity presupposed and therefore could not constitute personal identity. This move was destined to fail

however because it depended on the false assumption that the only way to distinguish between a delusive and veridical memory is by reference to the one and the same person who had the memory. This assumption is wrong, first, because the way to distinguish between the two sorts of memories is by an appeal to their causal histories, not to the person who had them, and, secondly, because there is nothing in my memories to testify that they were experiences had by me i.e. by one and the same indefinable self who is now experiencing them. For on the strict theory I am, as Reid puts it, neither thought, feeling, nor action, but rather something which thinks, feels, and acts. And yet when I examine my memories all I can find is thoughts, feelings, and actions, never this other something which is supposed to be their author. But if this self fails to make an appearance in my memories, then we are left without any explanation of how it is my memories testify that they were had by me.

Unable to support its claims by an appeal to experience it was not surprising that the strict theory moved on to a more a priori form of argument. Thus we found McTaggart claiming that since no description of any fact about myself implied that the description referred to me, then our knowledge of the self could not be knowledge by description but must be knowledge by direct acquaintance. The problem however is it does not follow from the fact that we do not know the self by description that

we must know it by acquaintance; for there is another possibility, namely, that we do not know the self at all. And since it is this last possibility that receives the support of experience, then it is plain that McTaggart's argument is of no avail. This is also why Madell's argument from the imagination had to run aground. Madell claimed that because I could imagine myself possessing a totally different set of properties than the ones I now have, I must be something other than my properties. But this, like all the other arguments for the strict theory, ignores the fact that I have no awareness of a self in the first place. And if I have no awareness of a self which now possesses certain properties, how could I begin to imagine it possessing other properties?

Finding no satisfaction with the strict theory we turned to Locke's memory theory and its offspring the psychological theory. By arguing that the self is constructed out of experience with its identity constituted by various psychological relations these accounts made some definite advances on the strict theory. Williams, however, uncovered a serious flaw in the psychological theory when he showed that it was possible for two present individuals to be psychologically continuous with one earlier person. For since two present persons cannot be identical with one past person - a relation which, if it held, would make them identical with each other - then it seems that psychological continuity cannot here be the

basis of personal identity. And if it cannot be the basis of identity in the situation where there are two present persons who display psychological continuity with a past person, then there is no reason to think it can be the basis of identity where only one person displays it. Williams' conclusion was that, in the absence of bodily continuity, the most we could say about someone who claimed to remember being an earlier person is that he did not really remember being the earlier person, but only mysteriously or clairvoyantly knew all about him. And here the argument might have rested had it not be for Wiggins' observation about the possibility of bisecting one person's brain and transplanting it into two separate bodies. For in this case the two resulting persons' claims to remember being the same earlier person would evidently not be an instance of clairvoyance. This possibility, however, raises its own problem. For now we must ask what has happened to the person whose consciousness has become thus bifurcated.

It was Parfit who attempt to address this issue by claiming that the various possible outcomes of Wiggins' operation were merely different descriptions of the same outcome. Thus the question of importance was not which description we ought to choose, but rather what ought to matter to me in a case of my fission. Parfit's answer was that what ought to matter is relation R: psychological connectedness and/or psychological continuity with the right kind of cause.

several
ambitions

Since this relation persists through fission then I ought to regard my fission as being near as good as survival (despite the fact that it is best described as dying). This account of the matter, unfortunately, turned out to be deeply confused. For not only was it wrong to call the possible outcomes merely different descriptions of the same event, but the existence of relation R was seen to be dubious: consciousness does not consist of overlapping chains of strong connectedness. Why Parfit thought it does is just because he was basing his theory on the old Butlerian account (which is why he also had to invoke the transitive notion of psychological continuity) rather than on the nature of experience.

From here our investigations took us to the bodily theory. Exponents of this view claimed that personal identity is constituted by the spatio-temporal continuity of a person's body. The obvious counter-example was the possibility of bodily transfer. Penelhum attempted to meet this objection by arguing that although it is permissible to say that bodily transfer could occur, i.e. that the identity of the individual who has undergone such a transfer would be determined by the memory criterion and not bodily criterion. ^{we} We need not accept this conclusion. For we could introduce the concept of retrocognition which would allow us to say that one person could remember having the past of another person. In this way we could stick to the primacy of the bodily criterion and

maintain that the identity of a person who has undergone bodily transfer is to be decided not by the person's memories, but by whose body the person has. This is because we can now redescribe the memories of the victim of bodily transfer as an instance of retrocognition of someone else's past. Although this was a notable attempt, the reason it was bound to go astray was simply because it relied on Butler's unacceptable idea that the only way to distinguish between veridical and delusive memories is by prior reference to a person, which for Penelhum meant reference to a body.

Williams contributed to the discussion by introducing the further possibility of brain-state transfer, i.e., the possibility that the information content of one person's brain might be transferred to the brain of one or more different persons. Persons who shared the same total brain-state, and thus memories and character traits, would be identical, said Williams, in the type sense, though not in the token sense of person. Token-persons, or particular instances of type-persons, are still bodies because any statements about them which could not be reduced to statements about bodies could still be made in terms of bodies belonging to certain person-types. But the problem with this procedure was just that if we need to refer to type-persons (i.e. non-bodily based persons) in our statements about token-persons, then the idea of token-persons is parasitic on the idea of type-persons and hence

even token-persons are not bodies. Williams thought he could buttress his position by claiming that to love a person was basically the same as loving a body. But this proved to be false because to be in love with someone is to enter into a complex of reciprocal desires; that is, my love for another must make reference to the other person's desires towards me.

The final reason for the failure of the bodily theory was that it only considers the body from the third-person perspective; i.e. as a spatio-temporally continuous object. But this is plainly not how we experience our own bodies, which appear to us from the first-person perspective as a point of view on the world. Consequently, the bodily theory of identity is one which I can never apply in my own case. And this alone is enough for us to reject it.

This brought us to an examination of the no-self theory which, I argued, is an eliminativist rather than a reductionist view of personal identity. It was in the work of Hume that we found an exposition of this view. However, to fully understand his position we had to defend him not only against his critics but also against those of his apologists who tried to construe him as a reductionist. This required, among other things, an appeal to the Buddhist doctrine of the two levels of truth which, it is arguable, was already implicit in Hume's theory. However, to account for the possibility of what are perceived

as instantaneous states of self-awareness we had to make a separate phenomenological enquiry into the nature of the constructed self-image. We then saw how, through a misunderstanding of the pragmatics of language, someone might come to mistakenly identify himself with the constructed self-image. Because the constructed self-image turned out to be a condensation of what Hume calls a succession of related ideas reflected upon in the imagination, we were able here to bring personal identity regarding the imagination into relation with that regarding the emotions.

I hope that with these arguments I have presented a reasonable case for the rejection of the idea of personal identity. There remain of course questions about what sort of influence the no-self theory will have on our ethical views. For if there is no self, is it permissible for us to continue to employ the apparently kindred notions of praise and blame, responsibility, and so on? For who is it we are praising and blaming? And if the no-self theory cannot accommodate such basic ethical notions, then does this not provide grounds for the theory's dismissal? Although a full consideration of these problems lies beyond the scope of this study, I think we can already see the direction from which they must be approached. In the second chapter, for instance, I referred to Butler's claim that if the strict theory were not true, then a person

should no more care about what will befall himself tomorrow than he would care about what will befall someone else; the implication being that since a person does care more about his own future than about that of another, the strict theory must be true. My response to this was to point out the fallacy in arguing from how people behave to the truth of the beliefs which lie behind their behaviour. Thus just as someone's believing in ghosts or behaving as if ghosts exist does not prove the existence of ghosts, in a like manner a person's behaving as if he has a persisting self does not prove he does. This criticism applies with equal force to the claim that because someone has ethical beliefs which apparently implicate the existence of the self, the self must exist. For it is obvious that a person's attachment to certain ethical views can be for reasons which are quite independent of those views. It seems, therefore, we should not take our supposed ethical intuitions as sacrosanct and proceed to erect a theory of personal identity that kowtows to those intuitions. Not only does such a procedure have to rely on the dubious assumption that there is a universally accepted set of ethical intuitions, but it also has to find a self where there is none. We would do better, I suggest, to first establish what we can about personal identity and then, in light of these findings, proceed to reconstruct our ethical theory. But this is the subject for a further study.

NOTES TO CHAPTER FIVE

1. A Treatise of Human Nature, edited by J. A. Selby-Bigge (Oxford, 1978), p. 525. Further references are to this edition and are included in the text. Unless stated otherwise, all italics within this and subsequent quotes are Hume's.

2. in The Philosophical Review, LXIV, 1955, p. 571.

3. 'Senses of Identity in Hume's Treatise', Dialogue: Canadian Philosophical Review, VIII, 1969, pp. 367-384.

4. Noxon, pp. 372-373.

5. Penelhum, p. 578.

6. Dialogue: Canadian Philosophical Review, III, 1968, pp. 239-254.

7. Ashley and Stack, pp. 244-245.

8. Penelhum, p. 578.

9. Quoted in K. N. Jayatilleke, Early Buddhist Theory of Knowledge (London, 1963), p. 319.

10. To arrive at this translation I have compared both Jayatilleke's translation (p. 361) and that of the same passage by F. J. Woodward in The Book of the Gradual Sayings (Angutta-Nikaya) or more-numbered Suttas (London, 1932), 1, p. 54.

11. Quoted in Steven Collins' Selfless Persons: Imagery and thought in Theravada Buddhism (Cambridge, 1982), p. 154.

12. In Sources of Indian Tradition, edited by Wm. Theodore de Bary, 1 (New York, 1958) pp. 103-105. The following translation comes mainly from this edition. However, the editor omits the important final remark made by Nagasena. This line, which I have included, can be found in The Sacred Books of the Buddhists, XXII, Milinda's Questions, I, translated by I. B. Horner, M. A., (London, 1963), pp. 37-38.

13. Quoted in J. P. Stern, Lichtenberg: A Doctrine of Scattered Occasions (London, 1963), p. 270.

14. Stern does not seem to me to be aware of this point in his discussion of Lichtenberg. See p. 270.

15. There are also the so-called 'second thoughts' in the Appendix to the Treatise (pp. 633-636) where Hume seems to express doubts about the truth of his account of personal

identity. I follow Norman Kemp Smith, however, in thinking that the passage in question is in fact a reaffirmation of his principles and his only doubts are with his theory's ability to account for states of self-awareness. This, as will be seen is not a difficulty for a more expanded version of Hume's account. For Kemp Smith's remarks on the Appendix see his The Philosophy of David Hume: A Critical Study of its Origins and Central Doctrines (London, 1941), pp. 553-560.

16. Kemp Smith, p. 171.

17. Kemp Smith, p. vi.

18. The Second Year: The Emergence of Self-Awareness (Cambridge, Massachusetts, 1981), pp. 148-149.

19. Kagan, p. 149.

20. For a discussion of primary process see The Standard Edition of the Complete Psychological Works, 5, The Interpretation of Dreams, edited by J. Strachey (London, 1964), pp. 588-609.

21. For the Dialogue in which this argument is given see The Collection of the Middle-Length Sayings (Majjhima Nikaya), I, translated by I. B. Horner, M. A. (London, 1957), section 233.

22. See J. A. B. van Buitenen, 'Studies in Samkhya II', Journal of the American Oriental Society, LXXVII, 1957.

23. 'The Transcendental World', Zen Notes, 1, 5 (New York, 1954), p. 12.

BIBLIOGRAPHY

- Ashley, L., and Michael Stack, 'Hume's Theory of the Self and its Identity', Dialogue: Canadian Philosophical Review, XIII, 1974
- Bankei, Bankei Zen: Translations from the Record of Bankei, trans. by Peter Haskel, ed. by Yoshito Hakeda, New York, 1984
- Brennan, A., Conditions of Identity: A Study of Identity and Survival, Oxford, 1988
- Berkeley, G., The Works of George Berkeley, Bishop of Cloyne, ed. by A. A. Luce and T. E. Jessop, 3, London, 1950
- Buddhaghosa, B., The Path of Purification (Visuddhimagga), trans. by Bhikkhu Nanamoli, Columbo, Ceylon, 1956
- Butler, J., The Analogy of Religion, Edinburgh, 1824
- Fifteen Sermons Preached at Rolls Chapel, London, 1949
- Chisholm, R., The First Person: An Essay on Reference and Intentionality, Brighton, 1981
- Collins, S., Selfless Persons: Imagery and Thought in Theravada Buddhism, Cambridge, 1982
- de Bary, W., ed., Sources of Indian Tradition, 1, New York, 1958
- Descartes, R., A Discourse on Method with Meditations on the First Philosophy, and selections from the Principles of Philosophy, trans. by J. Veitch, London, 1934
- Frege, G., Translations from the Philosophical Writings of Gottlob Frege, ed. by Peter Geach and Max Black, Oxford, 1960
- Freud, S., The Standard Edition of the Complete Psychological Works of Sigmund Freud, 5, The Interpretation of Dreams, ed. by J. Strachey, London, 1964
- Standard Edition, 7, Three Essays on the Theory of Sexuality, ed. by J. Strachey, London, 1964
- Friedman, M., The Hidden Human Image, New York, 1974
- Fromm, E., The Anatomy of Human Destructiveness, New York, 1973

- The Greatness and Limitations of Freud's Thought,
London, 1982
- Holstader, D. R., and D. C. Dennet, eds, The Mind's I:
Fantasies and Reflections on Self and Soul, Brighton,
1981
- Horner, I. B., trans., The Collection of the Middle-length
Savings (Majjhima-Nikaya), London, 1957
- trans., The Sacred Books of the Buddhists XXII,
Milinda's Questions, I, London, 1963
- Hospers, J. An Introduction to Philosophical Analysis, 2nd
ed., London, 1967
- Hume, D., A Treatise of Human Nature, ed. by J. A. Selby-
Bigge, Oxford, 1978
- Husserl, E., Logical Investigations, I, translated by John N.
Findlay, 1903
- Jayatilleke, K. N., Early Buddhist Theory of Knowledge,
London, 1963
- Kagan, J., The Second Year: The Emergence of Self-Awareness,
Cambridge, Massachusetts, 1981
- Kant, I., Immanuel Kant's Critique of Pure Reason, trans. by
N. Kemp Smith, London, 1950
- Kemp Smith, N. The Philosophy of David Hume: A Critical Study
of its Origins and Central Doctrines, London, 1941
- Llewellyn-Jones D., A-Z of Women's Health, Oxford, 1983
- Locke, J., An Essay Concerning Human Understanding, ed. by
Peter Nidditch, Oxford, 1975
- Lowe, M. 'Trends in the Development of Representational Play in
Infants from One to Three', Journal of Child Psychology
and Psychiatry, 16, 1975
- Madell, G., The Identity of the Self, Edinburgh, 1981
- McTaggart, J., The Nature of Existence, II, Cambridge, 1927
- Moore, G. E., Principia Ethica, Cambridge, 1951
- Noxon, J., 'Senses of Identity in Hume's Treatise',
Dialogue: Canadian Philosophical Review, VIII, 1969

- O'Flaherty, W., Asceticism and Eroticism in the Mythology of Siva, London, 1973
- Parfit, D., Reasons and Persons, Oxford, 1984
- Penelhum, T., 'Hume on Personal Identity', The Philosophical Review, LXIV, 1955
- Survival and Disembodied Existence, London, 1970
- Perry, J. ed., Personal Identity, Berkeley, California, 1975
- Quinton, A., The Nature of Things, London, 1973
- Oksenberg Rorty, A. ed., The Identities of Persons, Berkeley, California, 1976
- Sartre, J. Being and Nothingness: An Essay in Phenomenological Ontology, trans. by H. Barnes, London, 1957
- Saskei, S., 'The Transcendental World', Zen Notes, 1, 5, New York, 1954
- Shoemaker, S., Self-Knowledge and Self-Identity, Ithaca, New York, 1963
- Identity, Cause, and Mind: Philosophical Essays, Cambridge, 1984
- and R. Swinburne, Personal Identity, Oxford, 1984
- Sprigge, T. L. S., 'Personal and Impersonal Identity', Mind, XCVII, 1988
- Stern, J. P., Lichtenberg: A Doctrine of Scattered Occasions, London, 1963
- Strawson, P. F., Individuals: An Essay in Descriptive Metaphysics, London, 1971
- Streng, F. J. Emptiness: A Study of Religious Meaning, Nashville, 1967
- Thera, N., trans. The Dhammapadam, London, 1972
- van Buitenen, J. A. B., 'Studies in Samkhya II', Journal of the American Oriental Society, LXXVII, 1957
- Wiggins, D., Spatio-Temporal Continuity, Oxford, 1967

Williams, B., Problems of the Self, Cambridge, 1973

Woodward, F. J., trans., The Book of the Gradual Sayings
(Angutta-Nikaya) or more-numbered Suttas, London, 1932