BEYOND PERSONAL IDENTITY

BY JAMES BAILLIE

SUBMITTED FOR THE DEGREE OF PH.D.

DEPARTMENT OF PHILOSOPHY

UNIVERSITY OF GLASGOW

SUBMITTED FEBRUARY 1989

TABLE OF CONTENTS

SUMMARY

This thesis is concerned with what it is to be a person, and with what is involved in being the same person over time. I begin by making a survey of the major theories of personal identity, and mark some important divisions and distinctions between them, primarily between Reductionism and Non-Reductionism and, within this former category, between the Physical and Psychological Criteria, and argue that none of these has proved to be satisfactory. I stress the importance of the work of Derek Parfit, and in particular his shifting of the agenda away from the relation of identity to that of 'Relation R', and his claim that it is the holding of this latter relation - namely psychological continuity by any means - that contains 'all that matters' to us regarding the future, and not necessarily whether I survive. I show how this theory avoids the pitfalls that defeated the other theories, and propose various developments of it.

A critical eye is then cast over the methodology of thought-experimentation, so long the cornerstone of philosophical studies of personal identity, whereby conclusions are derived from considerations regarding what we would say if certain hypothetical states of affairs were to occur. The concept of 'theoretical possibility' is employed in order to determine the limits of applicability of such thought-experiments. Many of the most influential arguments and theories regarding issues of personal identity are found to be flawed due to misuse of this methodology, and their conclusions are judged to be unwarranted.

The remainder of the thesis is concerned with identifying and discussing issues regarding personal identity that remain, once a more modest methodological framework is imposed. These concern the nature and the limits of psychological unity and continuity, and focus on real-life conditions, both typical and pathological, and are rooted within scientific research (particularly the neuro-sciences) rather than imaginative speculation.

My conclusions are for the most part negative, arguing that not only the answers, but the very questions that have been traditionally posed regarding personal identity cease to be relevant, once the flaws in the framework that supported them have been exposed.

CHAPTER 1 INTRODUCTION

1.1 PRELUDE

I was initially drawn into investigations of personal identity through my fascination with the work of Derek Parfit, and his iconoclastic claim that personal identity itself was not essentially 'what mattered'. I saw my task as in developing a Parfitian theory, and, in doing so, I wholeheartedly adopted the method of thought-experiment, as employed by my mentor and so many of his illustrious peers and predecessors. However, in time I began to have serious misgivings over the use of this methodology. It seemed to me that some of the most notable thought-experiments were firstly too outlandish and secondly too vague and unspecified (and that these two factors were integrally connected) to admit trustworthy conclusions.

Wittgenstein famously described the situation of someone who has been bewitched by a misleading philosophical perspective as being akin to a fly trapped in a bottle, forever buzzing around in circles in the illusion of progress, but who only needed to stop and look up, and then the way out would come into view, and be seen to have been there all along. While identifying with this picture, a more accurate metaphor for my own predicament is that of a fly who has been enticed by a drop of jam, and who, finding himself stuck to it, is left with the only option of eating himself free. This is the project of the first half of this thesis - that of working from within the tradition of thought-experimentation, devouring its choice delicacies, even making a few of my own, and concluding that while the feast is very tasty and enjoyable, it is excessive and indulgent, and in need of the application of a few principles of sound nutrition. Within this context, my title of 'Beyond Personal Identity' makes reference to Parfit's dropping of the concept of identity in favour of an analysis in terms of psychological continuity, and also to my advocacy of the primacy of the concept of 'human being' over that of 'person'.

However, rather than urging an outright rejection of the practice of thought - experimentation, I am advocating the practice of extreme caution in its employment. Whilst some notable examples are whole-heartedly rejected, I keep an open-minded agnosticism regarding others, and admit that the situations described therein might conceivably be actualised. In the second half of the thesis I identify and investigate the central issues regarding personal identity once the

unrestricted use of thought-experiments is rejected. As we will see, I argue that many of the traditional issues dominating the subject are dropped, as they gained credence from this discredited approach, and so the whole basis of inquiry into personal identity will need to be reassessed.

In this introductory chapter I will give a brief overview of the dominant theories and analyses of personal identity, and describe some important distinctions and classifications that are employed in the construction of these theories. This will enable me to set the scene for the more detailed investigations of the following chapters.

1.2 Varieties of Identity

In any discussion of identity there are several different concepts which must be clearly distinguished, since confusion regarding them can lead to profound error. It is best to make these distinctions clearly at the beginning. Logicians regard identity as an equivalence relation, that is, it comprises the following relations :
reflexivity : a = a.
symmetry : if a = b, then b = a.
transitivity : if a = b, and b = c, then a = c.
However, the crucial property that distinguishes identity statements from all others is that they must obey Leibniz'Law, otherwise known as the law of the Indiscernability of Identicals : that is, if a = b, then whatever is true of a is true of b and vice versa. Put formally, this reads as $(x)\ (y)\ [(x = y) \supset (\emptyset x \equiv \emptyset y)]$.

The type of identity referred to in the above specifications is numerical identity, which must be distinguished from qualitative identity, otherwise known as 'exact similarity'. (When the term 'identity' appears on its own, I take it to refer to numerical identity). If a and b are numerically identical, then they are one and the same thing, whereas if a and b are qualitatively identical, then they are exactly alike in their intrinsic qualities and properties and in the relationships thereof - but it does not follow that they are numerically identical. Rather, this relationship between a and b might be expressed by saying that they are two tokens of the same type.

Many of the philosophical quandaries regarding personal identity arise from the running together of numerical and qualitative identity. The difficulties are compounded by yet another distinction, namely that between synchronic and diachronic identity. If a and b are synchronic-

ally identical, then they are one and the same thing (i.e. numerically identical) at a given time t; whereas if a and b are diachronically identical, then the relation of numerical identity holds between them over time. That is, they are both stages or 'time - slices' of the same temporally-enduring object.

In cases concerning the alleged holding of synchronic identity, the content of the question 'is a = b ?' often unpacks to reveal a question regarding sense and reference - that is, whether 'a' and 'b' are terms which refer to one and the same object, but (in Russellian terms) under different descriptions, or (in terms of Kripke & Putnam's causal theory) having two distinct causal chains of usage, both grounded in the same object. Recently, another set of problems relating to synchronic identity of persons has come to the fore, concerning such phenomena as split-brain surgery and Multiple Personality Disorder, concerning the issue of 'unity of mind' and the 'copersonality relation' amongst mental states i.e. under what conditions are mental states attributable to the same person at any given time ? I discuss these issues in detail in later chapters.

However, it is undoubtedly the case that the majority of the perennial questions regarding identity, and particularly regarding personal identity have been posed in regard to matters concerning the conditions of the holding of identity through time. The basic fact that everything in the Universe is in motion and constantly undergoing changes raises the fundamental question of how, and under what conditions, can identity through time be maintained in the absence or breaching of qualitative identity, in the face of such unremitting change. Clearly, no-one would suggest that such identity is preserved throughout <u>all</u> possible changes, so we must closely examine the issues regarding the criteria for identity through time for specific classes of entities.

It is wise to clear up some potential confusion regarding the ways in which Leibniz' Law relates to diachronic identity. To take an example : the newspaper, a copy of which was purchased by my father, that published the announcement of my birth to a jubilant world was, on that day (t1), clean and white. That newspaper is now (t2) dirty, yellow and crumpled, in a cupboard in my mother's house. Let us call the hot-off-the-press paper 'a', and the old rag 'b'. I want to say that a = b. But does not Leibniz' Law demand that if a = b, then any property that can be ascribed to a must equally be ascribable to b ? And surely a is white and b is not ?

Such an argument misapplies Leibniz' Law. I accept that if a paper is white at t1 and it is not white at t2, then there is something that is a true predication of it at t1 which is a false predication of it at t2. However, this does not constitute a contravention of Leibniz' Law, as the law does not require that if a = b, then
if a is F (white) at t1 then b is F at t2, but that
if a is F at t1, then b is F at t1.

1.3 REDUCTIONISM AND NON-REDUCTIONISM

In stating this important distinction among theories of personal identity, I will draw on the treatment by Parfit [1984], who amongst recent writers has discussed the distinction in the most depth.
A reductionist theory of personal identity is one that holds the following principles to be true :
1. The fact of the holding of personal identity through time consists purely in the holding of other facts, namely facts concerning physical and / or psychological continuity.
2. A person's existence merely consists in the existence of a brain and body, and the occurrence of a causally interrelated set of mental and physical states.
3. A reductionist can accept that persons exist, and have experiences, but they deny that persons are 'separately existing entities' beyond those existents specified above in 2.
4. These facts in which personal identity consists can be described 'impersonally', that is, without attributing them to any person; so that, while persons exist in the sense described in 3., we could give a complete description of reality that did not mention persons.
5. There can be cases in which the issue of a person's identity through time is 'unpuzzlingly indeterminate'. Of course, numerical identity is an 'all-or-nothing' relation that does not admit of degrees - as Shoemaker [1985] says,
"No ordered pair of entities in the world can be such that it is indeterminate whether its first member is identical to its second" (p450).
Rather, what Parfit is saying is that since persons are not separately existing entities, personal identity should be analysed in terms of physical and psychological continuity, and since these relations do admit of degrees, there will conceivably be cases where a clear-cut answer to the question of identity cannot in principle be given.
Although this factor of the indeterminacy of identity is often presented as a metaphysical matter, relating to some indeterminacy in the world

regarding the relationship between 'a' at t1 and 'b' at t2, it should, on the contrary, be regarded as a semantic matter, where an indeterminacy in the reference of a given sortal concept leads to the formulation of identity statements without determinate truth-values.

6. In such situations, and when all the facts regarding what is constitutive of a particular are known, then any question that remains regarding its identity is empty, that is, it is not a real question regarding different possibilities, but is a case where any apparently conflicting answers are merely different descriptions of the same state of affairs.

7. What matters as regards the future is not the continued holding of my identity per se, but the maintenance of the continuities given in 2.. For instance, Parfit argues that the relation that contains 'all that matters' is Relation R - psychological continuity by any means. Relating this to 6., Andrew Brennan [1988] sees the reductionism / non-reductionism dispute as concerning

"how much information is necessary for telling us all that matters in an identity problem" (p255)

To a reductionist, when all the facts are in regarding physical and psychological continuity, we know enough to determine all that matters even if the question of identity remains unresolved, whereas in such a situation, a non-reductionist would regard the big issue as remaining open until the matter of identity is settled.

A non-reductionist theory of personal identity can be defined in terms of its opposition to this reductionist creed. In particular, non-reductionism holds that persons are separately-existing basic entities apart from a continuing body and brain and sets of mental and physical states; personal identity is a 'further fact' that is irreducible to these other facts; the holding of identity is always a determinate matter, and is itself essentially 'what matters' as regards future survival. Parfit argues that once these conditions are combined, non-reductionism appears to be committed to the existence of something like a Cartesian Ego. In saying that questions of personal identity always have a determinate answer, a non-reductionist is not committed to the claim that we must be able in practice to determine the answer, but merely that this 'further fact' is out there, even if it transcends our practical limitations on verifiability. So, for example, a Kantian belief in a Transcendental Ego located within an unknowable Noumenal world would still count as a non-reductionist view.

1.4 REDUCTIONIST CRITERIA OF IDENTITY

Reductionism is by far the dominant perspective regarding personal identity, and most of the interesting debates have taken place within its confines, concerning which set of facts personal identity is best analysed in terms of. In this, the basic dispute is between forms of the Physical and Psychological Criteria. Both these theories have a long philosophical pedigree, but in their modern variants they are not committed to much of the metaphysical baggage carried by their older formulations. For instance, the issue of dualism versus materialism has been set aside, as contemporary reductionists are all committed to some form of materialism. Rather, the two criteria differ by focussing on different levels of description of events constitutive of identity, or, in particular cases, in opting for the priority of some such events over others in dealing with questions of identity.

The Physical criterion states that x at t1 is the same person as y at t2 if enough of x's brain continues to support a full human consciousness, and is now y's brain, and if no other person z exists at t2 who also has enough of x's brain to support such a consciousness.

For the moment, let us note that this is a sophisticated version of the criterion, with its emphasis on the brain as the essential bearer of conscious experience, in contrast with the Psychological criterion, which focusses on relationships between the experiences themselves. Note also that the Physical criterion is committed to the spatio-temporal continuity of the brain as a necessary condition of identity through time. In ch.4 I will attempt to refute this.

The Psychological criterion states that x at t1 is the same person as y at t2 if the relation of psychological continuity holds between them, and there is no other person z at t2 who is likewise psychologically continuous with x.

The most detailed formulation of this criterion has come from Parfit [1984], and my exposition will utilise some of his technical distinctions. Firstly, psychological connectedness consists in the holding of particular direct psychological connections between mental events. The paradigm examples offered by Parfit are between an experience and the subsequent memory of it, or the forming of an intention and the subsequent action. Connectedness is a relation that comes by degrees, depending on the number of direct connections that hold. Strong connectedness is defined as holding in cases where at

least half the number of direct connections made are preserved until at least the following day.

Psychological continuity consists of overlapping chains of strong connectedness. This preserves personal identity through time in cases where there is an absence of a suitable degree of psychological connectedness holding over longer periods of time than a day, and thus avoids the paradox threatened by Reid's [1785] example of the officer who remembers stealing apples as a child, but who, as a general in later life, remembers having led a cavalry charge as an officer, but not having stolen the apples. (I discuss this in ch.6) So, although most experiences occurring around t1 will be forgotten at t100, they will mostly be remembered at t2 (the next day), and most experiences at t2 will be remembered at t3,and so on, so we have overlapping chains of memories preserving our psychological continuity, and, thereby, our identity.

Different versions of the Psychological criterion disagree over the issue of which causal processes must be responsible for the maintainance of psychological continuity. The narrow Psychological criterion admits only the normal cause - one continuously functioning brain - as suitable. Thus it will give the same answers to identity questions as the Physical criterion. By contrast, the wide Psychological criterion allows any causal process that permits psychological continuity to hold. So, for example, in the teletransportation fantasies so loved by philosophers, (see ch.4), the wide Psychological criterion would permit my replica and I to be identified, but neither the Physical nor the narrow Psychological criteria would agree.

1.5 THE MENU

In chapter 2 I will continue to investigate the dispute between the Physical and Psychological criteria by offering a critical discussion of the major contemporary contributions to the debate.
In chapter 3 I discuss aspects of non-reductionism.
In chapter 4 I pick up the thread of ch.2 again, with particular reference to the issue of sortal- covered identity statements, and the question 'What kind of thing am I ?'.
In chapter 5 I call into question the methodology of thought-experimentation that has figured heavily in the preceeding chapters.

In chapter 6, I discuss the concept of psychological continuity with reference to the central concept of memory.
In chapter 7 I discuss the implications of split-brain surgery for the concept of mental unity.
In chapter 8 I continue this discussion by investigating issues deriving from Multiple Personality Disorder.
I finish by drawing together my conclusions.

As an entry into the contemporary debate between adherents of the Physical and Psychological Criteria, I will describe a puzzle set by Williams [1970] which goes directly to the heart of the matter. I will follow this with one influential response by Nozick [1981], and then with another, more fruitful reply, by Parfit [1984].

2.1 WILLIAMS' DILEMMA

Bernard Williams had previously advocated a Physical Criterion of personal identity, but in "The Self And The Future", the plot thickens. In this article, he describes a thought-experiment in which we seem bound to say that a person has 'swopped' a body and brain for another, and survives as the same person in this new body, and that therefore a Psychological Criterion looks justified. However, he then contrasts another thought-experiment which points firmly to a Physical Criterion. The problem is that this latter case appears to be very similar to the previous one, to the extent of being merely an alternative description of the same state of affairs, yet which leads to a contrary conclusion.

The examples involve that old and trusted friend of philosophers, the mad neuroscientist. In CASE 1, this highly disagreeable fellow has captured two persons A and B, and performs what might be called a mutual total brain-state transfer upon them. In other words, through some advanced technology, he manages to 'record' in some form their entire respective sets of memories, dispositions, etc., which he then erases from their respective brains, and then 'switches' them, so that the mental states that had been realised within A's brain are now realised within B's brain, and vice versa. A and B are the names of the persons prior to this operation, and the resulting persons are called the 'A-body-person' and the 'B-body-person', i.e. the A-body-person is the person whose present body was once A's body, and the B-body-person is likewise related to B.

Clearly, the big question concerns diachronic identity - i.e. are A and the A-body-person stages of the same temporally-enduring person, or are A and the B-body-person ? (As shorthand, I will use the more manageable expression 'is identical with', or '=' in tackling this issue - so the question is whether A = the A-body-person or whether A = the B-body-person.) Prior to the operation, A and B are told that

one of the resulting persons is to be given $100,000, whilst the other is to be tortured. They are asked to choose who will get what, by basing their decisions purely on the grounds of self-interest. If A and B both thought that the operation constituted a 'body-swop', then A would choose that the B-body-person receives the reward and that the A-body-person is tortured, with B choosing the converse. Williams plausibly argues that if the scientist withholds his decision from them until it is made manifest by the operation, whereupon he tortures the B-body-person and rewards the A-body-person, then the B-body person, having A's memories, will complain that that wasn't the outcome he had chosen, whilst the A-body-person will be relieved that the choice that he, as B, made was carried out.

The claim that A = the B-body-person, and that B = the A-body-person looks to be justified even in a case where both victims don't see the operation as being a 'body-swop'. For example, if A supports a Physical Criterion, he may choose that the B-body-person is tortured - but, says Williams, he as the B-body-person, will soon realise that he has made an unwise choice once the torture begins - and, specifically, he will recognise this unfortunate choice as being his.

In CASE 2, I have been captured by this same mad scientist, who plans to use me as the subject of his dastardly experiments. I am informed that I am to be subjected to agonising torture, but that before this ordeal, all my mental states - all my memories, traits of character, likes and dislikes, ect., will be erased from my mind by means of some electronic device, so that I will forget the fact that I am to be tortured. In addition, my lost 'mental set' will be replaced, via some form of programming, with another set of memories, attitudes, values, etc., qualitatively identical to those of some other living person. Now, Williams argues, it seems plausible to say that, from my point of view, anticipating these forthcoming events, the knowledge that I will undergo these psychological alterations does not remove my fear of this imminent torture, nor my conviction that despite all these changes, it is still me who is about to suffer. In fact, if anything, these changes will be perceived by me as adding to my ordeal, as now not only do I have physical agony to cope with, but also mental derangement.

However, by the Psychological Criterion, given that this case involves a complete break in psychological continuity, I should not be

egoistically concerned with the fate of the person who will occupy my body, since he and I are not the same person. But, as Williams says, our intuitions are clear that it is I who is to be tortured, so my fears are well-justified, and the Psychological Criterion, it appears, is not. Yet surely Case 2 is merely an alternative description of Case 1, stated in first-person-singular terms, so that, if I call myself 'A', the conclusion is that A = the A-body-person, in contrast to the previous conclusion.

The other difference between the two cases is that in Case 2, the second person, the 'mind-donor', is depicted as playing a fairly minor part, and the suggestion is that anything happening to him cannot affect the purported A:A-body-person identification. This argument is supported by two principles suggested by Williams in a previous essay [1957]. The first principle states that the question of whether x at t1 is identical with y at t2 can only depend on facts concerning x and y, and any facts about some other z are irrelevant to the issue. The second principle is that if x at ti is identical with y at t2 by virtue of some relation holding between them, then there cannot be any other z who is similarly related to x at t2. If any such z does exist, or even could have existed, then x and y are not to be identified.

However, in Case 1, what happened to the second person is crucial, because we seemed forced to conclude that I am identical to the person surviving in his body after the operation. Yet from the perspective of Case 2, even if we bring the two cases more into line and make the additional move in Case 2 of allowing my original 'mental set' to be realised within the body whose previous mind-set had been donated to me, this does not make any intrinsic difference to what is going to happen to me - I am still going to be tortured, no matter what happens to anyone else. So Williams has presented us with an antinomy, as we have been led to two mutually exclusive conclusions on the basis of the same premises. Williams admits himself to be puzzled, and offers no resolution.

2.2 NOZICK'S RESPONSE

Robert Nozick suggests a way of compatibilising our divergent responses to the two cases, by denying the applicability of Williams' two above-mentioned principles, which he regards as inapplicable not just to 'problem case' counterfactuals like the above, but to all

possible cases, so that, on the contrary, the issue of whether x at t1 and y at t2 are identical will always depend on who or what else is present at t2. Nozick offers the CLOSEST CONTINUER theory of identity (including personal identity), which states that :
x at t1 is identical with y at t2 if y's properties grow out of, are causally dependent on x's, and if there is no other z at t2 that stands in a closer (or as close) relation to x. Corollary to this 'closest continuer' relation is the 'closest predecessor' relation, whereby for such an x and y to be identified, x must be y's closest predecesser - y cannot more closely continue some other z existing at t1 than it does x. If x is y's closest predecesser and y is x's closest continuer, then x and y are said to be 'mono-related', and the identity relation holds between them.

Now Nozick thinks that as a matter of fact we always do make judgements regarding matters of identity in accordance with this theory (presumably subconsciously), as it provides answers to all hypothetical problem cases that are in accordance with what our intuitions advise us. However he backtracks significantly to say that the theory cannot by itself solve identity problems, since it does not specify *"what dimension or weighed sum of dimensions measures closeness" (p33).* Take the famous example of the ship of Theseus. In the course of time, a ship gradually has all its component parts replaced, so that eventually not one original part remains. Yet throughout this time, and all these repairs and replacements, it remained afloat and functioning as a ship, and still bore the same name. However, in the intervening period, the original ship-parts were collected and stored, and once the entire set had been gathered, a ship-builder restored them to their former roles, and relaunched the reassembled ship. So we now have two ships, each with a claim to be the same ship as the original one. So, in Nozick's terms, the two 'dimensions' are firstly spatio-temporal continuity as a functioning ship, and secondly numerical identity of parts. and Nozick's theory cannot tell us which of these considerations should take priority.

Also, the Closest Continuer theory can at most offer a necessary, but not a sufficient condition of identity, as y can be x's closest continuer, yet not be close enough to x for the identity relation to hold. This 'minimum closeness requirement', amd the dimensions along which it is measured will, says Nozick, vary according to the kind of object under consideration.

Nozick lists a number of classic puzzle cases regarding personal

identity to show that his theory gives solutions to them that accord with our intuitions. For example, if A at t1 has his brain-states copied and programmed into a clone body, then this clone, B, is not identical to A, as A's closest continuer at this time t2 is A himself. However, if such a brain-state transfer took place and A were to die at this point, then B would be A's closest continuer on account of psychological continuity, and thus would be judged identical to A.

Nozick never spells out what his solution to Williams' dilemma is, but it seems to be that in Case 2, A's concern for the A-body-person is due to the fact that this person is, in the absence of anyone else, A's closest continuer, whereas in Case 1, psychological continuity overrules this physical continuity, so that the B-body-person is A's closest continuer. This, of course, gets us nowhere, as it merely restates the problem, namely, given that the two cases are virtually alternative descriptions of the same state of affairs, how do they lead us to such opposite conclusions ? Why do different 'dimensions' take precedence in the two cases ? I will have more to say on Williams' poser later in this chapter, and also in chapter 5, where I argue that the real significance of the dilemma is very different from that which Williams intends. But firstly I must continue with my criticism of the Closest Continuer theory.

Not only do I deny that it is compatible with our intuitions in all possible cases, but also, I argue, that it gets itself into such a muddle in attempting to do so that it loses all credibility. One situation that exposes its weaknesses is that of a 'tie', that is, where <u>both</u> y and z at t2 are equally close continuers of x at t1, along the same dimensions. Nozick stresses that his theory is strictly a <u>Closest</u> Continuer theory, which does not allow identity to hold with a mere 'continuer none is closer than' (if such was permitted, x would survive in case of a tie) and that in such cases, <u>neither</u> y nor z is X's closest continuer, and therefore x ceases to exist.

So, for example in Parfit's story entitled 'My Division', in which A's cerebral hemispheres are removed, separated, and transplanted into the skulls of his fellow triplet siblings B & C, A ceases to exist. This is surely counterintuitive, given the physical continuity and the subsequent psychological continuity supported by each hemisphere. Nozick claims to be driven to his conclusion by the fact that all the other

alternatives seem to him to be even more inadequate. For example, it cannot be that both B & C are identical with A, as it would follow, as identity is a transitive relation, that B & C are identical with each other, and they are not. Neither can it be that oneof B and C is identical with A, as, since B & C are qualitatively identical to each other, any reason for identifying A with one of them must equally be a reason for identifying A with the other, especially given that both are causally related to A in exactly similar ways. Obviously one is not justified in just stipulating that A is identical with one of the two, but for no reason. He also rejects the solution proposed by Lewis [1983] (see ch2.9)which says that while B & C at t2 are different people, A at t1 also consisted of two different people.

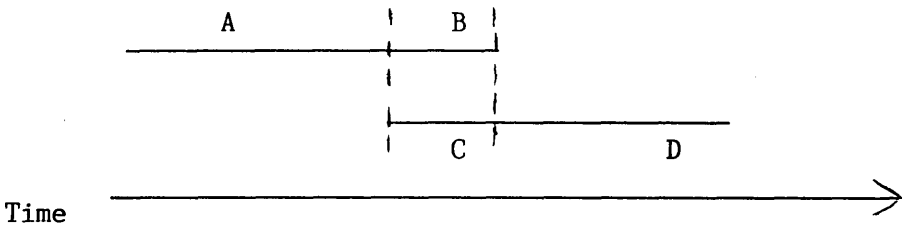
But surely Nozick's own solution is equally hopeless. If A is not identical-through-time with either B or C, then surely he should not be self-interestedly concerned with whatever happens to either of them. In other words, if he is informed that B will be subjected to torture at t2, he shouldn't be concerned for B in any way over and above the humanitarian concern for another person, or even the added concern for the well-being of loved ones. However, we are strongly inclined to say that A would, and should be egoistically concerned for B. Nozick tries to deal with this difficulty by saying that the degree of care that any x at t1 feels for any y at t2 will be proportional to the closeness between them along the appropriate dimensions, with the exception of cases where y is x's closest continuer, as x cares about his closest continuer in a special way, not proportional to closeness (as long as it satisfies the minimum closeness requirement). So Nozick seems to be saying that A will not be concerned with ceasing to exist, as what matters in the situation is that someone exists who continues A closely enough to have been identical with A if he had been x's closest continuer. So, in such a tie situation, A will be concerned about B and C equally, and in proportion to their closeness to him, but, had B been his closest continuer, yet intrinsically no closer to A than he is in the tie situation, then the degree of concern accorded to him would increase.

But how can it be that A ceases to exist in a tie situation ? If I were A, and my brain were split and transplanted, wouldn't I be able to affirm, from this new body, or bodies, "I do not exist"? This obviously can't be right. Nozick cannot mean just that A does not exist, but that he does not exist qua A. This leads to the possibility that if C were removed and vanished without trace so that I didn't know whether the

other transplant took place, then I, as B, wouldn't know whether I survived the transplant and whether I was identical with A.

Another problem that Nozick has to face involves cases of 'overlap', where, for example, one of my cerebral hemispheres is transplanted into a clone body, and I retain the other one, and I die at a later date, so that there is a partial overlap in our lifetimes. Now if I had died immediately after the hemispherectomy, the clone would be my closest continuer, and therefore identical with me. Yet, Nozick says, if I had survived for several more years, then I (the J.B.-body-person) would be my closest continuer within that time, and after my death, the clone would not be identical with me. But clearly the idea of a sharp dividing line specifying the overlap of lifetimes that cannot be crossed if the clone and I are to be identified is absurd.

Surely this example just exposes the Closet Continuer theory to the same problems as defeated the theories that Nozick rejects? He tries to get out of the trap by saying that the difficulty here is due to a tension between, on the one hand, our inclinations towards the Closest Continuer theory in its 'local' form (which is the theory already described) and what he calls its 'global' form, which states that x at t1 is identical with y at t2 if y is x's closest continuer and if there is no longer extended thing z which is more closely continuous with x than any other equally extended thing of which x is a part. So in an overlap situation, which can be depicted as below,



Where A represents myself before the hemispherectomy, B is myself after this operation, C is the clone while I am alive, and D is the clone after I am dead, we have a situation where
(i) B is A's closest continuer, and
(ii) D is the closest continuer of A+B, yet
(iii) C is D's closest predecessor, and
(iv) A is the closest predecessor of C+D,
so neither A+B+D nor A+C+D are mono-related, and cannot be stages of the same continuous entity. However, A and D are mono-related, and the global theory says that where B and C are relatively small

compared to the time-span of A and D, then A+D constitutes a single entity, albeit a temporally discontinuous one.

The first point that should be made about this attempted escape route is that Nozick can no longer claim to have the support of our intuitions regarding questions of identity through time. Our intuitions are rather simple fellows, and Nozick's theory grows more convoluted by the minute. More seriously, his ad hoc introduction of the global variant still does nothing to deal with the absurdities of the 'sharp dividing line' that scuppered the local theory. All it does, at best, is to provide a diagnosis of the Closest Continuer theory's inadequacies.

2.3 THE PSYCHOLOGICAL SPECTRUM

My own view is that if you accept that the thought-experimental 'problem cases' have a real bearing on the issues of personal identity, then any attempt to deal with these cases in terms of the concept of identity will fail, and that a more fruitful approach is to follow Parfit's line and analyse the cases in terms of Relation R or, as Andrew Brennan [1988] puts it, 'survival'. The Closest Continuer theory can retain its insights when it is interpreted as a theory of the conditions of survival rather than of identity - where, briefly, x's closest continuer is whoever survives x to the greatest degree, and where, in cases of a tie, there is, in Parfit's words, a 'double success' rather than both contenders 'cancelling each other out', as when the case is described in terms of identity. I will focus on Brennan and his concept of survival in detail in ch.4, but for the moment I will concentrate on Parfit.

Let us now return to Williams' Case 2, which pointed to a Physical Criterion, and consider Parfit's response. Parfit develops Williams' story to devise a whole range of cases (which he calls the 'Psychological Spectrum'), each involving the disruption of psychological continuity to some degree. Williams' case is at the far end of this spectrum, as it involves a total break with any psychological continuity. A near-end case would involve only a minimal loss of memories, character traits, etc., and correspondingly minimal addition of new ones. In between these extremes would be a vast number of cases, each involving a tiny degree more psychological change than its predecessor.

By application of the Sorites paradox, we can start with a near-end case, involving negligible change. If we agree, as seems to be extremely plausible, that (calling the pre-operative person A and the post-operative person B),that A and B are the same person, then, given that the spectrum can contain any number of possible cases according to how close we make the contiguous cases, then if the next case (calling the first one 'n' and the next case 'n+1') involves an infinitesimal increase in psychological change to what was involved in case n, then surely A = B in case n+1, as there is virtually no difference between the cases. Someone taking this line could draw on Williams' [1957] rule that something so fundamentally important as a change in, or loss of identity could not rest on such an insignificant matter as the minute difference between cases n and n+1. However, the paradox arises when we apply this procedure to each consecutive pair of cases right through the spectrum, for we seem to be forced into making the conclusion that A = B even in Williams' extreme far-end case where there is no psychological continuity between A and B. (Obviously these Sorites problems are not unique to matters of personal identity). At this point it looks as if Parfit is in agreement with Williams regarding the weakness of the Psychological Criterion. However this appearence is deceptive, and his argument so far is merely a prelude to his main arguments and their much more radical conclusion.

Parfit leaves the Psychological Spectrum for the moment. Yet it is instructive to note that he could have pushed the argument harder to allow the paradoxes to pile up. If we apply the Sorites rule again starting this time with Williams' extreme far-end case 2 (let us call it case n100), then, if a defender of the Psychological Criterion says that in such a case that A ≠ B, then he would seem to be obliged to make the same judgement in the second-last case n99, and likewise all the way back down the spectrum to case n. So here we have obtained the opposite conclusion to Parfit, by another application of the same technique.

2.4 THE PHYSICAL SPECTRUM

As a pre-taste of his own line, Parfit comments that the paradox he has exposed is not fundamentally due to any flaw unique to the Psychological Criterion, but is due to the underlying assumption that there is a determinate answer to questions of identity through time in all

possible cases. This is the assumption that Parfit attacks, arguing that at least in the middle cases in the Psychological Spectrum, any such questions are empty. He then turns the tables on supporters of the Physical Criterion by showing that a similar argument to that applied to the Psychological Criterion can be constructed to derive paradoxes from the Physical Criterion, if we accept this assumption that identity is always a determinate matter. A 'Physical Spectrum' is devised, wherein a scale of minute gradations of change are applied to A's body, where his brain and body cells are replaced by exact duplicates. So in the near-end case n we have someone, B, who is virtually both physically and psychologically identical with A, and so we would appear to have overwhelming reason to say that they were the same person. In the far-end case, the resulting person will be psychologically continuous with the original, but in no way is he physically continuous, as 100% of his cells have been simultaneously destroyed and replaced (as in Parfit's Teletransporter story - see ch 4) thus breaking a necessary condition of the Physical Criterion, so by this Criterion, we seem forced to say that A is not the same person as his 'duplicate', B.

If one adhered to a Physical Criterion, and also held that identity was always a determinate matter, then one would be committed to the claim that identity is preserved only if a critical percentage of cells remains unreplaced in cases such as the above. Even if the story was recast to involve only replacement of neurons, by claiming that the brain is the only part of the body that is relevant to the matter of identity, then the problem is unsolved, and merely shifts back a step to the claim that the holding of identity through time depends on the retention of this critical percentage of neurons (or back a step further, if we restrict the issue to some privileged set of neurons, such as those of the cerebral cortex).

But how are we to establish what this critical percentage of cells is, in a way that is not arbitrary and stipulative? No independent test could ever confirm or falsify the credentials of such a choice. The only possible independent test would have to be 'from the inside', that is, from B's own testimony. But, since in all cases in the Physical Spectrum, A is completely psychologically continuous with B, in all cases he will claim to be the same person as before. So following through the logic of Parfit's argument pushes us into

accepting a wide Psychological Criterion as appropriate for personal identity.

2.5 *THE COMBINED SPECTRUM*

Parfit employs another argument against the Physical Criterion, this time against unpublished arguments by Thomas Nagel (described in Parfit [1984]), which I will consider shortly. But first, mention should be made of Parfit's 'Combined Spectrum', which, as its name suggests, combines the essential features of both the Psychological and Physical Spectra, whereby, to varying degrees, Parfit's brain and body cells are replaced by exact duplicates of those of Greta Garbo, so that his psychological properties will also be accordingly replaced by hers. So the Combined Spectrum involves cases with various degrees of physical and psychological continuity holding. As expected, when Sorites techniques are applied as before, the paradoxes start to pile up, urging Parfit to recommend the utility of the view that since there is no 'further fact' that personal identity consists in over and above those of physical and psychological continuity, then in the above-mentioned cases there is no real content to the question of whether the pre- and post-operative persons are to be identified.

Geoffrey Madell [1985] has contested Parfit's claim that if some of Parfit's cells were replaced by cells identical to those of Garbo, then there is no fact of the matter as to whether the resulting person is Parfit or Garbo. Madell says that this impersonal interpetation of the thought-experiment is not *"the most natural and obvious way"* of understanding it, which is that it should be seen as a case of *"two minds in one skin, a pretty startling case of Multiple Personality, perhaps" (p107)*, so, given that the two sets of mental states, those of Parfit and of Garbo, will remain distinct and non-integrated, there is no one person who is identical with neither, and thus, he reckons, Parfit's claim that personal identity is a matter of degree loses its justification.

Now it seems to me that this argument from Madell misses the point, as he is not so much criticising Parfit's conclusions as describing a different story, one in which the token-Garbo neurons are added in such a way that the mental states embodied therein come to constitute an autonomous personality, distinct from Parfit's own, i.e. as the Garbo neurons are added, they somehow fail to become functionally integrated with Parfit's own, so that they establish a

nexus of memories, attitudes, etc., to the extent that they themselves constitute a 'mind'. Note that Multiple Personality Syndrome involves *"two or more personalities, each of which is so well-developed and integrated as to have a relatively coordinated, rich, unified and stable life of its own" [Taylor & Martin 1944 (p282)]*. But rather than offering the more natural and obvious way of describing the outcome of Parfit's story, he has altered it, introducing a different outcome, and offered the most obvious interpretation of it.

Madell himself is a non-reductionist, arguing that personal identity rests on a further fact irreducible to either physical or psychological continuity. However, I fail to see how his suggested outcome, as stated above, provides any support for such a view. Also, his non-reductionism would commit him to saying that a middle case in the Physical Spectrum, where the resulting 'mind' is formed using duplicate neurons, would lead again to a case of two non-integrated minds, and therefore another case of Multiple Personality, albeit a peculiar form, as such a divided 'mind' would be qualitatively indistinguishable from that of the pre-operative person.

2.6 THE PHYSICAL SPECTRUM AGAIN

Returning to the Physical Spectrum, it does seem implausible to say that the replacement of a few more neurons can have the tumultuous consequence of a change of identity. However, John Robinson [1988] disputes this claim, saying that such a process need not involve just 'one more small psychological change', since some of our psychological properties are essential to our identities, and others are inessential. If the neurons underlying these essential properties were replaced, then a change in identity would be effected.

However Robinson gives us no indication of which psychological properties he is referring to, nor any argument to support his claim that there are such properties. He merely admits that a satisfactory analysis of the issue is long overdue, and that this *"cannot be done independently of a theory about the nature, unity and continuity of consciousness" (p328)*. Agreed but, I stress, the onus is on him to identify these properties, or at least to offer some argument to the effect that some such properties must exist. In the absence of such submissions, his claims can be safely ignored as having no real

content.

Parfit considers another variant on the Physical Criterion, this time from Thomas Nagel, who, (as a Reductionist who thinks that I am essentially my brain), also holds that personal identity is the relation that contains 'all that matters', as Parfit puts it. Parfit tries to show the inconsistency of this position by another thought-experiment : I have been diagnosed as having an incurable and fatal brain disease, and an operation is necessary. I can have a choice of two operations:

The first operation, O1, is really a sequence of one hundred mini-operations, each involving the removal of one percent of my brain, and the replacement by the appropriate non-diseased neurons. The second operation O2 involves the removal of the whole brain in a single operation, and the transplantation of a whole new non-diseased brain.

In O1, each new percentage of replacement neurons is added to the original brain, and becomes structurally and functionally integrated with it, and so can be legitimately regarded as part of this brain, and this brain be legitimately regarded as preserving my identity throughout the part-replacement. (Incidentally, one can easily construct an analogue to the puzzle of the ship of Theseus by imagining that some scientist had kept the diseased cells, kept them alive in some way, and reassembled the original brain, perhaps in a clone body). However, the situation is different in O2, as the old brain has been destroyed. Parfit claims that Nagel must say that I survive in O1 but not in O2. But since in both cases there is full psychological continuity, with only the means of achieving this being different, surely this difference is a trivial matter, and how could it make such an impact as to mean the difference between life and death? Thus Parfit is led to reject Nagel's position.

Andrew Brennan [1988], while defending a Parfitian position, argues that we needn't describe O2 as involving the destruction of my brain, claiming that by his 'conditions for survival' (see ch.4), my brain survives to a high degree in its replica. The production of this replica is a copying process using my original brain as the prototype, so the two brains are causally related. We can make the additional point that my brain and its replica in O2 are two tokens

of the same type 'my brain'.

John Robinson offers a telling counterexample to Parfit's argument on the Physical Spectrum, and to his claim that a replica contains 'all that matters' to me regarding my future. He considers a case where my neurons are replaced not by artificially constructed duplicates, but by those from the brain of my twin, who is completely identical to me right down to the cellular level, so that, since our brains are identical, then we will likewise be identical on the psychological level also, so any such transplant will not effect my psychological continuity. (Even if we dispute his reasoning here, we can simply recast the example, and say 'and if we were psychologically identical'). Robinson says that while it is plausible to say that I survive in a near-end case, this would not be so in a far-end case when all my neurons are replaced, as this constitutes a brain transplant, with my twin's brain being relocated in my body, so surely the surviving person is him and not me. And since this takes place with the retention of full psychological continuity, then this cannot be constitutive of personal identity.

Note that in this case, as with Parfit's duplication cases, each neuron and its replacement are tokens of the same type - but the crucial difference in Robinson's example is that the two are not causally related. The twin's neurons are not copied from the prototype of my neurons. Robinson's case seems closer to those depicted in the Combined Spectrum. I will have more to say on these examples in ch.5, where I discuss the use of thought-experiments as a methodological tool.

2.7 MY DIVISION

Parfit offers another set of powerful arguments to back up his advocacy of Relation R. These are again based on thought-experiments depicting situations that are presently beyond our practical capabilities, but, he argues, could become a reality in the future. In the chapter entitled 'My Division', he starts by considering a case involving identical twins, A and B, who are seriously injured in such a way that A's body is damaged beyond repair but his brain is still functioning normally, whereas B has a seriously damaged brain but is otherwise uninjured. Now, by both the Physical and Psychological Criteria, if A's brain were transplanted into B's body, the resulting

person would be A, as the receiving of a new body can be seen more accurately as a limiting case of organ transplant. This highlights what is reckoned to be unique about the brain in matters of identity, (and shows the immediate plausibility of Nagel's position),that if B receives A's heart, lungs, liver and kidneys, the resulting person is B, but if A's brain is transplanted, the resulting person is A, and in fact it is more accurate to say that it is A who is the recipient here and not the donor.

Parfit now asks us to consider the fact that substantially less than a complete brain can support a full human consciousness - for instance a stroke victim can suffer the loss of function of an entire cerebral hemisphere, yet relearn the associated abilities to some extent by the remaining functioning hemisphere restructuring its 'division of labour' and taking them over. Likewise, people have survived hemispherectomy. Next, Parfit notes that a small percentage of people do not have a pronounced functional division whereby language-related functions are dominantly carried out by one hemisphere only. His examples in 'My Division' concentrates on some such person, whose linguistic functions are performed with equal input by both cerebral hemispheres.

Now imagine a case involving identical infant triplets A,B and C. In an accident scenario similar to the previous example, A survives with a healthy brain and a seriously damaged body, whilst B and C are seriously brain-damaged but otherwise uninjured. Realising that if nothing is done they will all die or at least have drastically impaired lives, surgeons decide to cut their losses. They remove and divide A's brain, so that one cerebral hemisphere is inserted into the skulls of B and C. By analysing the outcome in terms of the concept of personal identity, Parfit says that we have four possibilities. (In ch.2.9 I consider another, proposed by David Lewis [1983]).

1. A does not survive.
2. A survives as the B-body person, and is identical with him.
3. A survives as the C-body person, and is identical with him.
4. A survives as, and is identical with, both the B and C-body persons.

He argues that all these outcomes are deeply unsatisfactory, and that such cases are analysed much better in terms of Relation R.

Starting with option 1., we have already accepted that A survives if his full brain is transplanted into B's body. Secondly, we've

accepted that A would have survived with full psychological continuity if he underwent a hemispherectomy. Combining these considerations, we must conclude that if one of A's cerebral hemispheres were transplanted into B's body, the resulting person would be A. So, given this, how could anything happening to A's other hemisphere affect this survival? (Contrast with Nozick) if anything, we appear to have a 'double success', with A surviving twice over.

Options 2 & 3 can be treated together, as they are two sides of the same coin. The problem is that since any argument in favour of one possible outcome applies equally to the other, we can have no rational grounds for choosing one over the other. This leaves us with option 4. The problem here is that the result is contrary to the logic of the relation of identity, which is necessarily one : one - i.e. any x at t1 can be a stage of the same enduring object as only one such stage y at t2. Since identity is a transitive relation, then, if A = the B-body person and A = the C-body person, then the B and C-body persons are likewise identical, but this is not a coherent option. It is nonsense to say that A survives as one person with a divided mind and two bodies. The problem is not that it drastically distorts our concept of a person. I have no objections in principle about that, as I believe that the concept is vague and, as Brennan argues, capable of sustaining several incompatible developments. But this development is clearly a non-starter. The B and C-body persons could be separated at birth and never hear from each other ever again, so the claim that they nevertheless constitute one person is hollow and worthless, adding nothing to our understanding of the situation. This case is different from the one described by Parfit in the chapter entitled 'My Physics Exam' (see ch.7), where the plausibility of the claim that it involved one person with a divided mind came not only from the fact that only one body was involved, but because the division was temporary and reversible.

Another red herring that Parfit easily dismisses is based on Wiggins' [1980] example that the Pope's crown contains three smaller separate crowns, but is itself a crown over and above these three. Presumably the analogous claim is that the B and C-bodied persons are distinct people, but that there is a third person also, which is an integration of them. Parfit quickly refutes this, pointing out that such a 'person' would have no mental life over and above that of the

B and C-body persons. We can also mention here that the analogy fails to appreciate that natural and artefactual kinds are individuated according to entirely different principles.

John Robinson [1988] challenges Parfit's description of his thought-experiments, and his assumption that the two cerebral hemispheres (the two 'half-brains') together constitute one whole brain, and likewise his description of hemispherectomy patients as having survived with 'half their brain destroyed', since in such cases the brain-stem must also be considered : *"If we were to transplant the removed hemisphere into someone else's brainless body, we would not create a second person. It is only if we were to transplant the hemisphere into a body that already had some brain in it, the brain stem, that there would be another person as a result. Once this feature of the puzzle case is highlighted, then it becomes quite clear that the relation between a surviving subject before and after the removal of one of the hemispheres is not duplicated in the case where the surviving subject's hemisphere is removed and placed in someone else's skull. In the former case, but not the latter, there is the persisting brain stem to consider, and the psychological abilities and capacities that supervene on it" (p324).*

Can Parfit defend himself against this criticism? Perhaps he could say something like this : while obviously granting that the brain comprises a number of highly interdependent sub-systems, with the functioning of each part depending on the integrity of the other parts, so that the possibility of conscious experience will depend on the functions of the brain stem, it is a mistake to suggest that the brain stem is the bearer of any of the higher mental functions that are usually regarded as the elements of psychological continuity, and therefore of identity. On the contrary, the functions of the brain stem operate below the level of conscious awareness, dealing, for example, with matters of balance, control and movement, and it would take a strong argument, which Robinson does not provide, to say why these functions should be considered in matters of personal identity any more than, say, the functions of cleansing the blood or of metabolising fats. So when A's hemisphere is transplanted into B's skull, there is no reason to suggest that B's brain stem should be granted any weight regarding the question of identity, any more than does B's kidneys or liver. The resulting person is clearly A.

In reply to this, I have devised a thought-experiment that would

seem to count against Parfit, and add to the chaos that is being brought about by the use of such fictions. Imagine that the Scottish Football Association, in a fit of desparation, has abducted both myself and the Argentinian genius Diego Maradona. A neurosurgeon has been employed to perform a Parfitian brain-swop, whereby our respective cerebral cortexes are removed and placed in the other's body. Now Parfit and many others would be inclined to say that by both the Physical and Psychological Criteria, I would be the DM-body-person, who will spend his time pondering over weighty philosophical tomes, while the JB-body-person emerges with memories of the backstreets of Buenos Aires. However, the finely tuned balance and poise required of a top-class athlete are within the domain not of the cortex, but of the cerebellum, which has not been transplanted - so, whilst by the Psychological Criterion, Maradona would be the JB-body-person, his football manager would not agree with this judgement if his skills were of the mediocre level of the impoverished apprentice philosopher. From his professional perspective, the one who is Maradona is the one who plays like him, no matter what he does or doesn't remember.

One might reply that it is misleading to talk in this way, making such a clear-cut division between experience memory and procedural memory, saying that one can be lost or replaced by different contents, without the other being affected in the slightest. But, on the other hand, such a distinction is reflected in the plight of those suffering from amnesia due to herpes simplex encephalitis (see ch.6). One such victim, Clive Wearing, has suffered a complete loss of long-term experience memory, yet his practical skills, such as playing the piano, have remained intact. Of course, he doesn't know that he can play the piano until he actually does it, so perhaps it would be irrelevant that the DM-body-person would be dragged reluctantly from his books onto the pitch, protesting that he really wasn't all that good, if these skills came alive once the game started.

Leaving these doubts aside for the moment, let us return to Parfit and 'My Division'. Since he has rejected all possible outcomes described in terms of personal identity, he argues that we are forced to recognise that in such cases the concept is inadequate and hence inappropriate to describe what is going on. In such cases, even if we had all the relevant information, being able to describe the operations down to the finest detail, none of this would enable us to settle the

questions of identity through time, so, in Parfit's terms, such questions are empty.

Consequently, it is irrational to regard such an operation as being 'as bad as death', since the relationships between A and the B and C-body persons contain everything that really matters regarding survival, namely Relation R. The fact that in normal cases, Relation R does not take a branching form, is of no concern, as branching does not alter the intrinsic quality of the relation, nor does it contravene its 'logic', as Relation R needn't be 'one : one' as does identity. According to Parfit, we are led to the mistaken belief that personal identity is the relation that contains all that matters in survival because in all actual cases it coincides with Relation R. The value of the deployment of these thought-experiments is to devise situations where the relations diverge, and we have to choose between them.

2.8 THE 'ONLY X AND Y' RULE

Williams [1957] develops another argument that can be put against any Psychological Criterion. It arises out of a remark by Thomas Reid [1785] against Locke's claim that *"whoever has the consciousness of past and present actions is the same person to whom they belong"*. To Reid, it seemed that *"if the same consciousness can be transferred from one intelligent being to another..... then two or twenty intelligent beings may be the same person" (p114)*. Williams begins by acknowledging that identity is logically a 'one : one' relation, and so I cannot be identical with two future persons at any given time, and so if it were possible that two future persons were psychologically continuous with me, then psychological continuity cannot constitute a necessary or sufficient condition for the holding of personal identity, since a criterion for a 'one : one' relation must also take this form.

He adds that even if an advocate of the Psychological Criterion were to compromise and propose non-branching psychological continuity as a criterion, this will still not do, because any criterion of personal identity must satisfy two requirements (which, following Brennan, I have called the 'only x and y' rule). The first requirement (which I described in ch2.1) states that the issue of whether x at t1 is the same person as y at t2 must only depend on factors intrinsic to x and y - what happens to anyone else is irrelevant and cannot affect the issue. The second requirement states that since the matter of identity is of the utmost importance, then its answer cannot depend

on trivial matters.

So, by a Lockean theory, if I were to wake up one morning with memories of Napoleon's life, then I would be Napoleon. But, looking back at Reid's remark, it seems no more implausible to suggest that if this can happen once it can happen any number of times, so, for example, if you were also to acquire these memories, then you too would be Napoleon, by Locke's criterion. But we can't both be Napoleon, not only because identity is a one : one relation, but because it would follow that you and I would be identical. Reductio ad absurdum.

Parfit considers his Teletransporter case from the point of view of Williams' requirements. In it, I enter a machine which records the exact states of all my cells, destroys my body, and relays the full description of my cells to Mars, where this information is used to create a qualitatively identical body out of new matter, while preserving psychological continuity, since the entire process takes place at the speed of light. In such a case, by a non-branching wide Psychological Criterion, I survive as my replica. But what if my 'blueprint' were also sent to Io, a satellite of Jupiter, but stored there without making another replica? By such a criterion, I would only survive as the same person on Mars as long as no such other replica were created. But no intrinsic change would be undergone by the Mars-replica should the Io-replica be created. The point is that the question of whether or not I survive cannot, by Williams' requirements, depend on events in Io involving someone else. Williams would also add that since the Mars-replica isn't the same person as the Io-replica, we should deny that any singly-existing replica is identical with the pre-transported me.

Let me now make a slight diversion back to the previous example, which left us both strutting around in an advanced state of megalomania, eulogising Josephine. If we were to revise a Lockean theory, developing the concept of memory to allow it to operate through various different causal mechanisms (and calling it 'quasi-memory'), then we could avoid the contradictions by saying that I quasi-remember Napoleon's experiences, but that I am not the same person as Napoleon, and likewise in your case.

Returning to the main theme, Parfit hoists Williams with his own petard by using his argument to challenge the Physical Criterion. In fact he goes much further, arguing that no criterion of personal

identity could satisfy Williams' requirements in all possible cases, because the matter of identity will sometimes depend on what happens to others, or on some other trivial matter. He devises another thought-experiment involving our hapless triplets A, B and C, where A's cerebral cortex (or 'brain', as he puts it) is removed and divided, and each hemisphere is taken off to a separate theatre for transplant into B and C's bodies. A supporter of a non-branching Physical Criterion would say that if one hemisphere were successfully transplanted into the B-body, and the other operation were to fail, then A survives as identical with the B-body-person; but that if, on the contrary, the second operation were successful, then A does not survive as either resulting person. But such a claim violates Williams' requirements, as he is committed to saying that if A = the B-body-person, then anything happening to the C-body-person cannot affect this identity. To claim otherwise is to allow a situation where, if there is a one-hour delay between the two operations, A survives as the B-body person for one hour, until the first signs of neural activity in the C-body-person.

The only recourse for a defender of the Physical Criterion seems to be to modify it to prevent the possibility of it taking a branching form, for example by saying that x at t1 is identical with y at t2 iff y has a full human consciousness and more than half of x's brain. This criterion was suggested by Wiggins [1980], and since it clearly can only take one : one form, it satisfies Williams' first requirement, not his second. If we assume, in accordance with philosophical tradition, that by 'brain', Wiggins means'cortex', then possession of more than half a cortex is not a necessary condition of having a full human consciousness, as is shown by hemispherectomy patients, which opens the door for branching again. But also, we can see that a tiny difference in the percentage of neurons may separate a 'less than half-brain' from a 'more than half-brain' (i.e. 49.99.....9% v 50.00.... ...1%) and such a difference is trivial, if the two such brains function equally well. Otherwise we would have the situation of a hemispherectomy patient losing his identity if a minute fraction of his remaining hemisphere were removed or destroyed, without his mental abilities having been affected.

Because of these considerations, Parfit says that neither the Physical nor the Psychological Criterion can meet Williams'requirements

and, furthermore, such problems are applicable to any possible Criterion. However, if we examine the abovementioned cases in terms of Relation R, the problems and paradoxes fall away. By turning away from all talk of personal identity,, then in the 'My Division' cases, we don't bother even asking if A is the same person as either the B or C-body-persons, but rather say that A is R-related to both. (In shorthand this might be expressed as ARB & ARC) In advocating this view, we would hold that such relationships contained all that was of any importance in the identity relation. Such revised theories can satisfy analogues of Williams' requirements - i.e. ARB no matter what happens to C, and despite any other trivial events occurring or not occurring. In any cases (and these will be the overwhelming majority) that seem to suggest that personal identity is the only relation that can accommodate all that really matters regarding future survival, this will be because in each case it coincides with Relation R.

2.9 LEWIS : THE INDETERMINACY OF POPULATION

Another, more radical solution to the problem of branching is put forward by David Lewis [1983], who argues that in such cases of fission where one person at t1 has two persons at t2 with identity claims on him, this is not a case of one person 'splitting', as there always were two people there.

There is a long-standing debate between Lewis and Derek Parfit [1971 & 1976] concerning issues of survival and identity. Parfit has argued that what matters to us, in a self-interested way, concerning the future is the maintainance of my psychological continuity by any means (i.e. Relation R), and not my identity as such i.e. not that I need survive. He accepts that in practice the two relations will coincide, and that no problem cases where the two relations diverge have actually occurred. Such hypothetical cases are problematic due to the differing logical structures of the two relations, as identity is necessarily a "one : one" relation whereas Relation R is not - it can be "one : many" or "many : one", and if such cases of fission or fusion do occur, then the identity relation is inadequate to pronounce on the results.

Incidentally, in his original article [1971], Parfit seems to be discussing brain transplants (as in 'My Division') whereas Lewis is clearly discussing cases of fission arising from duplication - *"We demand to say that only one person entered the duplication center"(p63).*

In the following discussion, I discuss this latter form of fission. If, however, Lewis' arguments are adapted to apply to the former case, then my arguments against him would be those that I lay against Puccetti in ch.7.

Lewis' response is to agree with Parfit that what matters regarding the future is the holding of Relation R, but he also claims that what matters is the holding of identity through time, and that there is no contradiction involved in holding that both are what matters, as the formal discrepancy between the two relations only shows that they are different relations and not that they are mutually exclusive. Lewis rightly points out that identity and Relation R have different relata - Relation R is a relation between momentary person-stages, whereas identity applies to continuant persons who have different person-stages - it is not between person-stages as, by definition, no person-stage will survive in the future - but the question is whether the person will. However, according to Lewis, another relation can be derived from identity - one which, like Relation R, does hold between person-stages. This is what he calls the 'I-relation', which holds between all the stages of a single continuant person, and, since the two relations have the same relata, it is this relation that should be compared with Relation R in judging whether Parfit is correct in saying that Relation R, and not identity, contains all that matters regarding future survival. Lewis argues that if what matters is the holding of identity, then this corresponds to the claim that one's present person-stage is I-related to future stages. However, he goes on to claim that in all possible cases, any stage is I-related and R-related to exactly the same stages, and therefore, since they are necessarily co-extensive, then the I-relation = Relation R. thus there is no incompatibility in saying both that identity and that Relation R are what matter in survival.

We can thereby derive a non-circular definition of a person in terms of Relation R : Since a person is a maximal I-interrelated aggregate of stages (i.e. each stage is I-related to all other stages and to itself, and a person is not part of any larger I-related aggregate) then, since relations I = R, then a person can be defined as maximal R-related aggregate of person-stages.

Let us turn now to the aforementioned hypothetical problem cases.

These occur when Relation R takes a branching form, where one stage is R-related to stages that are not R-related to each other, or where such non-R-related stages are both R-related to a common future stage. these two situations are of, respectively, fission and fusion. In the following discussion I will concentrate on the case of fission. Let the single pre-fission stage be called S, and the two concurrent post-fission stages be called S1 and S2. In such a situation, S is R-related to S1 and to S2, but S1 and S2 are not R-related. Such a case seems to show that Relation R is intransitive, and since identity is transitive, then surely there will be possible cases in which we will have to choose between the two relations regarding which contains all that matters in survival.

Lewis responds to this challenge by saying that while identity is a transitive relation, the I-relation need not be, and that in a fission case, the I-relation will display relationships between stages that parallel those of Relation R, so S is I-related to S1 and to S2, but S1 and S2 are not I-related. Lewis permits this state of affairs as he conceives of cases of fission (and fusion) as involving a partial overlap between continuant persons, so, for example, if we say that S occurs at t1 and S1 and S2 at t2, then at t1, S is shared by two distinct people C1 and C2, where S1 is a stage of C1 and not of C2, and where S2 is a stage of C2 and not of C1.

So in order to sustain his identification of relations I&R, Lewis is committed to the claim that two (or, in theory, more) persons can share the same person-stage, so that fission is not a case of one person becoming or resulting in two persons, as there were two people there all along. As Lewis admits, such an overlap leads to 'overpopulation', since, in counting a population,or any subset, e.g. 'the number of people in room A at t1', we can either count the number of people who have stages at t1, or we can count the number of stages themselves. Now obviously in a world without such 'overlap' of persons, both methods of counting will always deliver the same answer, But if cases of overlap do occur there will be a discrepancy, with there being more persons than stages.

Having reached this position, and possibly to avoid its counter-intuitiveness, Lewis changes tack, arguing that we needn't do our counting by using the relation of identity, but rather by using the

relation of 'tensed identity' or 'identity-at-t', wherein x and y are identical-at-t iff both x and y exist at t, and their stages at t are identical. Tensed identity is not a relation holding between stages, but a relation between continuants which derives from a relation between stages. Yet, unlike identity per se, it is not a relation of identity among continuants, but a weaker relation that is an equivalence relation and an indiscernibility relation for the class of properties logically determined by the properties of the person's stage at t. So, in a case of fission where the split took place at t2, we can say that by counting by the relation of identity-at-t1 (i.e. before fission), there was one person present, but by counting by identity-at-t2, there were two.

Parfit [1976] challenges Lewis' parallels between the I and R relations in cases of fission, arguing that they cannot support his compatibilist position regarding identity and Relation R concerning what matters in survival. By Lewis' description of the case in terms of two persons, C1 and C2, C1's stage S is R-related to C2's stage S2, but since C1 and C2 are different persons, and thus C1 stands in the relation that matters (Relation R) to someone else, how then can the relation that matters be, or be compatable with identity? Parallel with S being R-related to S2, Lewis says that we have S being I-related to S2, but surely if we thereby have C1's stage S and C2's stage S2 as being stages of one and the same person, it is surely the wrong person, namely C2? Lewis wants to hold that both identity and Relation R satisfy conditions on 'what matters' regarding survival, so, concerning Relation R, given that S is R-related to S2, then S2 stands to C1 in the relation that matters to C1's survival - but if he wants to say in addition that identity is the relation that matters, he needs to show that both S and S2 are stages of C1, and, since S2 is not, then his argument fails.

Lewis, in a postscript to his original article, attempts to counter Parfit's objections by considering a case where at stage S1, before fission takes place, but after C1 and C2 are informed of their imminent fission, there is a desire for survival 'of the most common-sensical and unphilosophical kind possible', and, since S is a shared stage, then this desire is a shared one between C1 and C2. Now if it turns out that, shortly after the fission takes place, C1 dies and C2 continues to live for some considerable time, then C2 clearly has his pre-fission desire for survival fulfilled - but does C1? i.e. is Parfit

not correct that from C1's point of view, regarding identity, C2 is the wrong person?

Lewis replies that this would only by the case if C1's desire was that he, C1, survived, but that this was not the most straightforward desire that could have taken place, which is rather 'let us survive, or more precisely, 'let at least one of us survive', and so, on this model, C1's desire is satisfied. Lewis adds that in cases where C1 and C2 don't know about their imminent fission, then the singular desire 'let me survive' can be thought or uttered, but is unsatisfiable, as it rests on the false assumption of there being a 'me', a single person.

So, by Lewis, we seem to have a choice. Either we count persons by the relation of identity, with the result that in cases of fission there were two persons all along prior to their splitting, or on the other hand we count by 'tensed identity', which produces results that are more in line with our intuitions. Taking these two proposals in turn, I find the first option by far the less plausible, as not only does it seem to me to be an ad hoc move, but it is prone to problems that reduce the position to absurdity. Let us take a straightforward case of fission, in which we would want to say that one person at t1 results in two persons at t2. When I want to indicate their plurality, I will call them C1 and C2, and when I want to be neutral, leaving the question of how many persons 'share' or occupy the pre-fission stages, I shall simply refer to C. Now if there are two other persons present, D and E, who do not fission in the future, then Lewis must accept that D and E are 'different persons' at t1 in a vastly different way than C1 and C2 are at that same time. They have different bodies, different mental states, and all in all can be clearly distinguished and individuated at t1, whereas C1 and C2 clearly can not. When asked exactly how C1 and C2 are different persons, Lewis will reply that they are two different people sharing a set of stages - but when he is asked to explicate or to justify this claim, he can only talk in terms of their future fission - but nothing is thereby explained.

There is nothing intrinsic to C either physically or psychologically that could betray any difference at t1 between C and other singular persons, such as D or E, in such a way to reveal his plurality. In a world in which fission is a possibility, there is no way of telling at t1 or at any given time just how many people are present

within a given space. Since the number of people present isn't necessarily the same as the number of person-stages (which will be individuated by a physical criterion, i.e. counting bodies),we are left with a question that is unanswerable in principle, which should lead us to suspect that it is not a legitimate question in the first place and that the theory within which the question is grounded is incoherent.

Over and above these epistemological problems, there is the possibly deeper issue that the very matter of the number of persons present at a given time will depend on whether they undergo fission at some later time. In other words, the number of persons present at t1 will be a function of events at t2, t3,etc. Thus, Lewis seems to be committed to the possibility of backwards causation. To bring out the absurdity of this position, let us consider a couple of examples :

C is sitting in the replication machine (which is depicted more fully in ch.4) at t1. The switch that activates the creation of a duplicate is to be pressed at t2. However, the machine is a little unreliable due to bad wiring. Sometimes it works and sometimes not. By Lewis's theory, if the machine works at t2, then there were two persons sitting in the replication machine at t1, and if it is out of order at t2, then there was only one.

But what if C is sitting in a new super-replicator which can produce any number of 'copies' of C, at the rate of one per minute. It would follow that the question of the number of persons present at t1 would not admit of a determinate answer, as the number would be rising at the rate of one a minute, with, in principle, no upper limit.

Turning now to the option of counting persons by the relation of tensed identity, I have no objections to anyone using such a device, but I take a very different lesson from its employment than does Lewis, who seems to think that the fact that fission cases can be described in terms of such a relation somehow supports his main thesis that identity and Relation R are compatible answers to the question of what matters regarding survival[1]. I, on the contrary, note that 'identity at t' is a different relation to identity per se, and is not reducible to it, and argue that resorting to the usage of this new relation only confirms Parfit's argument that the logic of the identity-relation prevents it from being able to adequately describe cases of fission, whereas Relation R is clearly up to the task.

Fission cases can be described without the need for using the concept of identity. If such cases actually occurred, for example if a 'replication machine' were to be invented, then presumably there would be a satisfactory theory of how the process took place, and the terms involved would constitute the language in which such fission cases would be completely described, and any further questions regarding the identity of the pre- and post-fission persons would be empty.

CHAPTER 3 ASPECTS OF NON-REDUCTIONISM

3.1 BUTLER'S CHARGE OF CIRCULARITY

Contemporary versions of the Psychological Criterion can, of course, be traced back to John Locke [1694], and any such criterion has to face up to the charge of circularity, as originally proposed by Butler [1736]. This charge has been remarkably resilient, and I put its survival down to the fact that it isn't really clear what exactly Butler's point is, nor how to develop it in order to make it stick. I propose to consider several possible interpretations and extensions deriving from Butler, and argue that a neo-Lockean Reductionist theory can survive the attacks from all of them.

I will begin with Butler's classic statement that *"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" (p100).* In Butler's defence, the obscurity of this remark is totally derivable from the vague and ambiguous terminology of the writer he was attacking, namely Locke. However, we can make some inroads into his intentions. For 'truth', I think we can read 'the fact of my personal identity' and what it consists in. As for 'consciousness', he clearly equates this with 'knowledge', and in particular knowledge of personal identity. We can derive two interpretations regarding this knowledge and of the position that Butler is attacking. These correspond to the different interpretations of Locke's use of 'consciousness' (see ch.4).

On the first interpretation, Butler is denying that a direct present awareness of my personal ientity can be constitutive of it, since it presupposes it. Here, 'knowledge' corresponds to what Russell called 'knowledge by acquaintance'. So he can be seen as attributing to Locke the Cartesian doctrine that the self is directly revealed in the act of introspection, but then pointing out that such an act can only discover, but not be constitutive of such a self. On this first point, we need only be reminded of Hume's observation that one never observes 'oneself', but only a series of fleeting perceptions. However, Butler falls into another deep error. His argument is an instantiation of an implicit general claim that a direct perception or experience of x presupposes the prior existence of x independently of this awareness of x, and therefore cannot be constitutive of it. However, there are

counterexamples to such a rule. For instance, one cannot say that the experience of pain cannot constitute pain but can only discover it as it presupposes a pain already thereto be aware of. This is obviously wrong since there is no more to a pain than the pain-experience. To think otherwise is to reify the pain and drive a wedge between the thing itself and the experience of it. This is what Butler has done to the self (on this interpretation) and, in his defence, he is only highlighting an issue that Locke never really resolved.

Clearly Butler is committed to a view of the self as a thing, a substance, a res cogitans. But it is due to the unique nature of such a purported substance that Butler's model of 'knowledge presupposing truth' is not applicable to it. Unlike a case where, for example, the act of perceiving a car presupposes that there is a car already there to be seen, the act of perceiving a Cartesian Ego in introspection consists in the very act of the self perceiving itself - it is an irreducibly reflexive act, and thus, such an observer would not be merely discovering such a self, but would be, in so doing, manifesting its essential nature, Thus, such an act would be constitutive of the self and of personal identity, if such a substance existed (which, of course, I maintain it does not).

Also, as Wittgenstein [1958] remarks, when I am introspecting my mental states, I am not presented with 'myself' or 'my self' as an object. Such an act does not involve any identification or recognition of myself. For example, should I suddenly feel a pain in my leg, I do not verify the claim that 'I am in pain' by observing that someone is in pain, identifying the pained person as myself, and then conclude that I am in pain. Wittgenstein argues that the concept of identification can only be meaningfully applied in cases where there is the possibility of getting it wrong, i.e. misidentifying who was in pain - and there is no such possibility.

Turning to a second interpretation of Butler, we have the claim that our identity is revealed by our memories of past experiences, but that again such a discovery already presupposes that which is discovered. Again there is a logical flaw in the position that Butler is attacking on this interpretation, as ' being the same person as the person who experienced F at t1' is not a possible content of a memory. It is of a different logical type to anything that can be a possible memory, that is, a past experience. My point is parallel to Ayer's strengthening of Hume's attack on the notion of a direct acquaintence with the self,

whereby it is not just a matter of fact that when he examined the contents of his consciousness he could only observe a succession of perceptions, but it is so as a matter of logic, as such a 'self' is not a logically possible content of experience, and if this is true of the present tense, it is equally true of the past tense, in the case of memories of such experiences of the self, i.e. if it cannot be experienced, then it cannot be remembered.

Brennan [1988] follows Evans' [1982] extension of the abovementioned argument of Wittgenstein concerning the 'identification-free' nature of memory-judgements, whereby while I can be mistaken about the content of a memory, or whether it is a real memory rather than just an imaginative fantasy, I cannot be mistaken whether it is mine. Thus, the memory report 'I remember F' should not be analysed as

(i) Someone experienced F

(ii) Their having experienced F stands in the appropriate causal relation to my present memory

(iii) I experienced F.

Rather, as Evans says, *"Memory is not a way of possessing knowledge about an object of a kind which leaves open the question of the identity of that object. If a subject has, in virtue of the operation of his memory, knowledge of the past states of a subject, then that subject is himself". (p245).*

However, two points should be made here. Firstly, such an analysis leaves the whole Reductionism / Non-Reductionism dispute untouched - it leaves it open as to what it actually is to be a person. Secondly, as Brennan acknowledges, should it ever occur that someone has quasi-memories that were not 'real' memories, (see ch.3.2 below), then the former, discredited analysis would be appropriate to discuss such cases, where misidentification was a real possibility, but you were sure that your present quasi-memory was of your own past experience.

On either of my two readings of Butler, Locke is guilty of many things, but these needn't include circularity. This charge is due to Butler interpreting Locke as saying that both

(i) An experience is mine only if I remember it, and

(ii) Only if an experience is mine can I say that I remember it.

Read in this way the circularity is obvious, but Locke needn't be read as saying this. His case can be given in such a way that the identity relation can be given necessary and sufficient conditions that do not

include identity itself ;
x at t1 is the same person as y at t2 iff y remembers experiencing F at t1, and x in fact experienced F at t1.
The factor that distinguishes a genuine memory isn't the circular condition that the said memory is mine, but that there is a causal connection between the experience and the memory of it, whereby the experience of F causes a memory trace to be stored in the brain, and which can be brought to conscious awareness at a later date, as a 'memory of F'.

Returning to Butler's original statement, this can be interpreted as saying that 'I remember doing F at t1' is elliptical for saying that 'I remember my doing F at t1'. Now if he is just making a logical point about the conditions of application of our concept of memory, then this is highly debatable - but even if we concede that he has made a correct descriptive analysis of the concept of memory, we can still hold that such a concept is a ripe candidate for conceptual revision, as in its present state it cannot accommodate certain logically possible hypothetical cases involving the transfer of memories, which I discuss below.

Alternatively, if we interpret Butler as describing the subjective experience of remembering a past experience, then it is not the case that 'oneself' necessarily enters into the memory as part of the content of the remembered event. Whether or not this will happen will depend on the actual experience, and, in particular, on the extent of one's own active participatory role in the event. The extent to which 'oneself' is included in the actual content of a memory is in direct proportion to the extent of one's own physical involvement in the event. So, for example, when I remember my first school dance, I can 'see myself there' with clumsy movements, pounding heart, and clammy hands. When I remember this ordeal, I am the focus of my memory. On the other hand, we can acknowledge the power of a great performance or work of art to make the spectator 'lose himself' or 'lose the sense of himself', so that when I remember seeing Wim Wenders 'Wings Of Desire', I remember what happened on the screen, and that is all. I am not there in the memory. I only reemerge as the closing titles begin. Before that, my only contribution is that of a transparent 'point of view'.

Returning now to the claim that it is part of the conditions of use of our concept of memory that we can only remember our own experiences..... if so, we can then go on to introduce the new, wider concept of quasi-memory which isn't thus constrained and which therefore does not presuppose personal identity. According to Parfit's [1984] schema, I have a quasi memory of experience F iff

(i) I seem to remember having F

(ii) Someone had F

(iii) My apparent memory is causally dependent, in the right kind of way, on this past experience.

Note that this is virtually identical to the analysis of 'I remember F' that Evans criticises.

At this point, Non-Reductionists will argue that you cannot spell out restrictions on what will count as 'the right kind of cause' in (iii) without invoking personal identity (i.e. that F is my experience) and the analysis is therefore circular. But such a claim is mistaken. The whole force of any criterion based on memory lies in the fact that memory is a causal process. As I have noted, in the 'normal' case, an experience is encoded in the brain, and somehow stored as a 'memory-trace', which consciousness can be given access. Now, the claim that I could only quasi-remember my own experiences would depend on the assumption that this was the only causal pathway by which events could be recalled. However, if it were the case that memory traces could be transferred from one brain to another (and given our lack of a complete theory of neuroscience, we cannot discount such a possibility), so that for example it was discovered that specific memories of past experiences were located in precise sites in the brain, so that neuron-complex N, containing the memory of F, was removed from A's brain and implanted into B's brain, then it would seem to be the case that B would then be capable of having true quasi-memories of F.

Thus we have an example of an appropriate causal connection holding between experiences without the experiencer and the rememberer being the same person. Cases of fission would provide other such examples. We can also identify a causal connection between an experience and a quasi-memory prior to ascertaining the facts about identity. As Shoemaker [1984] says, if, in the case of a brain transfer, B at t2 has quasi-memories of A's experiences at t1, then the fact that he has the same brain at t2 as A had at t1 itself provides sufficient reason for saying that the quasi-memories were appropriately caused,

and this in turn is sufficient grounds for saying that A = B.

Regarding what would count as the 'right kind of cause' in his analysis of quasi-memory, Parfit later states that he would accept any sufficient cause. I agree with this, and would expand on it to say that such a cause would be acknowledged when we had some understanding of the processes involved which enabled the quasi-memories to take place. For example, in the case of a memory-trace implant, such an operation would be performed against the background of a neurologically based theory of memory, and thus of the cluster of laws underlying the process, which would ensure the regularity so that predictions could be made regarding y remembering F if x's neuron-set A (which stored this memory) was implanted in y's brain. Conversely, the occurrence of such a quasi-memory would be explained by such a theory.

No doubt, if such events ever became commonplace, our common set of beliefs regarding memory would gradually change so that it would cease to be part of our concept of memory that we could only remember our own experiences. In other words, the concept of memory would evolve into the concept of quasi-memory. It would just be that the 'natural' or 'direct' means whereby experiences were remembered could be described in terms of the neurophysiology of a single brain.

Similar points to those raised in the last few paragraphs would also apply to any replication cases, should they ever become fact. In such cases, if such a machine could record someone's brain-states, and reproduce them in another new brain, so that A's memories were 'programmed' into the brains of B and C, then these newly-constructed persons will have quasi-memories of A's experiences.

Of course, such fantasies may well be physically impossible in principle - we don't know. But one doubt what is worth raising comes from J.Z.Young [1987] *"The whole memory system may constitute one large map, organised around certain coordinates.....new features are placed appropriately on the map and retrieved by their relationship to features already there"(p171).* Given this, one might suspect that if the neurons that realised certain memories within the original brain were transplanted, the quasi-memories thus transferred would be non-operative in the new brain, as they would be in a limbo, being so alien that they would not become integrated into the 'memory-map'. If I can push Young's analogy further..... a feature on a map only exists as such if it is on that map. It loses its significance if it

is removed from that map and placed in isolation. Imagine being lost, and being handed a tiny pin-point scrap of a map and being told 'You are here'. Likewise one can suspect that memory traces would be similarly inoperative when removed from the base coordinate system within which they fulfilled their causal roles, and that they cannot be just stuck on to another entirely different causal network and be expected to be intact functionally...

However, this is an empirical matter, to be established in the event of brain-state transfers taking place. For the moment, let us be charitable and allow that memories can be thus transferred. Parfit constructs such an example, and argues that in such a case, one person can be said to have true quasi-memories of another person's experiences. A neurosurgeon has implanted copies of certain of Paul's memory-traces into Jane's brain, so that she seems to remember events that happened in Venice, yet she knows that she has never visited that city. However, she knows that Paul has been there, and Paul confirms that her quasi-memories correspond to certain of his experiences there, and so Jane can conclude that she is quasi-remembering an experience of Paul's. One can even imagine Paul 'testing' Jane in order to distinguish accurate recollection from mere imagination, i.e. 'OK, Jane, so you remember being in a gondola, and catching sight of a small red-coated figure..... what happened next?' This would be fairly easy where Jane and Paul could work in tandem. There are also cases where Jane could work out for herself whether a quasi-memory was of one of Paul's experiences or of one of her own. Parfit gives one such example, of Jane recalling shaving and seeing Paul's face directly in front of her in the mirror. This is clearly correct, although Parfit doesn't bring out its significance, which lies in its being an experience in which one is actively physically involved, and where one's perception of one's own body plays an integral role in the action, and therefore will be likewise 'featured' in the corresponding memory or quasi-memory.

On the other hand, towards what I have called the 'passive' or 'spectator' end of the scale, there will be quasi-memories that will not be so identifiable as such. but in such cases, it needn't be that such quasi-memories tell Jane what Paul's experience was like yet falsely tell her that they were of her experiences. Such a mistaken inference would not follow automatically, since she herself will be just as 'absent' from the content of the memory as is Paul. Such an

experience will be recalled as being, in Parfit's words, 'from the seer's point of view', with a conditional agnosticism regarding whoever had the original experience, tempered by the possibility of some evidence being made available at a later date that would identify the subject.

3.3 SWINBURNE'S 'SIMPLE VIEW'

Richard Swinburne is virtually unique amongst contemporary philosophers in that he attacks Reductionist theories of personal identity from the perspective of traditional Cartesian dualism. He describes his position as holding to the 'Simple View' regarding personal identity, namely that it consists in some basic fact that is not reducible to any other facts concerning either physical or psychological continuity. The Simple View holds that questions regarding personal identity always have determinate answers in all possible cases, whether or not we can ever come to discover these answers. He says that to think otherwise is to fall into the fallacy of verificationism, in failing to distinguish the question of what counts as evidence for the holding of personal identity, and the question of what personal identity consists in, and thus what it means to say that the relation holds. Thus he describes the Reductionist views (a.k.a. 'the Complex View') as empiricist theories.

Clearly an empiricist theory of personal identity will not suffice for someone who holds that all questions regarding personal identity must have determinate answers, since, as we have seen, there are many 'puzzle-cases' to which, on all the possible evidence, no such answer can be derived - so any such answer must be found beyond the realm of empirical knowledge. In arguing that all Reductionist theories commit the verificationist fallacy, Swinburne, in the original statement of his views [1974] backs up his claim by reference to other examples of this fallacy which are, contrary to his assumption, far from being analogous, as Madell [1981] points out. For example, in contrast to the problem of other minds, there is no such distinction between grounds for belief between first-person and third-person talk, as, unlike in this situation, no-one is saying that only talk of others' personal identity transcends all evidence I could have. Rather, the debate is about all cases, including one's own. Swinburne's second supposedly analogous case is our belief in the existence of the external world, and of objects. However, these constitute an explana-

tory hypothesis that best accounts for the regularities of sense-experiences, whereas the positing of some 'self' as a substance over and above the physical body and the holding of physical and psychological continuity has no such role, being unnecessary to account for the comprehensibility of experience for an embodied subject.

One way into Swinburne's 'Simple View' is to consider his adaptation of Williams' [1970] mad surgeon story as discussed in ch.1, in which he combines the original plot with a scenario similar to that depicted by Parfit [1984] in his story 'My Division', wherein I am told that my brain (or to be more precise, my cerebral cortex) is to be removed, and the hemispheres separated, with one each being transplanted into the cortexless skulls of bodies B and C. (Let us call me 'A'). After these operations are completed, one of these persons, the B or C-body-person, is to be tortured, and it is up to me, A, to decide who is to get what, prior to the operations taking place.

Swinburne says that according to the Complex View it should make no difference to me either way as to who gets tortured and who on the contrary is to be rewarded, as *"each person will be you to the extent that he has your brain and resembles you in his apparent memories and character..... both subsequent persons will be part you....(and you will).....in part suffer and in part enjoy what each suffers and enjoys". (p18, Swinburne [1984])*. Swinburne finds such an account incoherent, as *"how could you have reason for part enjoyment and part terrified anticipation, when no one future person is going to suffer a mixed fate" (p18)*.

In this final sentence, Swinburne seems to be pointing to the intransitivity that such 'branching' results in, and is tacitly saying that since B $\neq$ C, then $-[(A = B) \& (A = C)]$. Quite correct. However, in the above quotations, Swinburne makes his case at the cost of distorting his opponents' argument. A Reductionist like Parfit is not saying that under such circumstances, A is the same person as B and/or C to any extent, but rather that A <u>survives</u> as the B-body person and as the C-body person, as he is psychologically continuous with both of them. Thus it seems reasonable to say that, when faced with the thought of his future fission, A's mixed feelings are well justified, not because <u>he</u> or any <u>one</u> future person will experience such mixed fortunes, but because he stands in Relation R to two different future persons, whose combined experience will encompass

both the torture and the reward. This can be expressed another way : in fission, stages S1 and S2 are R-related to a common earlier stage S, but are not R-related to each other.

Swinburne goes on to say that we can make sense of the idea of A making the 'wrong choice', in other words, of finding out that he is being tortured, or conversely of making the right choice, whereas by Parfit's theory there is no real difference between the choices from A's perspective, and therefore no risk. Now it seems to me that we can only make sense of A making a right choice or a wrong choice in this way on the assumption that personal identity involves something beyond the physical and psychological continuity that relates A & B and A & C, and thus such apriori intuitions cannot without circularity be put forward as grounds in favour of such a theory. If, on the other hand, there is no such 'further fact' in which identity consists , then we cannot make sense of there being a risk involved in the choice, as there is nothing left to be at stake.

Swinburne's response to A's dilemma is to say that there is equal evidence for the claims that A = B and that A = C, but that all such evidence is, in principle, fallible, and the fact whether B or C is the same person as A is independent of such evidence. But, as I've said, Parfit's line is somewhat different from that depicted by Swinburne - it is rather to say that certainly we cannot decide on all the available evidence whether A = B or A = C, and obviously A cannot be identical to both, so the problem is insoluble. However, the situation can be re-described in terms of Relation R, and thus we can say that A survives equally in or as B and C.

Swinburne remarks that the weakness of Reductionist views in dealing with fission cases is in assuming that *"mere logic could determine which of the experiences had by various persons, each of which was to some extent continuous with me in apparent memory and brain matter, would be mine" (p20)*, whereas in his 'Simple View', their being mine was a further fact, over and above any evidential considerations.

However, yet again he distorts his opponents' position, as firstly the complex view, as I've said, needn't be cast in terms of which future experiences will be mine, and secondly, as Shoemaker [1985] says, the

Complex view doesn't rely on 'mere logic', but states that in the cases in which answers to questions regarding personal identity can be given (which isn't always), then these answers will be ascertained on the basis of empirical facts regarding the causal relations between experiences, and, in such cases, that is all that is to be said on the matter.

3.4 NON-REDUCTIONISM AND DUALISM

I will now turn to the matter of Swinburne's dualism and its relation with the Simple view. Looking at the components of the Simple View, it is clear that dualism cannot be derived from the indefinability or irreducibility of person identity, since, as Shoemaker [1984] says, when terms refer to structured wholes, i.e. trees, cars, one can hold that these terms are not reducible to , or definable in terms of the sum of their parts, but no-one would therefore want to hold a dualistic account of these.

Secondly, regarding the determinateness of questions of personal identity, Swinburne is mistaken to hold that any opposition to his theory must be based on verificationism since, as Parfit notes, one can deny that all such questions regarding the identity of nations, clubs, etc, must yield a determinate answer, and this denial isn't based on verificationist grounds. Thus further argument is required to show why the situation is in any way different in the case of persons. Rather, the denial that all such identity-questions must admit of a determinate answer is due to the vagueness of the sortal term involved, which allows an indeterminacy in the reference of all items falling under such a term, which in turn results in the indeterminacy of truth-values of statements involving such referring terms.

Turning to another set of arguments, Swinburne seems to think that the truth of a dualist theory follows from the fact that he can, coherently and without contradiction, imagine that he could continue to exist in a disembodied state, being *"able to operate on, and learn about the world.....without having to use a particular chunk of matter for this purpose...(where)...simply by chosing to do so, he can gradually shift the focus of his knowledge and control"(p24)*. He accepts that such a disembodied existence would be impossible for a purely physical being, so, given that he's shown that it is logically possible, this must be due to the truth of what he calls his 'wider Aristotelian principle' involving non-material stuff organised into a

certain form.

To begin with, I would dispute Swinburne's claim that the idea of disembodied existence is coherent. Imagination is a very deceptive faculty, rendering plausibility where it is not warranted. Certainly, the way Swinburne describes it, there is no blatant inconsistency, but this is only because he offers such a fragmentary sketch. If he were to attempt to fill out the description to include a full theoretical account of how disembodied existence was possible, then it would be a different story, as he would then be required to account for how it is possible that one could perceive the world without sensory organs, or act upon the world without any muscles - and, it seems to me, there is, in principle, no way in which such questions can be answered. Also, I'm not convinced that one can imagine being disembodied, in the sense of being able to grasp 'what it's like' from a subjective viewpoint, without tacitly introducing some elements of the physicality that is being denied. I can, at a stretch, imagine being invisible, or feeling numb, or weightlessness, but that's as far as it goes.

On a more serious note, Swinburne's argument is not logically sound. As Shoemaker [1984] argues, Swinburne's claims that one might survive disembodied do not constitute a de re claim that, regarding one particular person, it is possible that he might survive in a disembodied state - at most it says that it is possible that there be some person that could do so. Yet Swinburne goes from this lesser claim into the realm of the actual, saying that *"for any present person who is presently conscious, there is no logical impossibility ..... that that person continue to exist without his body" (p29)*, which does not follow.

Neither does it follow from my being able to imagine myself as disembodied that such a state is a de re possibility for me. Arnauld had long ago dismissed such an argument from Descartes, using a parallel case whereby I might doubt that a right-angled triangle has the property of having the square of its hypoteneuse equal to the sum of the squares of the other two sides - but it doesn't follow that such a triangle does lack this property, nor that it is an essential property of any such triangle. Likewise, it may be, despite my imagining in my ignorance, that I could exist without a body, that I am an essentially embodied being.

Next, Swinburne claims that 'laws of nature' do not necessitate that a certain person have certain memories or that his body be made of certain matter. He imagines a time c. 4000 million years ago, when the earth was a cooling globe of inanimate chaos, and argues that while natural laws would dictate the general evolution of the planet and the life-forms upon it, including the ways in which bodies would be formed that could support conscious life, *"what natural laws in no way determine is which animate body is yours and which is mine.... just the same arrangement of matter and just the same laws could have given me the body (and so the apparent memories) which are yours, and you the body (and so the apparent memories) which are now mine" (p25-6).*

However, yet again, this argument presupposes what it sets out to prove - that is, it assumes that there is a 'me' and a 'you' prior to the existence of the bodies we now 'occupy', so that somehow I am 'alloted' one and you another. In other words, Swinburne again assumes the dualist theory, and the existence of immaterial souls prior to the existence of physical bodies.

3.5 THE SUBJECTIVE VIEW

I will finally turn to another of Swinburne's arguments for the Simple View, wherein he claims that the ultimate irreducible nature of personal identity is proved from the fact that it is revealed in experience, and that no other theory can account for the copersonal relation between experiences. He claims that *"the continued existence of a person over a very short period of time is something that can often be experienced by that person..... without it depending on our knowledge of anything more ultimate.... the continuing of a person is a datum of experience" (p41).*

When we are tracking a moving object, it is not a complete account of our experience to say that we see it at place p1, then at p2, p3, etc., as such perceptions are not discrete units, but rather they overlap into one continuous awareness through time. He quotes Foster [1979] *"it is this double overlap which provides the sensible continuity of sense-experiences and unifies presentations into a stream of awareness"* - in other words, we are aware of all these experiences as mine - *"it is in the unity of the stream that we primarily discern the unity of the subject" (p176).*

Likewise, claims Swinburne, also among the data of our experience is the experience of oneself as being the common subject of simultaneous experiences, which are perceived as being 'mine' in a direct, non-inferred way, as this knowledge cannot be derived from any information regarding the experiences themselves, or from knowledge regarding their associated brain-states.

On this set of arguments from Swinburne, I am in agreement with Shoemaker that one can go along with these observations concerning 'first-person knowledge', but that such admissions leave the whole debate between the Simple and Complex views on personal identity untouched. It is certainly the case that such instances of self-knowledge have the property of being immune from error due to mis-identification, and it is a form of knowledge that is immediate and direct, and not derivable from any other information - however, we can hold that it is just a fact about the way in which human beings are constructed that being in a certain state, or having a certain experience directly causes the incorrigible belief that I am in such a state or am having such an experience, but it may still be the case that personal identity can be given a non-circular and substantial definition in terms of other relations. Also, as Shoemaker observes, Swinburne has yet to offer any argument to suggest that a dualist theory would provide any insights into the peculiar properties of self-knowledge, for merely saying that it consists in an unanalysable fact has no explanatory power at all.

A parallel point can be made regarding Strawson's [1959] argument that 'person' is a primitive concept, prior to such as 'individual consciousness' or 'Cartesian Ego'. Again, this leaves open the whole question of the kind of thing a person is, and of what are the conditions of identity for persons. The ability to correctly apply the concept of a person involves the ability to identify and individuate persons, which demands that we can recognise others as persons and attribute both physical and psychological attributes to them. Yet, even if Strawson is correct on this, the debate between the Simple and Complex views remains untouched, as it is still open to question just what persons actually are, and what can be said about them. As Brennan [1988] says, it doesn't follow from the fact that a certain recognitional capacity is primitive and that certain items are basic within this recognitional schema, that their nature is primitive and unanalysable, and that nothing more can be said about the conditions of identity through time

of items falling under such a concept. To make a comparison, 'flower' is a basic concept, relative to seed, stem, blossom, etc., - that is, the concept of a flower enables us to identify and individuate such a structured assembly of parts, yet what it is to be a flower is no more than the structured sum of these parts organised into a whole. Again, conceptual primitiveness does not imply ontological primitiveness.

Returning to this matter of our direct awareness of our identity through time, and all our experiences recognised as 'mine', Shoemaker points out that this is only the case if we assume that branching (i.e. fission) has not taken place. Even if we are in a state of reflective awareness of having experiences over a short period of time, where we are aware of having experience 1 followed immediately by experience 2 so that they overlap, then, if fission occured during this temporally-extended awareness, then the state of awareness itself branches, so that its earlier stages are not copersonal with its later stages.

The final nail in the coffin of Swinburne's dualistic theory of personal identity comes when we recall that there are classic objections to substance dualism that have never been answered - namely the problem of giving a theoretical account of how two mutually distinct substances could causally interact, and how a criterion of identity for minds can be devised that does not depend on a bodily criterion. Swinburne nowhere even acknowledges that these problems exist, let alone tries to solve them.

3.6 EMPIRICAL GROUNDS FOR NON-REDUCTIONISM?

This section is based on Parfit's [1984] argument entitled 'How a Non-Reductionist view might have been true'. He had already shown to his satisfaction that to hold a Non-Reductionist theory of personal identity basically amounted to the claim that identity centred on the continuing existence of a Cartesian Ego - i.e. one is committed to a substance dualism. However, rather than accepting the contemporary wisdom that the Cartesian Ego is an inherently incoherent concept, subject to insurmountable difficulties, Parfit surprisingly makes the concession that it is just a (contingent) fact of life that no such immaterial entities exist. He then goes on to suggest the kind of possible evidence that would be sufficient to justify the belief that such units of non-material substance existed as separate entities.

The example he offers is one designed to provide evidence for reincarnation. This is slightly peculiar, given his later characterisation of Buddhism (which also adheres to a belief in reincarnation) as a form of Reductionism, and therefore as advocating a quasi-Humean position, denying the existence of a self beyond this 'bundle of perceptions'. Incidentally, it has always been a mystery what exactly is reincarnated according to Buddhist theory, if there is no quasi-Cartesian Ego existing separately from our physical and psychological continuity. To suggest that the consequences of the Law of Karma continue, in the form of habits, desires, etc., seems to me to be committing the same logical fallacy as that ridiculed by Lewis Carroll in his tale of the grin surviving the Cheshire cat. The point being, of course, that propositional attitudes are not free-floating, self-subsistent entities, but functions <u>of</u> entities, i.e. conscious subjects of experience. But to return to Parfit, and his claim that evidence for reincarnation would provide evidence for the continuous existence of a Cartesian Ego, and hence for Non-Reductionism.

He constructs an imaginary case of a modern-day Japanese woman who experiences what seem to be quasi-memories of having lived many thousands of years ago as a Celtic warrior. From information acquired from these memories, she goes on to make certain predictions, for example concerning certain as-yet-unknown archeological or historical discoveries, which, following her instructions, are confirmed. Parfit says that if

1. There were many actual well-confirmed cases such as this; and
2. There were no <u>other</u> ways of accounting for such detailed and specific knowledge of such events of the distant past;

then one would have to concede that people can have true quasi-memories (or, simply, memories) of past lives, and, in the absence of physical continuity between the subject of the remembered actions and the woman herself, then we would be obliged to posit the existence of a purely mental entity that extends through time to cover both physical embodiments, continuing to exist throughout the thousands of years separating the Celtic and Japanese-bodied persons. And, says Parfit, if such an entity exists, then it is plausible to conclude that such an entity constitutes what this person essentially <u>is</u>. The only problem is, he claims, that we have no such evidence for the existence of such entities.

I will argue that there <u>does</u> exist what seems to be an abundance

of evidence similar in kind to that which Parfit would find acceptable, but that this does not commit us either to Cartesian dualism or to Non-Reductionism. Let us examine Parfit's two abovementioned conditions, upon which he thinks the dualist case depends. Firstly, contrary to Parfit's unsubstantiated assertion, there are countless well-documented cases, throughout virtually all societies, that could be used in the place of Parfit's hypothetical example. If such cases were fully described they would fill many volumes and it is neither necessary nor desirable that I do so here. A few notable examples will suffice. If the reader remains sceptical of the veracity of such sources, then he or she should just consider them to be on the same level as Parfit's hypothetical case, and consider my investigation as being into what we would say *if* such cases were true.

3.7 CASE HISTORIES : THREE STRANGE TALES

1. Consider the strange case of Lurancy Vennum, as described by Watson [1974]. At the age of 13 she fell into a cataleptic state which lasted for five years, which was followed by trances in which she spoke of 'angels and spirits'. The diagnosis at the time was that she was 'possessed' by two evil alien personalities. However, it was discovered that her own personality could be restored by means of hypnosis and, whilst in this state, she 'herself' said that the only way that these spirits could be kept at bay was for her to allow herself to become possessed by an 'angel' who wished to assist her. This angel was named as Mary Roff. It emerged that someone of that name had died, aged 18, a year after Lurancy was born. She had lived in the same town, although the two families had had no contact nor knowledge of each other.

After the particular hypnotic trance in which Mary Roff was discussed, Lurancy awoke with what appeared to be full knowledge 'from the inside' of Mary's life, down to the tiniest insignificant detail, remembering events, recognising people, etc. She even went to live with the Roff family, who treated her as their dead daughter, just as she herself regarded herself as Mary. This 'possession' lasted for over three months when suddenly, for no apparent reason, Lurency's original personality returned - she recaptured all her old memories which had been lost while living 'as' Mary, and her old personality was restored. She recognised her family again, as they had been strangers to her as Mary, and she returned home to live with them. She retained no memory

of her 'lost 100 days', and no such similar phenomena happened to her again.

This case shares many similarities with cases of Multiple Personality Disorder and psychogenic fugue, which I discuss at length in later chapters, However there is one crucial difference, in that in the above case, the newly manifested personality is neither newly created, as in the second personality of Mary Reynolds that I call MR2; nor in the role of an intraconscious personality observing her 'host', with access to all her thoughts and feelings, as with Eve Black in relation to Eve White (see ch.8). It rather appears that Lurancy, when acting 'as Mary', possessed knowledge that cannot be accounted for by any orthodox means, in terms of her own history. This is the crucial factor, once we reject Lurancy's own mediaeval spiritualistic interpretation of her experience in terms of spirits, angels and suchlike. This feature of unaccountably acquired knowledge is even more strikingly displayed in the following case.

2. This is the case of Imad Elawar, as described by Stevenson [1966]. As a very young child, Imad appeared to know many things that he hadn't been taught. He spoke of friends that his family didn't know, and which they dismissed as a child's fantasies, until one day he embraced a stranger in their Lebanese village, claiming that this man had once been his neighbour. The man replied that he had rarely been in this village before, as he lived in another village fifteen miles away, and thare was little contact between the villages.

Stevenson arrived to investigate the case, visiting this second village of Khriby with the 5 year old Imad. After several false starts, he discovered discovered strong qualitative similarities between Imad's quasi-memories and events in the life of one Ibrahim Bouhainzy, who had died there some years before. On examining Ibrahim's home, Stevenson noticed many small features and items that Imad had previously described in detail. Many of Imad's quasi-memories were confirmed [illegible] correspond to events in Ibrahim's life. Stevenson was satisfied that there was no fraud on the part of Imad's parents, as not only had they not known Ibrahim or anyone else who knew him, but they had taken no part in Stevenson's investigations and, in fact, had wrongly concluded, in disagreement with him, that Imad's quasi-memories were of the life of some <u>other</u> man. Taking all this into account, Stevenson argued that Imad's quasi-memories could not be explained away in terms of chance, deception, or ordinary memory.

3. Finally, consider the case of 'Rosemary' (in Watson [1974]) who, in Blackpool in 1931, fell into a trance and began utterring a strange language, unknown and unintelligible to her family, and which was eventually identified as an ancient Egyptian dialect. In this language, she voiced the claim to be Telika-Ventin, the Babylonian wife of Pharoah Amenhotep III, of the 14th century B.C., and said that she could communicate through Rosemary because Rosemary herself had been a handmaiden of hers in a previous life.

A friend of the family, Dr. Frederic Wood, copied down a number of these uttered phrases phonetically, and sent them to a leading Egyptologist, Howard Hulme. It is acknowledged that the only written records of the ancient language, in heiroglyphic form, represent only consonant letters, so no present-day expert knew exactly how ancient Egyptian actually sounded, as the vowel sounds could only be guessed at, by comparison with other, surviving dialects. WHen the vowel sounds were omitted from Rosemary's words, what remained was intelligible and informative to Hulme, who described the purely technical features of the speech that were so convincing, including *"period characteristics, survival of archaisms, grammatical accuracy, peculiar popular terms, ordinary elisions and figures of speech" (p199).*

3.8 A MATTER OF INTERPRETATION

One must tread a very fine line when dealing with such purported evidence. On the one hand, the history of parapsychology is by and large the history of charlatanism gradually exposed, and this encourages the suspicion that all such cases are fraudulent. When presented with such strong cases as I have described, the sceptic always has the option of regarding it as a conspiracy of deception on behalf of all the participants. On the other hand, such scepticism must have limits - they must at least give an indication of what would be acceptable to them as proof of the phenomenon under consideration. If, for example, any independent investigator, such as Stevenson or Hulme, is just assumed to be part of the conspiracy, or at least an unwilling dupe, then such sceptics cease to make empirical claims at all. If nothing can possibly count as proof for something, then nothing counts as disproof either. Also, as the odious neo-Nazi 'revisionist' historians show, it is possible to doubt the occurrence of any event, even that of genocide, no matter how incontrovertible the evidence seems - but this says more about the psychology of the doubters than about the

veracity of the events themselves.

Henry Sidgwick (in Wheatley & Edge [1976]) displayed a worthy attitude when he said, on behalf of the Society for Psychic Research, *"We believed unreservedly in the methods of modern science.... but we were not prepared to bow with equal docility to the mere prejudices of scientific men.... It appeared to us that there was an important body of evidence - tending prima facie to establish the independence of the soul or spirit - which modern science had simply left on one side.... evidence tending to throw light on the question of the action of the mind either apart from the body or otherwise through known bodily organs" (p77).*

C.D.Broad [1949] located paranormal events against a background of 'Basic Limiting Principles', a set of assumptions that underlie both practical activities and opinions and scientific theories, and which are either 'self-evident' or are *"overwhelmingly supported by all the empirical facts which fall within the range of ordinary experience and the scientific elaborations of it" (p10).* (He includes psychology within this latter category). Psychical research, to Broad, concerns 'ostensibly paranormal events' which, prima facie, conflict with one or more such principles. His Basic Limiting Principles include :

1) General Principles of Causation :
1.1 An event cannot have effects before it has happened.
1.2 An event 'a' at t1 cannot cause an event 'b' at t2 except if 'a' initiates a process of change which continues through the intervening period and contributes to the initiation of 'b', or if a initiates a structural modification persisting throughout, which cooperates with some change taking place, together causing 'b'.
1.3 An event cannot produce an effect at a remote place unless a finite period elapses between the two events, and that period is occupied by a causal chain of events with spatio-temporal continuity.

2) Limitations on the Action of Mind on Matter :- a mental event cannot directly produce any change in the material world except certain changes in its' own brain.

3) Dependence of Mind on Brain :- a brain state is a necessary condition of a mental state.

4) Limitations on ways of acquiring knowledge :-
4.1 One cannot perceive a material object or event except by means of sensations which it produces in the mind, via the appropriate causal chains.

4.2 It is impossible for A to know what experiences B has except
(i) via B's spoken or written accounts
(ii) via interpreting B's cries, movements, etc.
(iii) via seeing and making inference from material records and evidence
4.3 One cannot forecast a future event except
(i) by inference from present data from sensations, thoughts, memories
(ii) by inference via others' trusted evidence
(iii) by non-inferential expectations based on past associations, stimulated by a past experience.
4.4 One cannot know that a certain past event occurred, except where
(i) it is a direct memory
(ii) via another's memory
(iii) via material records and accounts
(iv) via inference from material evidence.

Both Parfit's hypothetical case and my purportedly real examples are in conflict with (at least) Principles 2, 3, 4.1, 4.2, 4.4. However, While Broad's schema has some merit as a purely descriptive analysis of the relation of paranormal events to our entrenched commonsense assumptions about the world and ourselves, nothing follows from this analysis as to how we should regard such phenomena - nothing forces us to deny their possible or actual occurrence just because of an incompatibility with this set of assumptions, as it may be that it is these principles themselves that may be in need of revision in the light of such events. After all, what is the status of these Basic Limiting Principles? Despite Broad's claim that they are 'assumed' to be true by scientific theories, they are not <u>implied</u> by any fundamental scientific theory, as is often taken for granted as when it is claimed that paranormal events are incompatible with accepted scientific theories. As Braude [1988] says, *"No well-supported global scientific theory (e.g. Quantum Theory, General or Special Theory of Relativity, Theory of Evolution) precludes the existence of <u>any</u> cognitive or intentional phenomena, normal or paranormal" (p278).* The position is similar with theories of perception, which merely *"<u>describe</u> the operation of familiar or known sense modalities.... it is not their business to legislate the full range of possible forms of information acquisition or organic interaction" (p278).* In fact Braude suggests that those who hold that paranormal phenomena are in any way special concerning an incompatibility with physical science *"have simply failed to appreciate the insurmountable obstacles to explaining <u>normal</u> intentional phenomena in physicalist terms" (p293).* While I would

quibble with his claim that the obstacles are, in the latter case, insurmountable, (who can foretell the future and how our theories will develop), I think Braude's point is correct.

Let us now return to Parfit's Japanese woman, and to his second condition regarding grounds for belief in reincarnation, and thereby in Non-Reductionism - that there is no other explanation for the phenomenon. Since I am prepared to accept the reality of my own examples, and since I reject Non-Reductionism, is the *onus not on me* to suggest an alternative explanation?

Ducasse [1961] observes that ESP is the classical counter-explanation of paranormal phenomena, where information is derived
a) telepathically from living people who know these facts,
b) by retrospective clairvoyant observation of the events, or
c) by clairvoyant observation of existing records or circumstantial evidence of past events.
However, as Williams [1957] has argued, in the absence of bodily continuity, we can have no corroborating evidence that could decide between the hypothesis of ESP and that of an identity claim, via a purely immaterial substance, and manifesting through memory. For the same reason, he argues, we cannot draw a distinction between identity and exact similarity in the case of token memories of character traits in the way that we can for material objects, i.e. bodies, where the issue is settled by the criterion of spatio-temporal continuity.

Adherents of ESP regard it as the most economical hypothesis, as it doesn't multiply entities unnecessarily, but relies on certain purported capacities of the mind of the 'medium'. However, such an explanation cannot cover <u>all</u> cases of paranormal cognition without attributing a virtually unlimited range of powers to the person concerned, and this is an ad hoc manoeuvre, as we have no other, independent reason for ascribing such powers. In fact, this ESP hypothesis doesn't qualify as an explanation at all, as there isn't the slightest suggestion of <u>how</u> the mind acquires this knowledge. In the absence of any acceptable theory, the claim that the mind acquires the information telepathically or through clairvoyance is akin to the fallacy of saying that opium puts you to sleep because of its 'dormitive virtue'.

I seem to be left with the option of classing paranormal events

as 'anomalies' in Kuhn's [1962] sense. While these anomalies have certainly been accumulating over the centuries, they are no great threat to any scientific paradigm, since one thing conspicuous by its absence has been any sign of a new paradigm or a developed body of theory that makes sense of these events. Any attempts that I am aware of are stuck at the level of unfalsifiable pre-scientific metaphysical dogma and woolly speculation.

As H.H.Price [1949] says, *"The theoretical side of psychical research has lagged far behind the evidential side, and that, I believe, is one of the main reasons why the evidence itself is still ignored by so many... highly educated people. It is because these queer facts apparently 'make no sense'... that they tend to make no permanent impression on the mind.... If we could devise some theoretical explanation, in terms of which the facts did make sense, it would be a great gain. Such an explanation is needed for its own sake, and it is also needed to get the evidence attended to and considered" (p20).*

So this is the gist of my response to Parfit's second condition - I need not find any alternative explanation, as nothing worth the name of 'explanation' has been offered in the first place. Like dualists generally, students of psychical research can offer no significant body of theory concerning the nature of this 'res cogitans' or of what psychophysical laws might maintain its interaction with a body. As Paul Churchland [1988] says, *"There is no settled core of theory whose past successes have unified the community behind it, whose current form has been shaped in response to past experimental failures, and whose experimental agenda drives the assembled discipline forward" (p233).*

So I admit that these paranormal events are mysterious, and not explained by science, but that they are equally mysterious from a dualistic perspective, which has the additional disadvantage of lacking even the beginning of any theoretical resources within which to construct an explanation.

CHAPTER 4 IDENTITY, SURVIVAL, AND SORTAL CONCEPTS

4.1 INTRODUCTION

David Wiggins [1980] claims that in all cases of material objects that fall under the extension of a natural kind term, where 'a' and 'b' are temporal stages of such an object, any identity statement 'a = b' that says that they are both stages of the same temporally enduring object entails firstly that there is a principle whereby 'a' and 'b' can be located on a single unbroken spatio-temporal track, and can thereby be identified as being stages of the same thing, and secondly that 'a' and 'b' can be identified under a sortal concept of the type that affords the best answer to the question as to what sort of thing the object is. I will cast doubts on both of these conditions in due course, but firstly I will continue to develop Wiggins' theory. To quote :

"a = b if and only if there exists a sortal concept f, such that

1. a and b belong to a kind which is the extension of f

2. to say that 'a is an f' is to say what a is (in the sense Aristotle isolated)

3. a is the same f as b ($a \overset{f}{=} b$)" (p48).

This second requirement refers to a special class of sortal concept, one that Wiggins calls a 'substance-concept', which can apply in a present-tense form to a given object throughout its entire life-span, and which specifies its 'form of life' and its 'principle of activity'. He distinguishes these substance-concepts from 'phased sortal concepts', which apply in a present-tense form only to specific and limited portions of the individual's life-span. So, for example, in the case of a human being, such terms as 'infant', 'teenager',etc., would be among the class of phased sortal concepts available.

It should be noted that whilst it is the case that 'a = b' entails that '$a \overset{f}{=} b$', we needn't be able to describe, or have any knowledge of the specifics of f, rather it is that we are merely committed to saying that 'a' and 'b' are the same something - that is, the natural kind term that is the appropriate substance-concept. The nature of f is a matter for empirical investigation, and 'f' will be a term within the taxonomy of an ideal scientific theory.

So what is the appropriate substance-concept that is implicitly invoked in issues regarding identity-through-time for creatures like myself? One would have thought that the answer would be unproblematic, but, looking back over the history of philosophy, a great deal of confusion and complexity has developed, which needs to be understood and disposed of. In order to do this, I will make a detailed study of the first serious attempt to deal with these issues, and one whose influence extends to the present day, namely that of John Locke [1694]. Once the errors in his analysis are out in the open, the way ahead will be clearer.

Locke was the first to recognise that there was a real issue here, and one that required to be resolved by philosophical analysis. Previously, the likes of Descartes would have regarded the matter as self-evident, in the knowledge that man was essentially a thinking being, a 'res cogitans', and therefore any naturalistic, biologically based sortal concept was out of the question regarding the issue of saying what such a being essentially is. Locke himself recognised the conflicts that were inherent in such a dualistic conception, but his own errors are equally rooted in dualistic assumptions.

Locke made a distinction between the issue of $\grave{a}(t1)$ being the same man as $\acute{b}(t2)$ and the issue of $\grave{a}(t1)$ being the same person as $\acute{b}(t2)$. He regarded the concept of 'person' as standing for *"a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing, in different times and places" (II xxvii 9)*. Such a definition makes an implicit reference to his criterion of personal identity, which consisted in the possession of an uninterrupted flow of self-conscious awareness : *as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person" (II xxvii 9)*.

However, as I have mentioned, Locke regarded beings like myself as falling under a second, quite different sortal concept, namely 'man'. Whilst the concept of 'person' is essentially concerned with a subject as a bearer of a rational and reflective consciousness, this latter concept of 'man' is what we would now call a biological classification, and as such focusses on the beings in question as living physical systems. It follows that the questions of whether $\grave{a}(t1)$ is the same man as $\acute{b}(t2)$ and of whether $\acute{a}(t2)$ is the same person as $\grave{b}(t2)$

must be distinguished as their answers are determined by reference to different criteria, namely, in the first case, whether there is a spatio-temporal track upon which both a and b can be located and thereby identified; and in the second case, whether b has memories of a's experiences at t1 and at all other times leading to t2.

Locke believed that it was possible in principle for the two criteria to diverge and thereby to yield different answers to questions regarding the identification of a and b. As he states in one of his most (in)famous passages, in the context of devising a thought-experiment in which two entire sets of memories were somehow exchanged between one body and another, *"should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, he would be the same person with the prince" (II xxvii 15)*. Yet, Locke maintains, such an embodied being would remain the same man as the cobbler.

So why did Locke make this distinction between 'man' and 'person', and why did he attach so much importance to it? This distinction of sortals, and the criteria of identity that underlies them, prefigures the division that is still debated in contemporary investigations, between the Physical and Psychological Criteria, with these now, of course, not implying substance dualism, but pertaining to different levels of functioning. Contrary to appearance, and in particular to the previous quotation, it would be wrong to wholeheartedly attribute to Locke the Cartesian doctrine that persons are essentially incorporeal souls, who are temporally and contingently embodied in an edifice of flesh and blood - but certainly some passages can be produced to suggest the contrary. Perhaps the safest thing to say is just to accept that here, as in so many cases, Locke does neither himself nor the reader any favours by his sloppy and ambiguous use of key concepts such as 'soul' or 'consciousness', and thus that it is a futile enterprise to try and maintain any one consistent interpretation. Perhaps it is the prerogative of a pioneer to paint in broad brush strokes, and leave the finer details to those who follow in their wake. Three main usages of 'consciousness' need to be distinguished :

(i) our normal waking state;
(ii) a state of reflective self-awareness; and, most importantly;
(iii) memory.

As for the concept of 'soul', sometimes, as in the last quotation, he uses it in a quasi-Cartesian sense, and elsewhere in a way that is vaguely synonymous with 'consciousness' in some rough amalgam of the

usages that I have just distinguished, without the implication of substance-dualism.

However, Locke also denies that there is any possibility of our attaining any knowledge of such a soul-substance : *"it being, in respect of our notions, not much more remote from our comprehension to conceive that God can.....superadd to matter a faculty of thinking.... than he should superadd to it another substance with the faculty of thinking" (IV iii 6)*, and he goes on to explicitly deny that such a soul is the essential bearer of personal identity. Due perhaps to a rational concern for his personal survival, he didn't risk provoking the wrath of the church and state authorities by offering an out-and-out denial of the existence of individual souls, restricting his argument to the claim that the existence, and the issue of the identity conditions regarding these souls are irrelevant to the issue of personal identity.

Continuing with his strategy of thought-experimentation, he argued that even if such souls do exist, (composed of some non-extended immaterial substance,) it is possible that one such soul could trans-migrate from one body to another, for example in reincarnation. So in such a hypothetical case, the present mayor of Queensborough may now have what was once the soul of Socrates, but if he has no consciousness (i.e. no memory) of being Socrates and of having his experiences, then by Locke's criterion, he is not the same person as Socrates. Here, Locke is moving away from a purely Cartesian position equating the immortal soul with the res cogitans, which was the bearer of memories, whether they be those of this present life, or of previous lives. Such a move was necessary if he was to maintain that memory was <u>constitutive</u> of personal identity. A Cartesian, on the contrary, would say that memory could only discover this identity.

Locke goes on to argue that if the same soul, in the same body, were to carry two separate and distinct alternating sets of thoughts and memories, there would be two persons in one soul : *"could we suppose two distinct incommunicable consciousnesses acting on the same body, one constantly day by day, the other by night.... I ask, whether the day and night man would not be two and distinct persons as Socrates and Plato" (II xxvii 23)*. Here, Locke dramatically anticipates modern discussions of Multiple Personality, and the challenge that such well-substantiated phenomena pose for our everyday beliefs regarding the unity of mind and personhood. Locke's main point appears to be that he regards the holding of memory <u>itself</u>, and not the possession of a

soul, as the necessary and sufficient condition of identity through time.

If this is Locke's theory, then insuperable problems result from this reluctance to regard the soul as the essential bearer of memory and consciousness. The point is that something must be the bearer of these functions - they are not free-floating, self-sufficient entities, any more than was the grin on the face of Lewis Carroll's Cheshire cat. So Locke is trapped on the horns of a dilemma - either he falls back and admits that the soul is the essential bearer of memories, in which case he is open to the classical, and I believe insurmountable criticisms of substance-dualism, or alternately, he attributes it to the brain. However, such a tactic would destroy his distinction between 'man' and 'person' (at least as prospective substance-concepts) and subsequently of any divergence in their answers to identity-questions. His 'prince and cobbler' story would be impossible, as the mental properties (associated with personhood) would be embodied by the neuronal properties, attributed to the man. Of course, the door is still open for related thought-experiments such as that from Williams [1970], as discussed in ch.2.

So I conclude that Locke's formulation of the man/person distinction is incoherent. The fundamental flaw in Locke's approach is that in focussing on memory and consciousness alone, without reference to any underlying substance or means of embodiment, he therefore places these functions in a vacuum, in a limbo, untouchable by any nomological factors which are essential in determining what is and isn't a possible state of affairs for them to enter into. It follows that Locke cannot locate memory within any theory, and hasn't a hope of explaining anything pertaining to it. I take up this task in ch.6.

As to why Locke placed so much importance on the man/person distinction, I think we have to return to 'consciousness' in its wider sense, as incorporating all the higher mental faculties, as involving the exercise of rationality, intentions, choice, etc.. Locke thought that 'man', as a biological catagory, could not possibly accommodate or account for such matters, nor, therefore, such areas of crucial importance for him as ethics and moral responsibility. However, like Descartes, he adhered to a mechanistic model of biological systems, and much of what must have been at his time an intuitively powerful

position regarding the impossibility of such systems possessing the ability to display reason has been robbed of its plausibility since the arrival of computers that can manipulate symbols in terms of logical relations.

4.3 NATURAL KINDS AND NATURAL LAWS

Let me now return to my question raised at the beginning of this chapter, namely what is the most appropriate sortal concept - the substance-concept - to use in application to beings such as myself? In common non-philosophical usage we have a number of candidates, such as 'man', 'person', 'human being', etc.. In such everyday contexts these terms are used fairly interchangeably, but in technical contexts we require more precision. My tactic will be to examine these concepts and the relationships between them, to shed some light on my question.

I regard 'human being' and 'homo sapiens' as bearing the same relation to each other as do 'gold' and 'the element with the atomic number 79'. That is, 'human being' picks out a class of individuals in response to outer visible characteristics, but also makes implicit reference to a more precise principle of classification, and it is this latter classification that the term 'homo sapiens' goes some way to supplying, as it locates the class of individuals as a species within the framework of a scientific theory. Thus 'homo sapiens' and 'human being' are natural kind terms (or at least surrogate natural kind terms), being classifications of a particular kind of animal. As such, they correspond to Locke's concept of 'man'.

Next, we come to the concept of 'person'. Is <u>it</u> a natural kind term? There is a strong argument that can be offered against such a claim, based on the fact that it fails to meet a crucial test that natural kind terms must satisfy. Wiggins, following Kripke and Putnam's causal theory of reference, argues that *"any would-be determination of a natural kind stands or falls with the existence of law-like principles that will collect together the actual extension of the kind around an arbitrary good specimen of it; and that these lawlike principles will also determine the characteristic development and typical history of members of this extension" (p169)*. Clearly, the terms 'human being' and 'homo sapiens' pass this requirement. however when we consider the concept of a person, it is not so obvious. As a piece of descriptive analysis, Locke was roughly on the right track

when he said that our usage of the concept of 'person' is essentially connected to the attribution of consciousness, in the wide sense of the term.

Starting from this insight, Wiggins has developed what he calls an 'animal attribute' model of personhood, in which he takes 'person' as a concept that has to be analysed firstly in terms of a natural kind component, which will name the particular species of animal that each individual can be classified as, plus a 'functional' or 'systemic' component, involving the attribution of higher cognitive faculties. So Wiggins' schema takes the following form :
"x is a person iff x is an animal falling under the extension of the kind whose typical members perceive, feel, remember.....; conceive of themselves as perceiving, feeling, remembering..... conceive of themselves as having a past accessible in experience-memory and a future accessible in intention, etc." (p171).
The dots here represent the class of intensional predicates. Another way of filling in these dots, in a way that seems to me to be completely compatible with Wiggins, is to list Dennett's [1976] 'conditions of personhood', namely being rational; being the subject of intentional ascriptions; the necessity of a certain attitude being taken to one as a moral agent; the ability to reciprocate this stance; being a language user; and the possessor of self-consciousness.

Regarding Wiggins' 'functional' component, care must be taken, since this term is also used in application to artefact-kinds, and while the concept of a person may share some significant features with artefact-kind concepts in the way that they differ from natural-kind concepts, persons are obviously not artefacts, and so the two categories should be kept distinct. In other words, a person is not 'for' anything - a person is not 'for' reasoning in the way that a knife can be said to be for cutting, as knives were created and designed by people in order to cut. Human persons, like all animal species, are in themselves functionless. However, their parts, i.e. their organs and bodily systems, can be legitimately regarded as having functions within the overall 'division of labour' involved in the process of life-maintainance - and it is in <u>this</u> sense that the mental attributes that Wiggins associates essentially with personhood can be regarded as functions, but not functions of persons, but of <u>minds</u>, and hence of <u>brains.</u>

So, according to Wiggins, 'person' is not a natural kind term,

but is a term that applies to a subset of animal kinds, with these latter kind terms being scientific classifications which are located within a corpus of biological and zoological theory. While I am for the most part in agreement with Wiggins here, I would guard against any tendency towards 'carbon-based chauvinism' regarding the attribution of consciousness, and I would leave it open, as an empirical matter, as to whether any future artefacts, silicon-based or otherwise, would satisfy any ideal Turing test and thereby deserve to be regarded as falling under the extension of the class Wiggins describes, and thus merit the attribution of personhood. From this perspective, 'person' can be seen, rather than as applying to a subset of natural kinds, as a cross-classification, cutting across the boundaries of all kind terms, both natural and, potentially, artefactual.

I would also add that we should regard 'person' as a secondary classification, for two reasons. Firstly, personhood is conferred by virtue of having some physical structure that is capable of supporting such psychological attributes constitutive of personhood. Secondly these psychological attributes are only contingently possessed by any member of any species. We can all sustain injury or suffer disease that would be sufficient to seriously diminish or even destroy such 'higher' functions, while the individual survives as a human being (or as whatever animal) by virtue of the continuation of the vegetative functions that keep the body alive.

'Person' cannot be treated as a natural kind concept since there are no lawlike principles that gather together all the individuals in its extension. This point is parallel to the argument of the functionalists against type/type identification of mental states and brain states. That is, what counts as a certain type of mental state is anything that is in the appropriate causal relations to certain stimuli and behavioural responses, and with other mental states; and since this specification can, in principle, be satisfied by more than one form of physical embodiment, then mental states cannot be identified with any one of these realisations, and hence psychology is irreducible to neuroscience. Likewise in our present discussion, what counts as a person is any subject that possesses the cognitive and self-reflective abilities as outlined by Wiggins and Dennett.

Thus we see the similarity between 'person' and artefact kind

terms. Not only must the class of animal-kinds that prospective 'person-tokens' fall under be kept open, but such a set of species that are so diverse in physical structure precludes the possibility of there being any nomological principles that could gather these individuals, and these alone, in the way that is required of a natural kind term. In classifying an individual as an example of a particular artefact-kind, we are not making any implicit reference to any set of nomological statements, nor to any theories composed of them, as the individuals in question cannot be gathered together as members of the given kind by reference to such laws. Nor can they likewise be collected by reference to some hypothesised inner structure underlying outer characteristics, since the materials and structures that various members of such a kind can be composed of may be totally different and divergent, and are in principle without limit, so no finite complete characterisation can be given of the physical embodiments of all persons at any time in such a way. Rather, such individuals can only be grouped together in the way we demand for artefact-kinds, that is, under a functional description, which is irreducible to any physical description.

Wiggins, an essentialist, puts it this way : each token person will have a real essence, which will be the essence of the particular species-kind he belongs to. Thers is no real essence as a person per se. To quote : *"Every person would belong to some natural kind that determines a sound Leibnizian principle of identity through change for some one kind of person (human-person, dolphin-person)..... indirectly this would be the real essence in virtue of which he was a person" (p172).*

As I have already mentioned, to classify individuals under a substance-concept like 'human being' involves a commitment to there being a set of nomological principles that gather the extension of that kind. In fact, according to Putnam [1970], it is a condition of the sense of such a predicate that such a set of natural laws hold for individuals that fall under such a term, and in such a way that they enable us to empirically discover what Wiggins calls the 'principle of activity' of that kind, including the characteristic forms of development for a typical member of that kind, and the limits of that development beyond which it ceases to be a member of that kind. Such a theoretical structure should provide us with principles of individ-

-uation for members of that kind, and enable us to identify and re-identify them, and thus provide us, in principle, with an answer to any query regarding the identity of any individual falling under such a description. It would seem to follow that the sense of such a sortal concept precludes us from even entertaining the possibility of any hypothetical example which involves human beings in situations that run contrary to the natural laws that contribute to the delimitation of the kind. For example, such fictions as depicted by Parfit [1984] concerning individuals whose bodies could fuse together, or those who could 'reproduce by natural division', i.e. fission, surely shed no light upon issues of personal identity as applied to humans, as such creatures could not be human beings. I will discuss this issue in detail in the following chapter.

4.4 ONCE AN F, ALWAYS AN F?

One other implication of Wiggins' theory is that any individual falling under a given substance-concept cannot change from being a member of such a kind to being a member of another one and still remain the same individual, i.e. once an F, always an F. Wiggins regards spatio-temporal continuity as a necessary but not sufficient condition of identity through time, as any such identity must be identity as something - it must be 'sortal-covered'. It goes without saying that such a condition does not cause any problems in the real world (unless one takes the Biblical story of Lot's wife literally !) but since a criterion of identity is intended to apply to all possible situations, it is worth seeing how such a restriction would apply to some counterfactual problem cases.

Parfit [1984] relates a science-fiction tale involving a 'Tele-transporter', a machine that facilitates interplanetary travel at the speed of light. The Teletransporter contains a 'Scanner' at departure point A, which destroys my entire body, while *"recording the exact states of all my cells" (p199).* This encoded information is transmitted by radio, at the speed of light, to a 'Replicator' at destination point B, which can, in principle, be anywhere in the Universe, and *"will then create, out of new matter, a brain and body exactly like mine" (p199).* Upon entering the Scanner, and once the mechanisms are activated, I will lose consciousness at point A, and wake up, seemingly instantaneously, at point B, as if emerging from a short nap. However, my original body is no more, as I will have woken up in a new body,

qualitatively identical to the original, down to the finest structures. And since the process involves the replication of my neurons, all my psychological attributes - my memories, my entire personality, will be unchanged.

Parfit then goes on to describe a second thought-experiment involving a new improved Teletransporter which incorporates a Scanner that can record the exact states of all my cells without destroying my body. However, it turns out that a design fault in the Scanner induces an irreversible cardiac disorder which causes the inevitable death of the original person within a few days, whilst the replica at point B is unharmed. Parfit notes that in this case there is some overlap between the lifetimes of what we might call (without begging the question) 'myself' and 'my replica'. So, in theory, during this time, we could talk to each other, and even see each other by means of two-way televisual phone systems. This fact of our coexistence means that I and my replica cannot be identified, since identity is a one : one relation. Our identification is also precluded by the branching of our respective tracks of spatio-temporal coordinates.

Another point, which Parfit doesn't fully bring out, is that immediately after the replica is created (and I am talking about minute fractions of a second), I and my replica will cease to be exactly similar. By the time it takes for my replica to phone back to me on Earth to assure me that he's arrived sfely, our physical forms will have diverged irreversibly. The two of us are living in different environments and receiving different stimuli, in response to which our (common) DNA will control and advise each cell in each respective body as to the production of appropriate proteins. Since each protein selection in turn becomes a factor influencing the selection of further proteins (again in response to surroundings and circumstances), myself and my replica will become increasingly dissimilar - so that if the design fault in the Scanner had been located and repaired prior to my entering it, and I had survived for another thirty years, then if after that time, I and my replica were to be compared, our qualitative similarity would fall far short of identity. It goes without saying that this divergence would take place on the psychological level as well. At most we would be like brothers - twins - and perhaps this is how we would come to regard one another.

Obviously these situations are not practically possible at the

moment. However, Parfit doesn't think that this fact invalidates their use as thought-experiments. He draws a distinction between the 'deeply impossible' - cases which would involve a contravention of the laws of nature - and the merely 'technically impossible' - cases whose impossibility is contingent on our technological limitations, which could be overcome in the future. One factor that Parfit doesn't deal with is the difficulty in telling whether a given hypothetical case is deeply or just technically impossible. If something is deeply impossible, then it is also technically impossible, but the reverse is not the case. For example, a heart transplant was a technical impossibility for an 18th century barber-surgeon, and, from his perspective, such an accomplishment may well have appeared to be deeply impossible. We, with hindsight, know that this is not the case, and that it was only deeply impossible in relation to his theoretical assumptions, which have in time been replaced. The history of science abounds with such misjudgements. All talk of deep impossibility can only be understood within the context of an underlying theory.

However, there are good reasons to be careful regarding the use of such thought-experiments. A cursory description (which is all that we are able to give in cases like Teletransportation) can easily fool us into thinking that we understand what is going on in such a situation - but in fact we are glossing over a vastness of ignorance, as we couldn't begin to describe more than a tiny fraction of what would be involved in such a case. I will return to discuss these issues in depth in chapter 5. But ignoring these doubts for the moment, I wish to investigate some implications of these thought-experiments. I will do this by devising three more thought-experiments, extending Parfit's line of inquiry. Let us call Parfit's original Teletransportation case T1, and his second, where the lifetimes overlap, T2.

T3 : I have been seriously injured in an accident, and my death seems to be inevitable. I am offered an alternative : space scientists have been attempting to colonize planet Z, in a far-off galaxy. The problem has been that humans cannot survive in the atmosphere of Z, and are restricted to specially insulated and controlled environments. Intelligent life has been discovered on Z. In fact, neurophilosophers have established that such Z-persons possess mental faculties just like our own, despite having a genetic structure that differs considerably from ours. In spite of this difference on the micro-level, Z-persons look just like humans on the macro-level. (Philosophers noted an uncanny likeness between these Z-persons and the subjects of

the twentieth century philosopher Putnam's 'Twin-Earth' story), Anyway, the offer is this : A Scanner will record all my brain states, thus recording all my memories, character traits, etc. My body is then destroyed. The information is sent to Z at the speed of light, and is programmed into a brain in a body made out of new matter, according to the blueprint of the body of a Z-person who looks exactly like me. It is in this new body that I will awake on planet Z.

T4 takes a similar scenario, except that it concerns planet Z', where Z'-persons, while again sharing the same set of mental attributes as humans and Z-persons, have a physical structure that is drastically different from our own both at the micro and macro levels, and so look nothing like human beings. It is as embodied in the form of a Z'-person that I shall awaken on planet Z', to be greeted by a hand-picked team of psychologists to help me get over the shock of my new appearance. All mirrors will be removed from the room.

T5 : Again I am terminally injured, and no alternative way out is open to me. Yet I am not despondent, as I know that this life is not the end for me. My psychological continuity will survive the death of this body and brain. I will lose consciousness, and awaken, as if from a sleep, in what seems [illegible] my physical body, and will see and recognise my friends and family who have also 'passed over'.

In T1 and T2, the traditional substance dualism has been replaced by a new dualism of matter and information in the form of my 'blueprint', including the exact specifications of the DNA molcules in the nuclei of my cells. T3 and T4 focus on the information comprising the exact states of my brain cells, and incorporates the functionalist claim that the same mental states can be realised by different physical systems. These cases also involve the survival of my psychological continuity through a change in kind - that is, the person that emerges from the Replicator is classed under a different substance-concept than is the person who stepped into the Scanner.

T5 is an adaptation of the Christian view of the afterlife, but in a form that is not fundamentally different from that of T1. God merely replaces the Teletransportation technology and somehow provides a means whereby such 'passed-over-people' can identify each other, by somehow re-embodying them or giving the appearance of having done this. Clearly, this version of the afterlife does not involve a commitment to

Cartesian Dualism. Indeed it seems to me that traditional Christian thought is metaphysically vague and underdetermined on the issue of what exactly survives the death of the physical body. Cartesianism is one development (and by now by far the dominant one) and, as I've argued, highly unsatisfactory because, among other reasons, it cannot account for how such pure disembodied souls can be identified and individuated. My story T5 is merely another development that avoids these problems. Like all the other 'T-cases', it involves the continuation of what Parfit calls 'Relation R', that is, psychological continuity by any causal means.

Marjorie Price [1974] devises a case involving Rover, a dog sent to Mars on a NASA mission, and who unfortunately becomes affected by Martian radiation, with the result that he gradually transforms into a living yet structureless blob, whose cellular structure does not even retain Rover's DNA. This amorphous mass is named 'Clover'. Now the question is raised as to whether Rover = Clover, i.e. are the pre-radiation Rover-stages and the post-radiation Clover-stages stages of the same enduring entity? By Wiggins' criteria we are forced to answer in the negative as, while they are certainly spatio-temporally connected, as one continuous 'track' can be drawn between them, there is no continuity under one same substance-concept. Clover is not a dog, and the only sortal concepts general enough to cover both Rover and Clover - e.g. 'organism', are excluded by their very generality from qualifying as substance-concepts. Price, on the other hand, does identify the two, at the cost of claiming that sortal-covered continuity is unnecessary for the holding of identity through time.

Another problem case is depicted in David Cronenberg's film 'The Fly', in which a scientist, Seth Brundle, devises a machine that is similar in principle (if not in scope) to Parfit's Teletransporter, whereby it records the genetic blueprint of the being inside the Scanner, and then creates a qualitatively new body out of new matter in the replicating chamber. However, it all goes terribly wrong. When Brundle attempts to transport himself across the room, a fly ventures into the Scanner with him, and the computer records the fly's genetic code in addition to Brundle's, and, since it has been programmed to only deal with one teletransported entity at a time, it integrates both genetic blueprints into one hybrid structure. Thus, when the scientist reemerges, it is with a genetic structure that is

neither completely human nor fly. To begin with, he looks and feels normal, but soon, grotesque mutations begin to take place as he is gradually transformed into a giant insect. His table manners decline dramatically. As these transformations begin to take hold, the scientist wryly re-christens himself 'Brundlefly'. So again we need to ask 'Is Brundle = Brundlefly?' By Wiggins' criteria, again we seem forced to say no, but this is counterintuitive. To show just how counterintuitive, compare it with a structurally similar example, T3, where my mind-set is programmed into a newly created Z-body, which, although exactly similar to my old body on the visible level, has an entirely different cellular structure. In this case, we would intuitively say that I survive - after all, not only do I look the same, but I also retain full psychological continuity. Yet here again, according to Wiggins, I and my Z-replica cannot be identified due to the lack of a common substance-concept between us.

4.5 CONDITIONS OF SURVIVAL

However, another option is open to us. Possibly Parfit's greatest contribution to the subject of personal identity was to turn the focus away from the question of whether x(t1) was the same person as y(t2), to examine instead whether Relation R held between them. As we have seen from Chapter 2, a major reason for making this shift was the existence of certain problem cases in which the question of identity was either undecidable or indeterminate. Secondly, Parfit argued that all that really mattered in survival was captured in Relation R, i.e. the existence of certain psychological states that are causally connected to one's own present psychological states. Unlike identity, which holds between an object and itself, and is therefore not transferable, Relation R can hold between two different beings. Also unlike identity, which is an 'all or nothing' relation, Relation R can hold to different degrees.

Andrew Brennan [1988] has substantially developed Parfit's theory, replacing Relation R with the 'survival-relation' (which I will call the S-relation). Like Relation R, the S-relation can occur between two different individuals, and can hold to different degrees. However, unlike Relation R, its application is not restricted to the psychological domain, but is used by Brennan to construct a general theory of what is involved in one thing surviving in or as another

thing, or in some future stage of the same continuing object. He regards this S-relation as being a more primitive relation than identity-through-time, and as being capable of covering and explicating 'what matters' regarding the future in cases of purported identity. He accounts for the S-relation in terms of a set of 'S-conditions'. Firstly he gives two conditions that necessarily hold when x survives in or as y :

Structure Condition : x and y must share the same structure. That is, their component parts must be of the same relative size, shape and positions to each other.

Causal Condition : x must play a significant and direct role in the production of y; specifically, x must be the 'prototype' of y.

A set of sufficient conditions under which x survives in y is obtained if we augment these with a third condition.

Matter Condition : (N.B. x's survival as y involves the holding of the S-relation to a higher degree than in x's survival in y.) x survives as y if, alongside the satisfaction of the Structural and Causal Conditions, y is constructed of matter of the same kind as x. (Note that this third condition cannot be a necessary condition of the holding of the S-relation, as, by it, x can survive in something else).

Before I go on to discuss whether I survive in or as my replica in cases T1-T5, mention must be made of another duality that employment of replication stories raises, namely that of types and tokens. At first glance, it is tempting to say that in all cases of replication, an item and its replica(e) are each tokens of the same type, with 'type' here not referring to a broad class of things falling under the same sortal concept, but in the narrower sense that any replica would be a token of the type 'Jim Baillie'. But the situation is more complicated than this. For a fuller account we need to look to Brennan's S-conditions. Brennan says that x and y are tokens of the same type if one survives (in his technical sense) to a suitably high degree as the other. To make this point clearer, let us look back at his Causal Condition, and employ yet another important distinction, between 'copying processes' and 'production processes'. In a copying process, one item is used as the prototype, the model on which the others are based. By contrast, an example of a production process would be a machine producing soupcans, each one qualitatively identical to the next, but where, taking any two

consecutive products, soupcan n and soupcan n+1, soupcan n does not survive as soupcan n+1, since soupcan n does not play any causal role in the production of soupcan n+1, but it is rather that they are both produced by the same causal processes deriving from the machine.

For Brennan, regarding copying processes, to say that two items are tokens of the same type is to say that one survives to a high degree as the other, whereas for production processes, it is to say that they have the same sort of structure and matter, and are produced by a common causal process, without themselves being causally related. In such a case, we have replication without survival. Such an account might fit Robinson's counterexample to Parfit's 'Physical Spectrum' argument (see ch.2. 6), where my twin and I are not causally related, but have been produced by the same production processes, via our parents. Another sort of case in which two items can have qualitatively identical structure and matter without one surviving as the other is in cases of 'adventitious copying' i.e. qualitatively identical items produced by entirely different causal processes. (It is easy to construct an analogue for Robinson's case here.) Brennan argues that strictly we shouldn't even call such a pair tokens of the same type, but rather just 'two qualitatively identical items', but this seems unnecessarily stipulative, especially since he himself has provided the means for distinguishing different ways in which items can be tokens of the same type, and has isolated the sorts of cases where the S-relation holds between the items.

4.6 TELETRANSPORTATION REVISITED

I will now return to my cases T1-T5, and re-examine them in the light of this S-relation.

In T1 (and T5), I clearly survive in my replica, as teletransportation is an example of a copying process, the prototype for my replica, (since all my structures, down to the finest level, have been copied). Clearly this analysis applies equally to T2. It is irrelevant that I also survive in my original body for a short time following the creation of my replica, as this result just enables there to be survival of myself twice over - I survive both in my original body, and as my replica.

T3 is more problematic in that it forces us to investigate the issue of structure in greater detail. If we say that I survive in my

Z-replica, are we saying, as Brennan suggests, that the Structure Condition is more important than the Matter Condition? Or can we make the point in another way by saying that structure at a different level - the macro level - is more important regarding the S-relation than is structure at the cellular level? The structure/matter distinction is not clear-cut, as matter, particularly on the organic level, is intrinsically structured.

T3 has similarities with pseudomorphism, a natural phenomenon discussed by Brennan, where the atomic ingredients of a crystal are gradually replaced by different elements which, although they form unit cells of a different structure (a unit cell is the basic structure of a crystal - crystals develop through repetition of unit cells), the overall crystal shape is very similar to the way it was before these changes took place, despite such changes in the constituent matter, e.g. although the atomic ingredients of fluorite (calcium and fluorine) have been replaced over a period of time by the ingredients of quartz (oxygen and silicon).

On the other hand, in the case of T3 we could focus on the fact that my psychological continuity remains intact despite having been teletransported into a body with a different genetic structure - in other words, we could say that what matters is the holding of Relation R. However, if we say that it is this feature which is of paramount importance, and that the form in which these psychological states are embodied is irrelevant, or is at best of secondary importance, then we must apply this view equally to T4. Another way of developing this Parfitian position and bringing it closer to Brennan is to return to the Structure Condition and note that on the psychological level there exists not just a bundle of perceptions, but an essentially structured bundle - the mind is essentially holistic, comprising a highly complex network of mechanisms for the input, processing, storage and retrieval of information. So instead of talk of psychological continuity, we can talk of the survival of this cognitive structure. In normal cases, the survival of the brain is paramount because it is the only physical structure in which this cognitive structure is realised, and so in such cases, the survival of a functioning brain is a necessary condition of the survival of the cognitive structure. However, the possibility of teletransportation looks as if it frees us (in principle) from this requirment.

But can T3 and T4 be accurately described as cases of copying? It seems strange to say that something composed of an entirely different type of matter can be a copy. But let us remind ourselves that copying, like the S-relation, can hold to different degrees, and guaging to what degree a Z or Z'-replica is a copy of me will depend on how we list our priorities. For example, we might say that my T1 replica is a better copy than my T3 replica (and therefore that I survive to a greater degree in my T1 replica) because we attach value to the retention of genetic structure as well as psychological structure. Likewise, we might say that my T3 replica is a better copy than my T4 replica due to the retention of my original physical appearance, which, while perhaps not as important as my genetic structure, still has great value to us, as is ironically shown by the fact that we go to so much trouble to look a certain way - we can easily imagine another case where I survive in a body that I'd much prefer to my original ! The overall important message, though, seems to be that the survival of our psychological structure is of utmost importance to us, and this corresponds to Parfit's claims regarding Relation R.

We can apply this S-relation model to the other problem cases. In both the Rover-Clover and the Brundle-Brundlefly cases, it is clear that the former survives in the latter to some extent, as can be seen by looking at the Causal Condition. Even in the case of Rover-Clover there is a continuous causally-connected chain of events governing the relations between each contiguous pair of time-slices throughout the process of mutation. This is also the case with Brundle-Brundlefly, which, in Brennan's terminology, was a botched attempt at a copying process, with the intended prototype being Brundle himself. Even despite the disastrous consequences, we can still accept that the information regarding his genetic blueprint played a major causal role in the creation of Brundlefly. Also, there is a high degree of structural and material similarity between Brundle and Brundlefly as he was immediately following teleransportation. Even though the seeds of his imminent mutation were beginning to develop on the cellular level, at this point his physical appearance was identical to that of Brundle on the macro-level. This urges us to ascribe the S-relation as holding to a high degree, and, of course, this claim is greatly strengthened when we recall that Brundle and Brundlefly are psychologically continuous. His memories remain intact, and while his character goes through a drastic set of changes (notably as regards

his preference in food !) the relation of psychological continuity is flexible enough to cope with these changes. In the course of time, Brundlefly goes through a succession of sudden and massive physical changes, but these are not sufficient to preclude the ascription of the S-relation, or even in some cases the ascription of identity through time - we need only recall the way in which a caterpiller is transformed into a butterfly. The important factor in both cases is that the changes are not random, but are causally induced.

I feel that these arguments by Parfit and Brennan are sufficient to tell us 'all that matters' regarding future survival. Once this content is subtracted, the further issue of identity is seen to be empty; and since the S-relation, unlike identity, need not be sortal-governed, we can solve such problem cases that Wiggins' theory runs up against.

4.7 IS IDENTITY SORTAL-RELATIVE?

My story T3 opens the door to another problematic issue, relating to debates concerning the relativity or absoluteness of identity. Those who hold to the thesis of the Relativity of Identity argue that the following situation is possible :

$[(a \underset{f}{=} b) \ \& \ (a \underset{g}{\neq} b) \ \& \ (f(a) \ \& \ f(b)) \ \& \ (g(a) \ \& \ g(b))]$,

whereas 'absolutists' like Wiggins deny that this is a possible state of affairs. This was precisely the situation envisaged by Locke in his passage about the prince and the cobbler. If we take 'a' and 'b' to refer to what I will call, in order not to beg the question, the cobbler-body-being at times pre and post-transmigration respectively, and 'f' to refer to the sortal concept 'man', and 'g' to refer to the sortal concept 'person', then Locke's argument amounts to the claim that a is the same man as b, but not the same person as b. I've already shown that Locke's argument is incoherent, given his treatment of memory.... but what if we were to update it and present it as a case involving a mind-swop, wherein the prince and the cobbler have been seized by our old friend the mad neurosurgeon, who records their respective brain-states by some electronic device whilst erasing them from their respective brains, and then switches them, with the result that the cobbler-body-being will now have the memories, character, etc., of the prince, and vice versa. In such a situation,

we are clearly inclined to say that the cobbler-body-being at pre-transfer time t1 is the same person as the prince-body-being at post-transfer time t2, yet we want to deny that identity as a man holds between these two, as this relation holds between the t1 and t2 stages of the cobbler-body-being, and between the t1 and t2 stages of the prince-body-being.

Shoemaker [1984] is troubled by such a possibility, and its apparent status as a counter-example to his desired advocacy of the absoluteness of identity. His only way out is to suggest that when we say that a person 'is' an animal (i.e. a man), this 'is' is not the 'is' of identity or of predication, but is rather the same usage of 'is' as is employed in sentences such as 'this statue is a hunk of bronze'. In other words, it would translate into something like 'is composed of the same stuff as', and thus the person and the man can share the same matter and the same spatio-temporal path, without being identified. I'm not sure how well this analogy travels, as in my neo-Lockean case we are not dealing with such a straightforward distinction between matter and form, but the more difficult distinction between matter and consciousness.

For my own approach, let me return to example T3. If I refer to myself as JB, and to my Z-replica as ZJB, then, given that there is complete psychological continuity between us, we would be inclined to say that JB was the same person as ZJB, but not that JB was the same man, nor the same animal as ZJB, since we are of different species, and, indeed, these species are located in entirely different systems of species-classification. So T3 corresponds to this :
$[(a \overset{f}{\neq} b) \;\&\; (a \overset{g}{=} b) \;\&\; (f(a) \;\&\; -f(b)) \;\&\; (g(a) \;\&\; g(b))]$.

So there appears to be a tension within my overall position, since, on the one hand, I want to hold that the most basic answer to the question 'What am I?' - i.e. what is the correct substance-concept to classify me as - is human being, yet on the other hand, in these problem cases, what seems to matter most of all is the retention of my psychological continuity, over and above the survival of my physical body, and so over and above my survival as a member of the species 'homo sapiens.' In other words, it seems that what matters above all else is that I survive as a person, or at least that some person continues to exist who is psychologically continuous with me. Still,

I think that my overall position is correct. Perhaps some tension is inevitable when we try to apply our concepts to such radically different counterfactual conditions. The important point is that my position is internally consistent, as I have shown that 'person' is not a substance-concept, and so, in such examples where what is under scrutiny is the question of what I <u>am</u>, 'person' is not in <u>opposition</u> to 'human being', as they are not mutually exclusive categories.

4.8. DISCONTINUOUS PERSONS?

I will now examine Wiggins' other condition for the holding of identity through time. He argues that for all members of all natural kinds, if $a \overset{=}{f} b$, then it must be the case that an <u>unbroken</u> spatio-temporal path can be drawn on which both 'a' and 'b' are located. This condition is commonly not regarded as being essential in the case of artefact-kind members, so, for example, if my watch is dismantled for repair, and lies in this dismantled state for some time (whether a day or a year, it makes no difference) and later reassembled, we would still regard it as being 'the same watch'. In the absence of any nomological factors pertaining to such artefact-kinds per se, and of any 'real essence' in the form of inner structure, we have nothing to go on beyond the facts of common linguistic convention. It is just a fact that we <u>do</u> regard it as the same watch, and it is hard to see any utility in any alternative convention whereby we would regard dis-assembly as being a breaching of the conditions of identity through time in the way that Wiggins (and virtually everyone else) thinks it does in the case of natural kind members.

On the contrary, I maintain that there <u>are</u> possible cases where we would say that identity was preserved in cases of the disassembly of such a 'natural' object, and I will show this even in the most challenging case, namely human beings. We only need to recall the rate of development of surgical techniques over the past few decades. Surely if one bisected a worm and stitched it together again using microsurgery, the 'survivor' would be the same worm as before? In the case of human beings, we commonly tolerate such examples of splitting such as the loss of organs and limbs, and if one's arm was severed, and sewn on again by microsurgery, issues regarding identity would not be thought to arise, as these would only be raised if the <u>brain</u> was tampered with. But imagine a case in which I am in need of brain

surgery that involves the removal of my cerebral cortex, and the temporary bisecting of it by the cutting of the corpus callosum, while work was done to it and to my sub-cortical brain structures, while my body is placed on a life-support machine. Once these operations are done, my corpus callosum is microsurgically repaired and the intact cortex is reattached to my brain-stem. This, I believe, would be a case in which we would be overwhelmingly drawn to say that I was the same person after the operation. . If, in the face of this, . . one protests that the one : one logic of identity has been breached by the splitting of the cortex, and that therefore we cannot say that my identity has been maintained, then, as I have suggested, all the more reason to drop all talk of identity in favour of the S-relation.

4.9 WHAT AM I?

So, in summary, many considerations have led me to the conclusion that the substance-concept that applies to a being like myself is 'human being'. I am not, however, urging a linguistic reform, urging that we should stop talking of 'persons' and 'people', because these concepts work perfectly well in their everyday contexts, where they are virtually synonymous with 'human being', 'man', etc. It is only when they take on their more philosophical sense that care is needed.

There are further reasons why I think that this specialised sense of 'person' not only does more harm than good, but is quite unnecessary. On the one hand, it admits less than the full extension of the class of human beings, and on the other hand its own extension stretches beyond the class of human beings. In the first case, large numbers of humans are born, or will become so handicapped that they will always lack the cognitive functions by which Wiggins & Co. specify personhood, and thereby, such unfortunates will be humans, but will not qualify as persons, and any talk of granting them some status as 'honorary persons', or treating them 'as if they were persons' is just a desperate ad hoc manouvre to get out of one of the unpleasant consequences of this form of the human/person distinction. Not only is such a denial of personhood repugnant to me, but we have no need to make such a pronouncement in such cases, as we can describe the situation perfectly well without making this sortal distinction, and without reference to this rarified sense of 'person'. They are just mentally handicapped humans (or 'persons', in the ordinary sense of

the term). Also, it is all too easy to see how such a distinction could be used to form the basis of arguments to the effect that 'persons' per se have more intrinsic rights than mere human beings.

In the second case, looking at the cross-classificatory aspect of the concept of 'person', we would be forced to speak of 'human-persons', 'dolphin-persons', and so on for any other species that we come to regard as displaying the appropriate cognitive faculties. (Martian-persons, anyone?) But again, such awkward ways of speaking are unnecessary because, should we discover that members of various species typically display the range of cognitive abilities previously thought to be unique to humans, then in time these abilities will be accommodated into the meaning of each particular substance-concept. To be more precise, it will become part of the conception (to use Frege's terminology) or, (as Putnam puts it) the stereotype of the concept - that is, it will become part of our theory, our network of significant beliefs about typical members of the given kinds, so again the suffix '-person' falls away as unnecessary.

CHAPTER 5 METHODOLOGY MATTERS

5.1 USES OF THOUGHT-EXPERIMENT

This short chapter is strategically placed around the midpoint of the thesis, and represents a kind of watershed, separating and distinguishing two different methods of enquiry, which could loosely (too loosely, in fact) be described as being, respectively, primarily 'conceptual' and 'empirical'. In this chapter, I want to cast doubt on the whole foundation of the enquiries, and therefore on the conclusions, of the previous chapters, and to suggest another, more fruitful way in which the issues of personal identity can be studied. This will involve a rejection of the methodology that has been adopted by most of the major figures in the field, and thus it will not be surprising that I will arrive at a quite different conception of just what the main 'issues' are regarding personal identity.

It will have been observed that in the preceeding chapters I have followed a fairly conventional approach to the subject, relying heavily on the use of thought-experiments, in accordance with a tradition that goes back as far as, at least, Descartes and Locke. One might say that, in essence, nothing has really changed since then- after all, what are Williams, Parfit & Co.'s neuroscientifically induced 'mind'swops', if not Locke's prince and cobbler soul-swop, clad in pseudo-scientific trappings? Surveying the copious literature, it seems possible that one can, with sufficient ingenuity, devise a thought-experiment that purports to provide support for <u>any</u> theory regarding personal identity, whilst in the near-certainty that an equally convincing counter-example lies around the corner. After a while one begins to wonder whether any progress has actually been made, and whether such techniques are appropriate to the tasks. The most noticeable trend in recent work has been the sheer proliferation of thought-experiments, with this expansion being not only in quantity, but in diversity and ingenuity of theme, which can seem to stretch our powers of credulity beyond breaking point. Yet such outlandish fictions are still taken as having some bearing on life as it really is. Most philosophers have not argued for this, being content to trade on the assumption that the events described in these stories were in some sense 'possible' and thus any criterion of personal identity must be able to accommodate them.

Recently however, Derek Parfit [1984] has taken a more sophisticated line, arguing that these bizarre counterfactuals have the power to [illegible] sharply [illegible] many of our most deeply rooted beliefs regarding the concept of a person, and concerning what it is to be the same person through time; and since these beliefs underlie actual events in the real world, the employment of such thought-experiments is justified. As well as this capacity for the edification of concepts, such a method can provide opportunities for conceptual revision. As Wilkes [1988] describes it, by *"stretching a concept into the unknown one may find out more precisely what it is to which we are committed..... if the concept fractures under the strain - if, that is, we would not know what to say in the hypothesized circumstances - then too the scope and limitations of the term's range and extent become clearer" (p5).*

To adequately assess such claims, we need to examine the applications of thought-experiments in more detail. In what follows, I will draw upon the recent work of Kathleen Wilkes [1988], which, if correct, has the iconoclastic consequence that a great deal of influential recent work on personal identity is rooted in fundamental error. After describing and assessing her arguments, I will then return to survey some notable thought-experiments as employed in previous chapters, re-evaluating them and their conclusions from a new perspective.

The use and justification of thought-experiments can be stated quite simply. Suppose that we want to test some theory, or to find out the consequences of such a theory being true; or again, suppose we want to investigate the range of application of a given concept. We can do this by asking a 'what if....?' question. In other words, we imagine that a certain state of affairs has taken place, and then try and work out the implications of such an event, and then reassess the theory, concept, or whatever, in the light of these judgements. Or, using 'possible worlds' terminology, we can say that such events have happened in some possible world. However, it is essential that these events take place in *"a world like our own in all relevant aspects, except for the existence in that world of the examined phenomena" (p2).* In other words *"the possible world is our world, the world described by our sciences, except for one distinguishing difference. So we can know or assume everything else that it is relevant to know, in order to assess the thought-experiment" (p8). (All quotations in this chapter are from Wilkes [1988]).*

However, it is not the case that 'anything goes' in the employment of thought-experiments. They are quite literally experiments, and as such are bound by the same methodological canons as all experiments which aim to tell us about the world. One major constraint that all thought-experiments must conform to, whether in philosophy or in the sciences, is to state, or be able to state the background conditions against which the event under scrutiny takes place. This rule is essential because the thought-experiment is specifically designed to draw out the consequences of a given event, or factor within an event, and this can only be done if we can ensure that the same consequences could not be due, in sum or in part, to some other auxiliary factors. So, since we require that the possible world in which the event takes place be like our world with the exception of this one alteration under scrutiny, we must ensure that the background conditions to this event are well-specified, and remain consistent with conditions in our world.

When such requirements are satisfied, then the derivation of the consequences of the experiment is a relatively straightforward matter (although it is not absolutely immune from the possibility of error). Such successful thought-experiments tend to occur mainly in the natural sciences. In contrast, the history of thought-experiment in philosophy and in particular as applied to the subject of personal identity has been, according to Wilkes, one long catalogue of disaster. To see why, we must return to the above-mentioned requirement of the specification of background conditions. Let us by contrast examine the situation in the natural sciences, where we not only see the satisfaction of this demand of strictly defined experimental conditions (in particular, since the experiment often takes place within a closed system) which allows a complete description of the relevant background, but we also have the advantage of an explicit and developed theory governing the experiment, couched in theoretical terms that are strictly defined, which allows any assumptions and entailments regarding the experiment to be correspondingly clear-cut. Thus, under such circumstances, it is a straightforward matter to determine whether or not a particular factor is relevant to the results achieved.

5.2 ABUSES OF THOUGHT-EXPERIMENT

By contrast, the concepts in which theories of personal identity are couched are not capable of such strict definitions, and so any

inferences that we do attempt are not straightforward, but are problematic and without a firm basis. We are left at the level of deciding 'what we would say if.....' with no trustworthy means of evaluating such conclusions over and above our intuitions, which are. of course, conditioned by our beliefs, So [illegible] *"when we are dealing with the rich and riotous chaos of commonsense concepts, we are dealing with terms that generally do not pick out natural kinds, and so there is no body of explicit theory or shared and agreed generalisations about them; we are rather dealing with implicit and partial, rough and ready assumptions. Hence the importance of intuition grows in direct proportion to its precariousness" (p16).*

But the most serious problem with the typical philosophical thought-experiment lies not [illegible] the vagueness of natural language, but [illegible] in the fact that such experiments fail to sufficiently fix the background conditions. In the absence of such specifications, the experiment fails to be adequately described, and this has the crucial consequence that we do not know if the hypothetical situation depicted within it is a real possibility.

We are now getting to the heart of the matter. Very often, philosophers have made an illicit jump from the fact that they can, in a sense, 'imagine' or 'conceive of' some state of affairs, and this does not appear to them as self-contradictory, to the claim that such a situation is a real possibility.Thus Descartes reckoned that he could imagine his continuing to exist as a thinking being while being disembodied, and he took this feat of suspended disbelief to support the claim that such a situation could actually happen. In fact, rather than describe his error as being in some jump from 'imaginability' to 'possibility', it is more accurate to say that whatever the mental imagery involved in such a thought-experiment, one cannot imagine being disembodied. It only seems as if you imagine this due to the cursory description given to the contents of the experiment. Once you try and think it all through, things look less obvious - for example, what would such an existence be like? How could you move without a motor cortex and muscles? How could you perceive without sense-organs, and so on. As Wilkes says in her general point about such cases, *"the fact that we may not have identified all the relevant laws makes no difference, except to mislead; since we are largely ignorant of these and their interconnections, it of course seems easy to imagine the transformations; the obtrusive facts are not there to obtrude" (p31).*

Likewise, it is easy to be fooled into thinking that you can imagine a bar of iron floating in a pool of water. But once the physical properties of both these substances are taken into account, and once the background conditions are specified to preclude the introduction of invisible strings suspending the bar, unusual anti-gravitational forces, etc.,it soon becomes clear that no matter what the mental image you are having looks like, it is not a correct account of the behaviour of iron or water.

It is not enough to say that the event described in the thought-experiment is logically possible - it must be 'theoretically possible' - in other words, a satisfactory description of the background conditions provides us with a theory of the kinds of objects involved, and this structure enables us to decide whether the hypothetical event is a real possibility for the kinds of object involved. Wilkes describes this theoretical possibility as providing *"something between stringent essentialism and loose conventionalism, something that will allow us to insist..... that the human species is a kind governed by law, while not denying that some of these laws may fail to hold of individual members of the species." (p28).* I will attempt to develop this idea further by returning to the ideas of Wiggins and Putnam, discussed in the last chapter.

We can follow Putnam in saying that natural kind terms are 'law cluster concepts', where there is a determinate set of 'core facts', most of which must hold for a given object to qualify as a member of that kind. This cluster of laws will limit what is theoretically possible for a member of that kind, and help determine what presuppositions and implications are relevant to any thought-experiment involving a typical member of that kind. Wilkes develops this position, pointing out that such laws governing biological kinds do not apply in isolation, totally independent of each other, but on the contrary are closely interrelated. So, *"it will be rare that we find an isolated breach of a single law. It is far more likely that such a violation will have consequential effects upon other laws, where such laws are either at the same sort of level (in the same or a different theory) or where they are more fundamental laws that describe the operationa and the limitations of such higher level laws as are supposed to be violated"*, displaying the fact that *"the physical, as well as the mental, is holistic, with laws arranged in a systematic hierarchy of mutual dependence" (p30).*

So, with all this in mind, let us examine a typically adventurous thought-experiment devised to reveal our deepest beliefs regarding personal identity, namely a popular one in which we are asked to consider what we would say if people could divide like amoebae - so, if A split to form B & C, what is A's relation, as a person, to both B and C? As neither/the same person as B nor C? As B but not C? As C but not B? I needn't repeat all the problems with each choice.

Now it should be fairly clear that we cannot leave the description of this hypothetical state of affairs at <u>this</u> level, that is, superimposing this radically different form of human development over <u>our</u> world, with its cluster of laws governing biological processes, and assume that everything/else goes on as [illegible] normal. Firstly, the background conditions haven't even begun to be described, and, as I have said, since such a development is taken to occur outwith the strictly defined limits of a scientific experiment, such a statement of the background conditions will need to include a significant proportion of the possible world. Indeed it is hard to specify a limit on the amount of information that is relevant to the description of such a thought-experiment that we would need to know before we could derive reliable conclusions from it. When we attempt to spell out all that is required, the whole enterprise descends into farce. I can do no better than to quote once again from Wilkes : *"It is obviously and essentially relevant to the purposes of this thought-experiment to know such things as : How often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be induced, or prevented? Just as obviously, the background society, against which we set the phenomenon, is now mysterious. Does it have such institutions as marraige? How would that work? Or universities? It would be difficult, to say the least, if universities doubled in size every few days, or weeks, or years. Are pregnant women debarred from splitting? The <u>entire</u> background here is incomprehensible. When we ask what <u>we</u> would say if this happened, who, now, are <u>we</u>?" (p11).*

This last question brings us back to David Lewis, and the question of how many people there are in a room at a given time, if such people are subject to fission. I argued in chapter 2 for the incoherence of his views. Wilkes comes to the same conclusion by a different route.

We can now relate all of this to the point about the holistic nature of the law-cluster that limits the real possibilities for

members of natural kinds. It is, I think, clear that any possible world in which persons could split like amoebae would be highly different from the real world in ways that we could not even begin to conceive of, and thus we don't know to what extent, and in what ways, the background conditions in such a thought-experiment are consistent with conditions in the real world. The most that we can say with any certainty is that they will differ greatly - but this is not sufficient to put us in a position to determine the theoretical possibility of the phenomenon posited, or of these conditions themselves. A related point is that any beings that could divide like this would need a physical structure that is radically different from our own, and thus they could not be of the same species as ourselves, and may be individuated along entirely different principles to us. So no matter what we would want to conclude about the identity-conditions for such creatures, we cannot use it to derive conclusions about ourselves - but, of course, this was the whole point of constructing the thought-experiment in the first place.

5.3 HUMAN FREEDOM AND NATURAL LAWS

It might be argued that all this talk of clusters of laws that constrain human possibilities misses a crucial point. Such an analysis seems to locate human beings in the position of passive slaves of immutable laws which strictly and absolutely delimit their range of possible activities and developments. It might be argued that such a depiction doesn't take into account the fact that human beings are active participants in a dynamic, changing universe. It is becoming clear that human knowledge and expertise (and abuses thereof) have reached a level where we ourselves now constitute an active initiatory factor in determining the development of our species, and that of other species. Take genetic engineering for example. Or consider the consequences of the desrtuction of the tropical rain forests, including the 'greenhouse effect' which could possibly involve large climatic disruption in the near future. While I don't imagine that we will start splitting or fusing, or anything as outrè as that, who can say what forms our species may evolve into in order to survive in the face of such challenges? And who is to say, by some principle, that such future beings would not be human?

WHile our physical structures (anatomical, physiological,biological

etc.) place us under significant practical limitations at any given time, they do not do so completely. In other words, any Wiggins-style 'principle of activity' that applies to human beings is not an absolute, static and determinate affair, but is open-ended, and we ourselves play a part in its development. To talk in Sartrean terms for a moment, our principle of activity is waiting to be actualised by us, and does not have any 'essence' prior to this.

I make these rather melodramatic remarks as a prelude to discussing one prominent philosophical thought-experiment that is far away from the excesses of amoeba-like fission, fusion, time-travel and the likes, and focusses on a more credible case that involves the possible application of human ingenuity in initiating certain events that would seem to force us into making radical revisions to our beliefs about personal identity and unity. It is of course an empirical matter, but it seems to me that there is one type of splitting that could be shown to be within the realm of real possibility in the forseeable future, namely the transplantation of intact neural structures, (note that I do not say 'brains').There has been an extraordinary rate of development in the practice of transplant surgery, in the techniques of microsurgery, in overcoming the problems of rejection, and so on, all within my lifetime. For example, thirty years ago a successful heart transplant was considered an impossible dream - yet it was a dream that was realised in the '60's, and since then the techniques have gradually been improved and refined to the extent that at the present time a single person can survive the transplant of the heart, lungs and liver. Add to this the fact that numerous experiments have been performed in which various parts of the brains of salamanders have been exchanged, and, as Wilkes informs us, the entire heads of monkeys have been swopped. Given such astounding achievements, all which [illegible], until recently, would have been probably judged to be 'deeply impossible', we can surely have no apriori justification preventing us from accepting that human brain-transplants, or hemisphere transplants, look as if they are theoretically possible.

Now do all these considerations go against the position regarding the validity of thought-experiments that Wilkes has proposed? I do not think so. I believe that any decent account of the conditions of legitimacy of thought-experiments regarding personal identity must be able to accommodate these points. We can render compatible the two

seemingly divergent strands by saying the following :

1. What is and what is not theoretically possible is a function of a given background theoretical structure - there is no 'theoretical possibility' as seen from an outside, theory-free standpoint.
2. No theory, no matter how well corroborated, is immune from error, even fundamental and widespread error.
3. So if some state of affairs S is judged, on the basis of theory 1, to be theoretically impossible, and theory 1 in time becomes replaced by theory 2, it does not follow that S is theoretically impossible according to theory 2.
4. It is also the case that we have no unified, complete and satisfactory theory of human beings that could, even within the scope of the freedom that I've argued for, lay down precise limits upon what is a theoretically possible state of affairs for human beings to enter into for all future times. We have a variety of theories, in physics, chemistry, biology, psychology, sociology, and so on, at various degrees of maturity.

Taking all these points together, we have to say that we just don't know if the situations depicted in philosophical thought-experiments are real possibilities in the real world. Some may well be so. However, it does not follow that any thought-experiments based on such situations can yield trustworthy conclusions about this world, or regarding the theories or the concepts employed by us in the real world, since we still lack the ability to fill out the background conditions adequately. It is one thing to say that we don't know that the situation depicted therein is impossible. It is another thing entirely to say that it actually is a real possibility, and, as Wilkes says, *"ignorance is a poor justification for any experiment, scientific or philosophical" (p20)*. So we have no real grounds for saying that such hypothetical situations are real possibilities, and, on the contrary, surely the sheer fact that we are, on occasion, unable to even begin to flesh out the background conditions is, if anything, evidence against its possibility, despite the limitations of any theories we currently employ.

5.4 THOUGHT-EXPERIMENTS REASSESSED

I will now, as promised, return to some of the most influential

thought-experiments used in the study of personal identity, and re-assess their worth in accordance with the new line taken in this chapter. One useful division that can be made among these experiments is between those involving some form or degree of brain-transfer, and those involving a mind-transfer. In other words, the first sort of experiments involve the removal and subsequent replacement of neural structures, i.e. the 'hardware', whereas the second sort concentrates on the 'software', that is, the actual contents of the mind/brain.

I will begin with the first sort, and take issue with experiments involving 'brain-swops'. Now this term can clearly mean several different things. As I have mentioned, philosophers are often far too lax in their talk of brains, equating the brain per se with the cerebral cortex. Now if we literally mean brains, whole brains, that are to be removed and exchanged, then it is clearly not enough to remove the cortex. The brain stem will also need to be removed. And what of the spinal cord? This is integrally connected with processes required for the possibility of conscious experience. Can it be removed? The central nervous system functions as one overall system, and any distinction between those parts that can be called 'brain' and those that can not will be rather grey. But over and above this conceptual issue, we don't know what would happen if all these structures were exchanged and put in new bodies. We are not monkeys, and so any results regarding the swopping of monkey heads is not conclusive evidence that a brain-swop or head-swop is a real possibility for human beings. And even if it succeeded on the basic physical level, and both humans survived the operation, we don't know what state their higher cognitive faculties would be in. Granted, we do have a strong hunch that person A would wake up in the body into which his brain was transplanted, but such guesswork is not a strong enough foundation to base a philosophical theory on.

On the subject of cortical or hemispherical exchange (e.g. 'My Division'), I have two points to make. Firstly, one of the major premises used in constructing this thought-experiment is wrong - a split-brain patient - one in whom the corpus callosum has been sectioned - does not have 'two minds' and is not 'two persons'. Furthermore, the brain is not split, but rather the means for direct intercommissural information exchange is removed. But any fragmentation or division in consciousness resulting from this operation takes place against a

background of great unity, both structural and functional, via the intact subcortical structures. (I discuss this in detail in ch.7). Secondly, we have the problem of ignorance again - we do not know what would happen if, say, one cerebral hemisphere was transplanted into a cortexless skull. We do not know if the memories, character traits, etc. of the person whose hemisphere it was will be preserved intact, or if those of the person whose body and subcortical brain systems would dominate, whether they would integrate, whether they would conflict, or, more likely, whether death would ensue.

Thirdly, let us examine thought-experiments involving brain-exchange on a cell-by-cell scale, such as in Parfit's Physical and Combined Spectra. Such stories seem to me to be based on at least one false premise concerning the nature of memory storage in the brain. It is assumed that memories come in discrete quanta, one (or however many) located within each neuron. So the neuron itself is assumed to be the container of memories. But strong experimental evidence suggests that this is not so, and that long term memories are stored in the form of an altered electrical resistence in the synapses between neurons. So we cannot 'atomise' the brain in such a way as to achieve any neuron : memory correspondence, as there is an irreducibly relational factor between neurons regarding memory.

Regarding experiments where a qualitatively identical brain or body is created from the original prototype, I have little more to say than to merely stress our virtually complete ignorance of what would be involved in this, and so we have no way of telling if it is theoretically possible to do so. Wilkes makes the good point that a qualitatively identical replica means precisely that, and so all the original's structures would have to be reproduced, down to the sub atomic level.

I will now turn to mind-swops. Again, the illusion of coherence within such tall tales is purely down to our ignorance of any of the 'obtrusive facts' that might get in the way. We cannot state the background conditions of such experiments with any clarity, and so the experiment is inadequately described, and therefore not only do we not know if it describes a real possibility, but we cannot use such an experiment to derive conclusions about personal identity in the real world. So we can now see the root cause of all the confusions and contradictory beliefs that resulted from Williams' [1970] two cases of

mind-swop. Both contrary conclusions we are led to by the respective cases are unwarranted and untrustworthy, as they are based on different but equally inadequate descriptions of the conditions involved in such an experiment. Our intuitions are conditioned by our beliefs, and in this case, we are not in a strong enough position to acquire a set of well-justified beliefs concerning the pair of thought-experiments. The fact that the two cases that elicited the opposing responses were alternative descriptions of the same state of affairs merely exposes how out-of-its-depth our intuitions are in such a situation.

As will be expected, if the undisciplined use of thought-experiments is curtailed, and if the 'problem cases' which could only exist within the confines of such fictions are deemed irrelevant to real-life issues, then clearly we will emerge with a very different conception of what the central philosophical issues regarding personal identity are, and some well-pedigreed problems will be made obsolete. In the following three chapters I will deal with some issues regarding personal identity without thought-experiments.

6.1 LOCKE'S CRITERION

I will begin this chapter with a detailed discussion of John Locke's theory of the relationship between memory and personal identity. This may appear as if I am trying to smash an egg with a sledgehammer, devoting far too much time and energy in attacking a position that can be easily dismissed. However, it should be borne in mind that Locke's theory has been the starting point for virtually all contemporary forms of the Psychological Criterion, and whilst these latter theories involve significant modifications and improvements on Locke, I will argue that they share some fundamental errors with him, rooted in an erroneous conception of what are the essential issues regarding memory and personal identity. I will also argue that some contemporary neurologically based theories of memory share some mistaken assumptions with Locke, and so I am justified in paying his theory serious attention.

John Locke [1694] explicitly associates the holding of personal identity, and indeed of personhood itself, with the ability to remember one's past experiences : *"As far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person" (II xxvii 9); "That with which the consciousness of this present thinking thing can conjoin itself, makes the same person, and is one self with it, and with nothing else, and so attributes to itself and owns all the actions of that thing as its own, as far as that consciousness reaches, and no further" (II xxvii 17).* (my underlining)

To start the discussion, let us examine the second of these two quotations. Anthony Flew [1951] noted that this passage can be interpreted in two very different ways, depending on how the term 'can' is understood. On the one hand, it can be interpreted in the 'logical' sense, which could be spelled out as meaning 'can without contradiction', and on the other hand it can be read in the 'factual' sense, as meaning 'can as a matter of fact'. Flew argued that Locke's position (or positions) can be made more explicit by recasting it as the claim that having memories of one's past experiences constitutes the necessary and sufficient condition for the holding of personal identity through time. Looked at in this way, turning first to the interpretation

involving the 'logical' sense of 'can', we have the claim that x (t1) is the same person as y(t2) iff x and y are both persons and y can (logically) remember what x experienced at t1.

However, this clearly will not satisfy the requirements of a criterion of personal identity, as it is far too lax. Too many hypothetical cases would be able to satisfy such a criterion, where we would certainly not want to confer identity. For example, there is no formal contradiction involved in an imaginery case wherein various laws of nature are broken or overruled so that I now remember experiencing certain events as they were originally experienced by Napoleon. But, as I have already argued in previous chapters, any purported criterion of identity that can accommodate such counterfactuals leads to absurdity if its implications are followed through. The problem can be expressed like this : If I remember experiences that were undergone by Napoleon, then by the above criterion, I must <u>be</u> Napoleon. If I then remember experiences as undergone by Nelson, then I must be Nelson. But if so (since if x=y and x=z, then y=z) it follows that Napoleon and Nelson must be the same person, and since this conclusion is false, it follows that at least one of the premises is wrong. I needn't go on to list all the contradictions issuing from such an identification, i.e. one both being and not being one-armed in 1805, etc.

If, on the other hand, we try to rehabilitate the criterion by fixing the meaning of 'remember' so that, as a matter of logic, one can only possibly remember experiences that are one's own, then the resulting formulation is a tautology, and, as such, cannot function as a <u>criterion</u> at all, since a criterion must actually say something substantial, it must have some empirical content. Such a move would throw the baby out with the bathwater. Problems would be glossed over at the expense of precluding the possibility of making any significant statements about the relevance of memory to personal identity. And anyway, even if we were landed with such a restriction on the meaning of 'memory', 'remember' and related terms, we could then go on, as contemporary philosophers have done, to introduce the more general concept of 'quasi-memory' (see ch.3), which is free of such stipulated logical restrictions, and we would be back at square one, with the same old problems rightly reinstated.

Turning now to Flew's second formulation of Locke's criterion, we have the claim that x (t1) is the same person as y (t2) iff x and y are both persons, and y can, as a matter of fact, remember what x experienced at t1. (Incidentally, if we are being strict, x and y are not persons,

but person-stages). However, in stark contrast to the previous formulation, this one is clearly far too strict, as it excludes real-life cases where we are bound to accept that x and y are stages of the same person. In fact, such a criterion would exclude all actual cases of purported identity through time, since we are all constantly forgetting some of our experiences, and in order to satisfy this criterion, we would have to remember all our past experiences, and clearly none of us, not even Luria's celebrated mnemonist, can claim to do this. In fact, if we spell out the demands of this criterion, it seems that not only must one remember every experience that has occurred throughout ones life, but one must retain this all-encompassing memory at every moment of ones life, if the criterion is not to fall prey to the paradoxes that Reid exposed (see ch.6.3). Since I at t2 will only remember at most some of my experiences up to and around t1, yet, for Locke on this interpretation, what constituted my identity and person-hood at t1 was the possession of all the experiences at that time, then Locke cannot identify me at t2 with any person at all at t1. Rather, he only offers us a theory of how some experiences undergone at t1 are taken on by a person at t2, and subsequently, as Mackie [1976] so eloquently puts it, what we have is *"hardly a theory of personal identity at all, but might better be described as a theory of action appropriation" (p183).*

6.2 'MEMORY' DISMANTLED

An immediate response to my objections is to say that they merely highlight the fact that there was a virtual absence of theory in Locke's day concerning the nature of memory, beyong it being a 'storehouse of ideas' in the mind. Indeed, despite the enormous advances that have been made since Locke's day, we still do not have a complete and satisfactory account of memory. Most theories assume the existence of a 'memory trace' or 'engram', it being the ultimate physical embodiment of each specific memory, these being representations of past experiences, and each being in some way a 'structural analogue' (in Martin & Deutscher's [1966] words) of its corresponding experience, but there remains great conceptual obscurity regarding exactly how such a structural isomorphism can take place. It also remains the case that no memory trace has ever been discovered, no doubt due in part to this conceptual difficulty entailing that we don't really know what it is we are looking for. The memory trace remains a theoretical entity, inferred to exist because to deny its existence in some form would

mean that memory involved 'action at a distance', and it is a deeply entrenched belief (and one of Broad's 'Basic Limiting Principles') that there can be no direct causation through a temporal gap. We require a continuous causal chain, and there must be some physical basis for the continuity of this process, bridging the gap between the time of the experience and its later recall, and, whatever this physical basis is, its unit has been called a 'memory trace'. Given our ignorance regarding memory, we have tended to base our explanations of memory on analogies with things that we do understand, from Plato's impressions in wax, through Penfield's videotape recorder playbacks, to more contemporary computer models in terms of input - storage - readout.

Various distinctions and divisions of types of memory have been suggested, and, at the very least, these catagories are useful in describing and ordering diverse phenomena, and cautious employment of them will enable me to take the traditional philosophical arguments relating to memory and personal identity considerably further. Let us begin by dismantling the old unitary concept of memory, and talk instead in terms of a tripartite division of registration, storage and retrieval. Looking at the distinction between information storage and retrieval, one might argue for a reconstructed neo-Lockean criterion, saying that we needn't demand that any stored memory should be able to be retrieved at any given time, but merely that we should insist that all information regarding past experience is stored in the form of memory traces throughout one's entire life, whether or not these memories can in practice be retrieved, and thereby recalled to conscious awareness at any given time.

Locke, of course, would find such a revision incomprehensible, since his primitive model of consciousness could not accommodate the concept of information that was stored yet was beyond the reach of conscious awareness, since he had inherited from Descartes the idea that all that one knows must be known-to-be-known by the knower : *"that consciousness which is inseparable from thinking, and, as it seems to me, essential to it, it being impossible for any one to perceive without perceiving that he does perceive..... consciousness always accompanies thinking" (II xxvii 11).*

The introduction of this storage/retrieval distinction will enable a reconstructed Lockean theory to move away from its Cartesian assumption of the epistemically privileged position of first-person

psychological reports, and accommodate Freud's insight into the existence of unconscious mental contents whose existence is unknown to the subject. But even without mentioning this Freudian distinction, it is quite obvious that this last statement quoted from Locke is totally wrong. To see how, let us distinguish two mental states :
i) The state of being happy at t1. This involves the form of 'consciousness' described in passages such as when he defines the self as *"a conscious thinking thing..... which is sensible or conscious of pleasure or pain, capable of happiness or misery" (II xxvii* 17).
ii) The state, at t1, of being aware of being happy at t1. This is the form of consciousness referred to in II xxvii 11 above, where Locke seems to say that this state ii) necessarily accompanies state i).

Such a ii)-type state is dependent on the occurrence of what we might call the 'first-order' mental states of type i). It is also a higher, more sophisticated type of state than a i)-type state, in that, for example, while we would ascribe i)-type states to higher animals, we would not do so for ii)-type states. But for my purposes, the main point is that Locke is wrong to say that such ii)-type states are "inseparable from thinking" (i.e. from i)-type states), but, on the contrary, are occasional fleeting states that are absent for the vast majority of waking life. This is not sufficiently recognised, partly due to our internalisation of Cartesianism and partly, I believe, due to a 'trick' relating to the nature of ii)-type states, namely that drawing them to one's attention brings such a state into existence. For example, if I say to you 'What are you thinking about?', and you reply 'My dinner', and I then ask 'Are you aware of thinking of your dinner?', then, presuming that you understand my question, you become thus aware. But this awareness soon slips away unnoticed.

6.3 MEMORY STORAGE

But let us return to this proposed revised criterion, that the continuity of memory that is necessary to preserve identity requires only that the memories continue to be stored, whether or not they are retrieved. Is this a credible criterion? Can it cover everyday cases? The problem is whether there can be any other evidence, apart from the retrieval into conscious awareness, that memories are stored. Until the time of death, it can never be conclusively known that a memory of a given experience was not irretrievably lost to conscious awareness - the most that we can ever say is that it has not been remembered yet.

It is undoubtedly true that we store an enormous amount of information more than we can retrieve at any given moment - but is it possible that every experience is stored in some memory-system for one's entire life, and can, in principle, be retrieved if one is presented with the appropriate cue or stimulus? The major problem with such a claim, when offered as a serious scientific hypothesis, is that it is clearly unfalsifiable in principle, since the non-retrieval of any information can always be put down to the lack of suitable stimuli, rather than its not having been stored in the first place. But even if there is no possible evidence that could falsify the theory, is there any evidence that might be put forward in favour of it?

It is sometimes suggested that Wilder Penfield's [1958] experiments, where precise points on the exposed temporal cortex were stimulated by live electrodes while the patient was conscious, [illegible] provide such evidence, as they resulted in the surprising phenomenon that patients experienced extremely vivid sensations when the electrodes administered stimulation. These experiences were thought by Penfield to be recollections of past experiences. From this, Penfield theorised that every individual memory is stored in some stable and specific location in the temporal cortex, and, secondly, that if a given point were stimulated, whether artificially or 'naturally' within normal brain function, then the appropriate embodied memory would be brought to conscious awareness.

However, both Penfield's experimental results and his theoretical model have been discredited. He thought that whenever we pay conscious attention to something, we simultaneously 'record' this experience, which is then stored in some structurally isomorphic form within the temporal cortex. So he saw memory in terms of a tape-recorder model, where every experience is stored completely and, given the appropriate stimulus, can be 'played back' in full. But Penfield's actual results are far too humble to warrent such dramatic conclusions. Firstly it should be kept in mind that all these experiments were carried out on epileptics (in fact the phenomenon was discovered by accident while operating on patients to reduce seizures) and so any findings relating to these people should not be extrapolated in order to propose theories regarding the functioning of normal brains. Secondly, the aforementioned phenomena were extremely rare occurences - out of 520 patients, 40 reported having some sensation whenever the electrode touched the cortex, and out of these, only 12 could be put forward as possible cases of genuine recall - that is, under 3% of his total. Thirdly, of this 3%,

no independent check was ever made to corroborate the claim that these experiences were genuine memories. Fourthly, the results can be explained within other equally plausible theories, for example that they were genuine memories, but that their recall had nothing to do with the electrode, but that they were somehow associated with what the patient was thinking about at that time. Fifthly, there is always the possibility (albeit an unverifiable one) that if another point on the cortex had been stimulated instead at exactly the same time, the same experience would have been recalled.

Penfield himself made an observation that would seem to disprove his theory, when he noted that if the same point was stimulated twice, a few minutes apart, the same experience was not summoned twice, and indeed on occasion different experiences were brought about. This would at least go against a 'static' theory of memory traces. It has long been a controversial issue in the neurosciences whether the memory trace has a static/structural basis or a 'dynamic' one, or whether these are combined, with short term storage being dynamic, in the form of a patterned current continuously running between neurons, and long term storage being static, based on an altered electrical resistance in the synapses between neurons.

In relation to this whole issue, in studies of people who have had substantial areas of their cortex either removed or significantly damaged - i.e. up to 40% - there seem to be no specific and selective gaps in memory, as would seem to be implied by any theory positing a precise, static 'one : one' correlation between memories and points on the cortex. So perhaps there are many tokens of each memory trace, distributed throughout the cortex. Finally, from an evolutionary standpoint, it would seem to be unnecessary to have a brain that stored everything, every last insignificant detail. It makes much more sense to have a brain that actively selects rather than passively records, that can sort out the wheat from the chaff (however imperfectly) and retains information that appears to be useful and relevant to the future.

6.4 PARFIT'S PSYCHOLOGICAL CRITERION TESTED

Returning to Locke, one of the most noteworthy attacks on his theory, and the one most responsible for the modern modifications to

it, is Reid's [1785] argument that, according to Locke's theory, *" a man may be, and at the same time not be the same person that did a particular action" (p114).* This argument involves the famous example of the officer who had been flogged as a boy for stealing apples from an orchard, who had later captured a standard on his first military campaign, and who had ended his career as a general. The paradox arises that *"it is possible that, when he took the standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging" (p114).* It follows from Locke's theory, says Reid, that, given that both he and Locke mean by 'being conscious of', being able, given the appropriate cue, to remember the relevant experience, that the schoolboy was the same person as the young officer, and the officer the same person as the general, but that the general was not the same person as the schoolboy - but since the transitive nature of the identity relation demands that these two be identified, given the previous two identifications, it seems that by Locke's theory, the general both is and isn't the same person as the schoolboy.

This argument succeeds against Locke's theory. More contemporary forms of memory-based criteria of personal identity, versions of the Psychological Criterion, are designed to cope with Reid's objection. They do so by rejecting Locke's restriction of the criterion to the holding of <u>direct</u> memory connections between experiences and their future recall, and replacing it with a model of psychological continuity involving overlapping <u>chains</u> of memory-relations. So, returning to Reid's counter-example, and calling the boy, the officer and the general a,b and c respectively, c remembers b's experience, and b remembers a's experience, and at all times in between there is a continuous chain of memory connections that is sufficient to maintain identity despite the loss of any direct memory connections between a and c. Such is Parfit's Psychological Criterion, discussed in ch.1, which I will restate, in a shorter, clearer form :
x (t1) is the same person as y (t2) iff x is psychologically continuous with y and with no-one else at t2.

Clearly this criterion can, unlike Locke's, accommodate such commonplace phenomena as forgetting, since our normal rate of forgetting is not great enough to threaten either of the criterion's two integral relations of psychological continuity or psychological connectedness. It can also deal with another common phenomenon which I have never

seen mention of in any philosophical discussion of personal identity, but which is fatal to Locke's theory, namely infantile amnesia. In virtually all cases, adults can remember virtually nothing of the first three years of life, so it follows that by Locke's theory that one is not identical as a person with the infant from which one developed. However, even though I now have no direct memory connections with myself at one year of age, I can be sure that some form of strong connectedness held then between successive days, albeit in some pre-conceptual way, i.e. recognising familiar faces, places, objects. But I accept that there is an oddity in even applying a psychological criterion in the case of the very young, i.e. the new-born. Surely infants and adults encode and interpret the world in very different ways. What if Piaget [1971] is correct in stating that cognitive systems are not staticly structured systems that remain in essentially the same form throughout one's life, but are subject to change and development, proceed from infancy to adulthood by a series of stages that are qualitatively very different from each other, and that this cognitive development is not a uniformly gradual process, but is subject to leaps. While it might not be exactly right to describe such leaps as paradigm shifts in ways of worldmaking (as to talk in these terms would seem to imply a change in explicit theory, which would imply the use of language and the ability to think in concepts, whereas an infant's early stages are pre-conceptual), it might be held that a leap from one stage to the next results in the child seeing the world in a fundamentally different way, for example coming to understand that objects exist independently of one's perception of them.

Does the existence of such profound changes in our modes of representing and interpreting the world pose a threat to our notions of psychological continuity? This might be so if the qualitative changes took place in instantaneous 'quantum leaps', but this is not what Piaget is arguing for - it is a more gradual process than this, and the rate of change in conceptual development is still compatible with the holding of psychological continuity. Also, these stages proceed in an invariant order, with a rich network of causal connections underlying this process. In connection with this causal aspect, Piaget stresses that the child plays very much an active role in his own cognitive development. This is in opposition to another feature of Locke's treatment of memory, and indeed Penfield's (it is remarkable how similar their assumptions are), with their passive 'storehouse' model whereby experiences are somehow copied and filed away for future reference. However, as the noted neuroscientist J.Z.Young [1987] has

recently remarked, in direct opposition to such a theory, the brain *"is not simply a storehouse of records, but a very active organiser" (p17)*. Likewise, as Frederic Bartlett [1932] says, *"The first notion to get rid of is that memory is primarily or literally reduplicative or reproductive" (p32)*. In various tests, Bartlett showed that people do not merely register events passively, but interpret them according to previous knowledge, values and interests, and that when information is stored, it is not left in one permanent form, but is constantly undergoing modifications, reevaluations and simplifications as information is gradually lost, so that what is remembered at a later date will not be an exact copy of what was initially experienced, but a 'version' of it.

So far, then, Parfit's Psychological Criterion looks to be in good shape. It faces a more serious challenge when we come to consider various forms of amnesia, which are sometimes seen as undermining the continuity that such a criterion requires. I will now discuss some forms of organic amnesia (that is, where there is a known physical cause, as opposed to psychogenic amnesia, where the trouble is psychologically based).

Traumatic amnesia is a common consequence of receiving a severe blow to the head, for example following a road accident. It has three main phases. The first stage is post-traumatic amnesia, commencing when the subject first regains consciousness. It is characterised by a general disorientation, of not knowing where you are, what day it is, how you came to be where you are, etc.. This is followed by a period of retrograde amnesia, where one's equilibrium becomes more reinstated, but where events prior to the accident cannot be recalled. At the beginning of this stage, this amnesia can stretch back to cover a period of many years. However, in time, memories are gradually regained although the moments immediately prior to the accident can never be recalled. The common explanation for this is that the information has not been sufficiently consolidated, as the trauma of the accident has interfered with the transfer to long-term storage. The third stage, anterograde amnesia, which overlaps somewhat with the second stage, is characterised by tiredness and by difficulties in concentrating, which result in problems in learning and retaining new information.

6.5 THE SLEEPING PILL

Given the limited and temporary nature of this form of amnesia, it is clear that Parfit's Psychological Criterion will not be threatened by it. However, Parfit constructs an argument involving retrograde amnesia to draw some startling conclusions. This occurs within his thought-experiment entitled "The Sleeping Pill" (in [1984]), in which he makes a challenging comparison between the predicaments of such an amnesiac and the subject of a teletransportation scenario. The story begins with the well-acknowledged fact that some drugs, notably some sleeping pills, can produce retrograde amnesia, whereby I will remain awake for one hour after having taken the pill, but on the following day I will not remember anything that happened during the half-hour before I fell asleep the previous night. So if I've taken such a pill just over half an hour ago, the person who wakes up tomorrow will be psychologically continuous with me as I was before taking the pill, but not with me now. I am, according to Parfit, presently on a 'psychological branchline' which will come to an end when I fall asleep. During this half-hour, I will be psychologically continuous with 'Myself-in-the-past', but not with 'myself-in-the-future',, and so the relationship between 'me now' and 'me tomorrow' is like my relationship to someone else, in as much as any intentions that I form now will not be acted upon tomorrow, and any bright ideas I have at present will not be remembered the following day, unless I resort to public means of communication, such as writing myself a note.

Parfit says that this predicament is analogous to a variant of the teletransportation story T2 (let us call it T2') in which my replica is not created until I die. He suggests that the situations are structurally identical in the sense that if we give the names a, b and c respectively to, firstly in the sleeping pill case, myself before taking the pill, myself after it has taken effect but before I have fallen asleep, and myself when I wake up the next day, and secondly in T2', respectively to myself before entering the Scanner, myself in my original body after the teletransportation of my genetic blueprint, and my replica, then in both cases we have psychological continuity between a & b, and between a & c, but not between b & c.

If Parfit is correct about his assimilation of these two cases, then we seem to be landed with a difficult dilemma, since, if the cases are structurally isometrical, then, given that in T2' we do not want to ascribe stages a & b & c to one person, then we seem to be equally

constrained against doing so in the sleeping pill case, but this is enormously counter-intuitive. On the other hand, if in the sleeping pill case we do admit all three stages as being of the one same person, we seem bound to do so in T2', which again is most undesirable.

Andrew Brennan [1988] argues that such a dilemma is an illusion, since Parfit is mistaken in comparing the two cases. He analyses the cases in terms of his 'survival conditions' (see ch.4) and argues that in the case of T2', 'a' survives both as 'b' and as 'c', but that 'b' does not survive as 'c', since 'c' is not causally related to 'b'. However, both 'b' and 'c' are causally related to 'a', 'b' through physical continuity, and 'c' being an example of a copying process, using 'a' as the prototype. By contrast, in the sleeping pill case, the amnesia experienced by 'c' is causally dependent on events happening to 'b', as the pill has interfered with the brain's processing functions, so that any experiences occurring at that time cannot later be retrieved by 'c'.

Brennan is correct in pointing out the asymmetry between the two cases. However, it seems to me that his analysis of the sleeping pill case, despite any intrinsic merits it may have, has moved away from anything that can with any accuracy be understood in terms of a psychological criterion, as his analysis invokes causal processes that fall under a physical description of events. If it were to be fleshed out, his analysis would be in the language of neurology and pharmacology and not of psychology. This is no problem for me, but it may be for Parfit.

An alternative way of dealing with the problem of retrograde amnesia in the sleeping pill case and similar situations is just to say that the Psychological Criterion is loose enough to accommodate such lapses in memory, as psychological continuity still holds to a high enough degree to maintain identity, given the continuity between a and c. This position is strengthened considerably when we reject the privileged status of memory-connections within the Psychological Criterion and widen it to include such factors as consistency of character, of likes and dislikes, the continued ability to use language, to perform learned tasks, all of which are unaffected by retrograde amnesia. To combine this with Brennan's argument, we can say that we have a theory of amnesia which can be inserted as a clause within a larger theory of persons based on a wide version of the Psychological Criterion.

6.6 *VARIETIES OF MEMORY*

One point that I keep returning to is the need to discuss memory, and therefore any Psychological Criterion, in terms of a far more conceptual scheme than has usually been employed in philosophical treatments of the issues of personal identity. Far too often, philosophers have unwittingly followed Locke & Co. in taking a unitary view of memory. However, it is now well-acknowledged that memory involves not one single unitary system, but a number of complex interconnecting systems, each with a different specialised function, but cooperating in the shared general task of storing information for future use. It follows that all talk of someone 'losing his memory' is rather misleading, if not incoherent, as such a statement tacitly assumes this discredited unitary model. Amnesic syndrome is not a general or overall loss of memory function, as in such a case one would in all probability be dead or at most confined to a 'vegetable' existence. Nor does it involve a general deterioration in memory function, but a selective impairment in long-term storage, where some systems can be severely damaged while others are relatively unaffected.

The following discussion of severe forms of organic amnesia will make use of the well-known distinction between long-term and short-term storage. For our purposes, short definitions will suffice. Long-term storage refers to all information that has been stored in such a way that it is potentially retrievable over times greater than a few seconds, whereas short-term storage comprises the set of systems facilitating the temporary storage of information required to perform another more inclusive mental operation - for example, the practice of mental arithmetic requires the temporary storage of numbers which can be retrieved at the appropriate stage of the operation, after which they can be discarded. Likewise, in understanding a sentence, one needs to be able to remember the first words in the sentence when the last words are being spoken or read.

Organic amnesia is caused by damage to the limbic system of the brain, in particular to the medial temporal lobe of the cerebral cortex, and a subcortical region called the diencephalon, including the hippocampus and mammiliary bodies. Major contributary factors include encephalitis, brain tumour, stroke, and Korsakoff's syndrome, the last of which I will now discuss. Korsakoff's syndrome is a result of chronic alcoholism and subsequent thiamine (vitamin B1) deficiency.

One problem in ascertaining the degree of amnesia suffered in such cases is that the syndrome is usually accompanied by other intellectual and cognitive impairments. However, in general some major memory systems remain unaffected. Short-term storage is usually intact, as are linguistic abilities and motor skills. SUch patients retain semantic memory, that is, general knowledge, but in severe cases this can be restricted to information acquired before the onset of the disease. For example there are numerous reported cases who, on being asked who is the present Prime Minister, will reply 'Harold Wilson', or some other name that would have been the correct answer had the question been asked at the time relating to the portion of the past still accessible to the patient's memory. This peculiar phenomenon is due to the fact that Korsakoff patients have great difficulty in learning (and therefore remembering) any new information - so they suffer from both anterograde and retrograde amnesia.

Korsakoff patients show some residual learning capacity, which shows that procedural memory (information in long-term storage relating to complex skills that can only be given a partial description, e.g. riding a bike) isn't fully destroyed. SUch patients can also retain the ability to perform various tasks and skills that they acquired prior to the disease. THis was strikingly exhibited by Clive Wearing, the subject of a Channel 4 documentary 'Prisoner of Consciousness'. He was not a Korsakoff case, but his amnesia was caused by the herpes simplex encephalitis that affected his frontal and temporal lobes. He was, and remains an accomplished musician, and continues to read music, to play, arrange and conduct just as before. The peculiarity of his predicament was dramatically displayed when he was invited to play on a church organ, on an instrument that he had regularly performed in the past. He protested vehemently that he had never seen it before in his life and that he'd never played any musical instrument before. He was astonished when the music began to flow from his fingertips.

Returning to Korsakoff's syndrome, the other major area of disfunction is in episodic memory, that is, 'autobiographical memory', the ability to remember past experiences from the inside. Again, this can remain intact for periods prior to the onset of the disease, so that, for example, the patient may retain vivid memories of his youth, to the extent that he may believe himself still to be that age. By contrast, with reference to times since the disease took hold, (and sometimes for years before this), his past is a stranger to him. For

example, in a severe case, if you were to meet such a patient for the first time one morning, and left, returning the same afternoon, he may not recognise you, nor remember having met anyone. It is therefore understandable that such patients are not aware of there being anything wrong with them - how could an amnesic remember that he couldn't remember? Perhaps this is a blessing. Two case histories, recorded by Oliver Sacks [1986], will serve to display the horror of such a predicament.

6.7 TWO CASE HISTORIES

Jimmie G. suffered from severe retrograde amnesia due to Korsakoff's syndrome. When Sacks first met him in 1975, Jimmie thought that the year was 1945, and that he was still 19 years of age. When Sacks confronted him with his mirror image and asked if the middle-aged man reflected there could be a youth of 19, Jimmie went into a blind uncomprehending panic, which, perhaps mercifully, was extremely short lived, since the experience was forgotten within a few seconds, once his attention was distracted. In general, he had no sense of the passing of time at all. As Sacks observed, *"he is a man without a past, stuck in a constantly changing, meaningless moment" (p28).* On being interviewed at the hospital where he had lived for many years, he asked of Sacks 'What is this place,doc? Do I work here?'. He possessed no semantic or episodic memory concerning anything after 1945. Beyond this date, there was no certainty, and he was left with only guesses, instantly forgotten, to try and make sense of his situation. Sacks recalled Hume's [1739] account of the inner life of man as essentially comprising *"a bundle or collection of perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement" (Bk.1 ch.4 p6),* and ruefully reflected that such a picture is realised not in people at large, but in such damaged cases as Jimmie G., the 'Humean being'.

William Thompson presents an even more extreme case of Korsakoff's syndrome, having, unlike Jimmie G., no 'base', no recall of any part of his life, nor knowledge of any facts. He has no idea who he is, adopting and casting aside personae rapidly and randomly, and likewise seeing any other person addressing him as being up to a dozen different people in the space of five minutes, moving from one construction to another, apparently unaware of any discontinuity, and

delivering each 'version of reality' with equal force. In fact, to him each of these conflicting accounts were equal in value (i.e.virtually valueless), and such basic distinctions as true/false, real/unreal, relevant/irrelevant, important/trivial cease to apply to his ontology. To quote Sacks, *"for him they were not fictions, but how he suddenly saw or interpreted the world....its radical flux and incoherence could not be tolerated, acknowledged for an instance" (p104).* He could perceive there to be nothing wrong with himself, precisely because he was lacking any stable viewpoint or perspective from which to make such a judgement. In his state of permanent Korsakoff psychosis, he was *"continuously creating a world and a self..... such a patient must literally make himself (and his world) up at every moment" (p105).*

6.8 WHAT IS THE REAL ISSUE?

Do these tragic cases present serious problems for Parfit's criterion? Firstly, turning to the relation of psychological connectedness : since short-term storage remains intact, one could say that most Korsakoff patients possess a fair degree of psychological connectedness in the extremely short term, i.e. moment to moment, although even this may be denied in the case of William Thompson. By Parfit's definition (see ch.1), such a relation would only count as connectedness to a very low degree, since the relation of 'strong connectedness', which is required for identity to be maintained, requires that at least half the direct psychological connections are retained into the following day. This condition would definitely disqualify William Thompson. In the case of Jimmie G., one could argue a case for his having psychological continuity of a sort, on the basis of his well-established memories of events and experiences prior to 1945. However, it would be impossible to decide if this number of memory connections surviving day to day would count as sufficient to maintain his identity (apart from anything else, we do not have strict criteria for counting such things). However, if we broaden the Psychological Criterion as I have suggested above, and include in the equation such factors as continuity of character, of tastes, of values, then Jimmie G. is seen as possessing a much higher degree of psychological continuity than was previously judged. This approach gains support from the testemony of Clive Wearing's wife, who stated that, throughout her husband's illness, *"I've never lost touch with the 'Cliveness' of Clive..... his soul, his person is unchanged..... he's still the same man"*. I understand her to be pointing at least in

part to the forms of continuity I have mentioned. However, even after all this, William Thompson still fails to make the grade, as in his case the flux extends to even these newly included aspects of mental life. He cannot with any accuracy be said ... even to have a personality, likes and dislikes, opinions, etc., as such concepts assume an underlying consistency that he does not possess. He cannot be even said to have a self (in the sense that I will explicate in ch.8). So even my new Broadened Psychological Criterion fails to accommodate him.

A criterion of personal identity must be able to apply to all cases where identity holds, and any version of the Psychological Criterion will have insuperable problems in achieving this, since psychological continuity is not realised by one unitary system but by a complex structure of interrelated systems, and, basically, anything that is structured can be dismantled, to varying degrees, often leaving only the empty shell. No version of the Psychological Criterion can accommodate all of the disparate mishaps, catastrophes, and malfunctions that can afflict us. Any criterion with any real substance will fail some severe cases, such as William Thompson.

Let us return to the beginning of the chapter, to Flew's distinction between the 'logical' and 'factual' uses of the term 'can' regarding the ability to remember past experiences. We saw that both usages failed to provide a satisfactory criterion of personal identity within a Lockean framework. Flew was mistaken in thinking that he had exhausted the possibilities inherent in the term 'can'. What we need here is the 'can' that limits what is theoretically possible, in the sense described in ch.5. For this, we need an account of how memory works, and, within this, an explanation of when and how these functions break down. I have supplied an outline of this in this chapter. The introduction of this 'theoretical possibility' into the arena forces us to ask fundamental questions regarding what a criterion of personal identity is, and what it is for. For example, from Locke through to Parfit and his contemporaries, their various formulations of the Psychological Criterion are located within their theories of personal identity in terms of their opposition to, and their juxtaposition to a Physical Criterion. This conflict, this 'either/or' structure developed because it was considered 'possible' that the two criteria could diverge and provide different answers to questions of identity. Hence it was a very real question as to which of the two criteria was correct or had primacy, and which covered what was essential to

personhood and identity. But if such a divergence is not theoretically possible, this whole enterprise has been one long red herring.

Turning back to Clive Wearing, Jimmie G. and William Thompson, what exactly are we asking when we raise doubts as to whether their personal identity is maintained over time? For in one sense there is no mystery, no doubt as to who they are. We know who they are. There is only a 'philosophical' question here if the Physical and Psychological Criteria are put in opposition to each other. If we do so, the problem is that by the restrictions imposed by any substantive Psychological Criterion, someone like William Thompson fails to be anyone. (Of course, one can argue that, in terms of the 'philosophical' sense of the term 'person', he is not a person, but merely a human being, but for reasons given in ch.4, I will ignore this easy option). Once a cease-fire is declared between the criteria, we can clearly see that to ask something like 'Given the holding of physical continuity between them, is this person addressed as 'William Thompson' at t1 the same person as addressed as 'William Thompson' at t2?' is just the wrong question to ask. In Parfit's terms, it is an empty question, it being possible for one to say 'yes' and for another to say 'no' (i.e. 'he's not a person'), yet be in full agreement about the 'facts of the matter', both neurological and psychological. There is no philosophical mystery here concerning identity, just the philosophical error of asking the wrong question.

5.9 Appendix : *PSYCHOGENIC FUGUE*

However, matters are more complicated when we consider another form of amnesia, namely psychogenic fugue, a condition which acts here as a 'bridge' between the amnesic disorders discussed in this chapter and the dissociative disorders to be discussed in chapter 8. 'Psychogenic fugue' is characterised as covering cases where a person typically disappears, wandering far from home for days or weeks on end, during which he has no recall of his life before his flight - that is, he suffers a complete loss of episodic memory (but with his semantic and procedural memory unaffected), but with this amnesia remaining unrecognised by himself. Such fugue conditions typically end as suddenly as they started, with the person suddenly 'snapping out of it', often after sleeping, with the memories of his past life prior to the fugue fully restored, but with his memories of times during the fugue state now lost.

William James [1890] describes a typical case of psychogenic fugue, wherein one Rev. Ansel Bourne of Providence, Rhode Island, disappeared on 20/1/1887, and nothing was discovered concerning his whereabouts until 14/3/1887 when, in Norristown, Pennsylvania, one A.J. Brown, who had arrived in the town some six weeks before and had opened up a general store, awoke one night in panic. Brown had recognised himself as Ansel Bourne, and not knowing where he was, how he had got there, what he was doing there, remembered anything about the previous few weeks in Norristown 'as' A.J. Brown. The last thing he remembered was withdrawing some money from a bank in Providence on 20/1/1887.

Three years later, William James hypnotised Bourne in order to retrieve these lost memories, but when Bourne recalled them in a hypnotic trance, he, 'as Brown', was amnesic regarding his life 'as Bourne', saying that whilst he had heard of him, he "didn't know as if he had ever met the man". James confessed that *"I had hoped by suggestion, etc, to run the two personalities into one, and make these memories continuous, but no artifice would avail to accomplish this, and Mr Bourne's skull today covers two distinct personal selves" (p393).* James assumed that Bourne's state of fugue was a spontaneous hypnotic trance, since only by putting him under hypnosis did he succeed in recovering his lost memories 'as Brown'. However, it doesn't follow from this success that hypnosis produced the amnesia in the first place.

But it is James' claim that the man before him constituted 'two distinct selves' that I wish to investigate in chapter 8, in order to see whether any sense can be made of a literal, non-metaphorical interpretation of this claim. This will lead me to investigate the fundamental concept of 'unity of mind'. This is also the central issue of the following chapter, on the implications of split-brain research.

CHAPTER 7 COMMISSUROTOMY AND THE UNITY OF MIND

7.1 'THE BRAIN' DISMANTLED

I want to turn now to consider another contemporary area of investigation that has been recognised as having profound and far-reaching implications for the study of personal identity and the central concept of 'unity of mind'. In particular, the whole question of synchronic identity, resting on a notion of 'psychological unity-at-a-time' has been called into question by experimental findings in the field of split-brain research.

Looking back over the history of philosophical inquiries into personal identity, one of the most noticeable developments taking place in the present century has been the convergence of the two traditionally opposing positions, the Physical and Psychological Criteria, due to their common acknowledgement of the unique significance of the brain to their inquiries. Firstly, in the case of the Physical Criterion, the brain was regarded as the one organ or physical structure that one absolutely could not lose and yet continue to exist, or to exist as the same person. It was assumed that, given that the continued identity-over-time of a material object requires that there be an unbroken spatio-temporal path by which this object and it alone could be identified, then in the case of living human beings, the essential 'traveller' along this path is the brain. In other words, I go where my brain goes. I can lose my hair, my arms, my heart, lungs and liver, yet still remain me, but the brain is the exception to this rule. Thus, in thought-experiments, it was argued that if A's brain were transplanted into B's body, the resulting person would be A, and the operation would be more accurately described not as B having a brain transplant, but as A having a body-transplant, with this situation being merely the limit to common cases of organ transplant. For their part, supporters of the Psychological Criterion conferred special status on the brain as being the physical means whereby all distinctly psychological phenomena - notably memory - were realised.

In all such discussions, the brain was treated as a single unitary organ in the sense that it was regarded as being the indivisible bearer of consciousness and of selfhood. In other words, it was taken for granted that the functional integrity of the brain could not withstand division or substantial diminution, and that no

less than a complete and fully operative brain could succeed in embodying a full human consciousness. However, once a more detailed and systematic knowledge of the workings of the brain began to be acquired, this assumption began to be undermined, with the result that, in Nagel's [1971] words, it began to look possible that *"the personal, mentalist idea of human beings"..(and in particular "the idea of a single person, a single subject of experience and action") ..... "may resist the sort of coordination with an understanding of humans as physical systems, that would be necessary to yield anything describable as an understanding of the physical basis of mind" (p147).*

Although, as we shall see, this unitary conception is so deeply entrenched that it may be impossible in practice to prevent ourselves from thinking of ourselves in terms of a single unified subject of experience, this inability of ours does not negate, defuse, or diminish the deep conceptual problem that Nagel is pointing to.

7.2 COMMISSUROTOMY DESCRIBED

Turning now to split-brain research, it should be noted, of course, that 'split-brain' is something of a misnomer, smacking slightly of journalese. The expression refers to commissurotomy, which, strictly speaking, does not result in the bisection of brains (as the brain is a highly complex organ, and many of its component parts remain intact after the operation) but is restricted to the bisection of the main connection between the two cerebral hemispheres.

There are various forms of commissurotomy, all involving the cutting of the corpus callosum (literally 'thick-skinned body'), a large transverse band consisting of around 800 million nerve fibres, which directly connects the cerebral hemispheres, thus facilitating direct communication between them. (see figure 1). In a complete commissurotomy, the entire corpus callosum is sectioned, along with the underlying hippocampal commissure, one fornix, the anterior commissure, and the massa intermedia of the thalamus. This results in the division of the cerebral cortex, with the cerebral hemispheres being connected only indirectly via the brain-stem and subcortical routes, and thus causing the splitting in sensory and motor functions to be discussed shortly.

A central commissurotomy involves the severing of the corpus

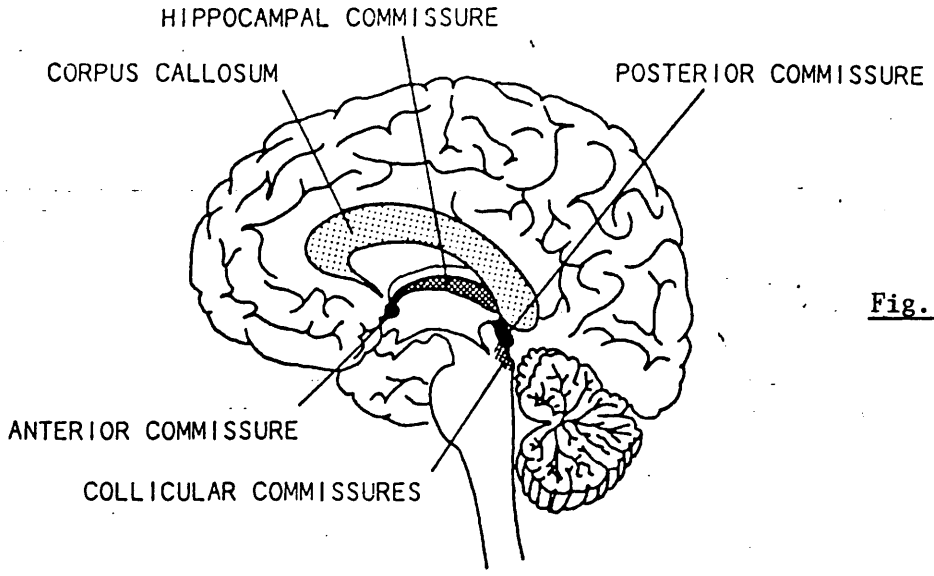


Fig.1

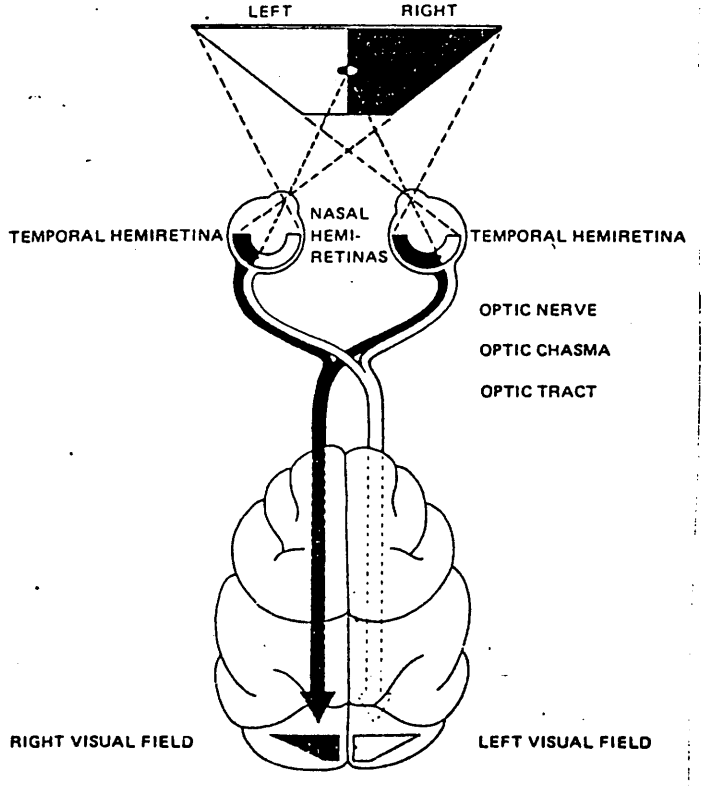


Fig.2

Both adapted from Gazzaniga & Ledoux [1978]

callosum and the hippocampal commissure, resulting in a tactual split but not a visual split. A frontal commissurotomy involves the sectioning of the anterior portion of the corpus callosum, the anterior commissure and one fornix, and results in relatively little splitting of functions, even under experimental conditions.

At the present time, complete commissurotomies are rare, as modern diagnostic techniques (EEG, CAT scans, PET scans, CT scans, nmr scans) can precisely chart the area of the cortex that the epileptic seizure (which warrants the operation) is issuing from, and thus we need only section the corresponding part of the corpus callosum that would relay the discharge to the other hemisphere. However, it was complete commissurotomy that was the common form of the operation for some time, and it has been this operation that has been the focus of philosophical interest. Accordingly, in the following pages, unless otherwise stated, I will use the term 'commissurotomy' to refer to the operation in its complete form. I will, however, return to the more limited forms of the operation later, as they form crucial evidence for my final assessment of the philosophical implications of split-brain research.

Commissurotomy was introduced in the 1940's as a drastic measure to control grand mal epileptic seizures, which originate as an electrical disturbance in a particular site in one cerebral hemisphere, spreading across it to the other hemisphere via the corpus callosum. Since the disturbance increases in magnitude as it proceeds, the idea behind the operation was that by cutting the corpus callosum, the disturbance would not only be limited to its hemisphere of origin, but its magnitude would thereby be severely curtailed. In this respect, the operation proved to be successfull beyond all expectation, as attacks not only became less severe, but also less common, and in some cases disappeared entirely.

Despite the operation's undoubted success regarding epilepsy, it was in time discovered that it had some peculiar consequences, which I discuss below. However, these side-effects were not immediately apparent. Much of the literature on the subject tends to exaggerate the cognitive and behavioural consequences of commissurotomy, as they understandably concentrate on the anomalies. Rather, the fact is that such patients' everyday behaviour appears, both to the patient himself and to observers, to be virtually unaffected most of the time. Indeed, close investigation by clinical researchers at the time did

not discover any peculiar results. This led many neurologists of the day to ascribe only a minor role in brain function to the corpus callosum - or, indeed, to wonder if it had any function. One wit suggested that it was only there to transmit epileptic seizures between the cerebral hemispheres.

By the 1960's it was discovered that under certain controlled experimental conditions, the responses of split-brain patients was found to be highly unusual, and alarming in its implications. Specifically, when sensory input was restricted to one hemisphere alone, and a response to this stimuli was requested from that hemisphere, it appeared, in Sperry's [1968] words, that such patients possessed *"two separate spheres of conscious awareness, two separate conscious entities or minds running in parallel in the same cranium, each with its own sensations, perceptions, cognitive processes, learning experiences, memories, and so on"(p296).*

However, despite a misleading gloss of unanimity that is sometimes presented by neuroscientists and philosophers on this subject, it must be kept in mind that patients' responses were by no means uniform, even under the most controlled conditions. Large differences in performance have been recorded between different patients and also in the same patient over a period of time, especially when comparing recent post-operative responses with those of a few years later, when a marked improvement has frequently been noted. It seems that in time some direct communication can be re-established between the hemispheres despite commissurotomy. I will return to this fact later on, but for the moment I will confine the discussion to cases that suggest, prima facie, that the operation has resulted in two separate minds in one body. In order to investigate these cases satisfactorally, we need to be acquainted with some physiological details concerning the ways in which sensory stimuli are transmitted to the normally functioning brain.

7.3 THE EXPERIMENTAL BACKGROUND

Tactile stimuli originating in a given side of the body are transmitted to the cerebral hemisphere on the other side (that is, contralaterally) with the exception of stimuli to the head and neck, which are directly transmitted to both hemispheres,
With visual stimuli (see figure 2) the retina is functionally subdivided in such a way that the left sides of both retinae, which scan

the right side of the visual field, send impulses to the left hemisphere, and conversely, the right half of both retinae, scanning the left half of the visual field, send impulses to the right hemisphere. Auditory impulses from either ear are transmitted to both hemispheres, with the signal to the contralateral hemisphere being the stronger. Olfactory impulses are only transmitted ipselaterally - that is, each nostril sends impulses to the hemisphere on its own side of the body.

Regarding motor control, each hemisphere controls the movements of the opposite side of the body, with the exception of the head and neck, which are controlled by both hemispheres together. Each hemisphere has some small degree of ipselateral control of the body, but this is negligible in comparison to the other's contralateral control. and is overruled by it in cases of normal brain function. However, this ipselateral control should not be dismissed as being an inevitably small and inadequate facility under all circumstances. It comes into its own in cases where the contralateral hemisphere has been removed or significantly damaged. Under such circumstances, the remaining hemisphere can be trained to develop greater and more refined ipselateral control. A similar development has been observed in some cases of commissurotomy, with a gradual but significant improvement in ipselateral control being developed in the years following the operation. In general, it appears that the brain's ability to adapt and take on roles and functions usually assigned to missing or inoperative substructures is at its strongest in children, tending to decrease with age.

For the vast majority of people, the areas of the brain that are responsible for the various aspects of linguistic abilities are located mainly in the left hemisphere. Thus, should someone suffer a stroke incapacitating this hemisphere, he will be left speechless and with vastly impaired linguistic comprehension, although in time he can learn to retrieve some of this function, to a degree, by means of the right hemisphere. Many left-handed people have their linguistic functions primarily located in the right hemisphere, and there is also a small percentage of people in whom the neural regions responsible for these functions are distributed fairly equally between both hemispheres. Incidentally, the hemisphere responsible for controlling linguistic functions is called the dominant hemisphere.

I will now go on to describe some common forms of experiments

that have been performed on split-brain patients, during which the controversial responses which have been the subject of so much philosophical debate were discovered

The patient is sat in front of a blank screen, onto which a tachistiscope flashes signals to a particular side of the screen. The patient is asked to stare at a spot right in the centre of the screen, ensuring that his head and neck movements are minimised. These procedures ensure that the signals are only directly accessible to one hemisphere. The signals are flashed for 200 milliseconds, which is just long enough to allow them to be registered by one half of the retinae, corresponding to one half of the visual field, but not long enough to allow any saccadic eye movements that would bring the signal within the range of the other half of the retinae and thereby deliver information to their associated hemisphere. These signals can take the form of words, simple images of familiar objects, or small patches of colour. In this last case, two colours are normally involved, two which are easily distinguishable from each other. (So, for example, red and green is an ideal coupling, as opposed to, say, red and orange). One each of these coloured patches is restricted to half the visual field, so that, for example, the left hemisphere has direct access to the red patch only, and the right hemisphere has access to the green patch only. The left hemisphere is not, and cannot be directly aware of the colour flashed to the right hemisphere, and vice versa.

In the majority of cases, where the left hemisphere is dominant, colours flashed to the right of the screen can be reported verbally, but not colours flashed to the left side of the screen, since the right hemisphere is mute. Similar results apply when words are flashed to the respective sides of the screen, and also in cases where objects are touched, unseen, by the hands . When the object is held in the right hand, the patient can verbally attest to his recognition of it, but not if it is in his left hand.

However, if a word, i.e. 'pencil'. is flashed to the left side of the screen, the left hand will pick out a pencil from amongst a pile of concealed objects, while the patient claims that he saw nothing. LIkewise, if two different words, corresponding to two different objects, are flashed onto a screen, one on either side of the mid-point, the two hands, hidden from the patient's sight by curtains, will search through a pile of unseen objects, seemingly oblivious to

the other and its search, and unconcerned with the object of the other's search, until they finally locate their respective targets.

Another set of experiments, by Levy, Trevarthan and Sperry [1972], used composite or 'chimeric' imagery : While the patient continues to look straight ahead at the mid-point of the screen as before, a composite image of two half-faces (the left side of one face and the right side of another) was flashed upon the vertical axis of the mid-point of the screen,so that each hemisphere only had direct access to one half-face. When asked to verbally indicate what image was seen, patients would report seeing the one sent to the left hemisphere, but when asked to point to the one that was seen, they chose the face sent to the right hemisphere. One further, unexplained oddity was that in both cases, the patient would report seeing whole faces - somehow each hemisphere managed to 'complete' the partial image into the more familiar whole form.

7.4 MINDS, BRAINS, PERSONS

As Wilkes [1978] argues, when theorists say that these results show that split-brain patients have two distinct minds or centres of consciousness, they are assuming that to deny this interpretation would involve a breach in the Law of Non-Contradiction - i.e. that the same person at time t both knows and does not know that p, or both sees and does not see q, so that the person both knows and does not know that the word 'pencil' was flashed within his field of vision. Given that the experimental results seem to provide cases where, say, the left hemisphere knows that p and simultaneously that the right hemisphere does not know that p, such theorists conclude that the left and right hemispheres must logically constitute two minds, two subjects of experience.

Certainly we feel strongly tempted to say that such patients have two of something... but to go further than that raises problems. As Patricia Churchland [1986] says, our old, commonsense 'folk-psychological' concepts such as 'mind', 'self', 'centre of consciousness' are so vague and theoretically undefined that we don't have clear rules as to how to individuate them, as we don't [illegible] know exactly what we are counting. As with other conditions that I discuss later, such as multiple personality, alexia, and visual agnosia, these pre-scientific mentalist concepts are unable to provide an adequate

explanation, or even description, of such conditions. However, it is instructive to draw out the confusions that are an inevitable consequence of such attempts, and therefore to question the privilege of some of our deepest assumptions regarding the mind.

As I elaborate in the following chapter, a single mind is attributed to someone not as some mysterious extra entity put forward to explain the otherwise baffling coherence of one's mental life, but merely as a mark that such coherence holds to a significant degree. However, when it comes to specifying precisely the minimum degree of coherence required, we soon realise that there is no black-and-white dividing line, due to the vagueness in the application of the concept. However, we are still able to offer paradigm cases that would most definitely fall under the heading of 'one mind', and those that are more problematic. As I argue at length, 'unity of mind' is always a matter of degree, being a mark of the amount of integration of the set of one's mental contents. Before returning to the main body of experimental results regarding split-brain patients, I wish briefly to comment on a couple of side-issues concerning forms of mental division that are not sufficient to warrant the ascription of two minds to one body.

One of the peculiar side-effects of commissurotomy that was occasionally displayed in everyday non-experimental settings was the simultaneous manifestation of two sharply conflicting emotions, with the two dissociated hemispheres seeming to be in opposition. This was exhibited in the much-documented case of the man who both embraced his wife with one hand and pushed her away with the other. However, it is often conveniently ignored by philosophers seeking to use this case that the patient in question had extensive damage to his right hemisphere, and thus any results concerning him cannot be employed to derive conclusions about all split-brain patients, as his peculiar behaviour may be due to this brain damage rather than to the commissurotomy. Regarding this patient, Gillett [1986] remarks that *"it would be implausible, on a moment's reflection, to claim that one of his hemispheres loved his wife and the other one hated her"(p227).* Gillett makes this assertion not on empirical grounds, but because he thinks that to make such a claim involves a category mistake. But this misses the point, as such a claim is no more peculiar than to say, in the case of someone with an intact cortex that 'his brain loves his wife'. To be sure, we don't commonly ascribe emotions, or propositional attitudes, to brains or half-brains, but to persons.

However, as I've previously argued, given the inadequacy of our commomsense mentalistic concepts in dealing with such anomalous cases, we can make such ascriptions in the case of split-brain patients, due to the fact that we lack any fully appropriate ways of describing the state of affairs, and this is as good a way as any, and better than most, and should therefore be regarded as a 'caretaker' form of expression, in the absence of a better one.

The mode of expression that Gillett derides has the merit of highlighting the obvious emotional conflict. However, we all experience conflicting emotions simultaneously without our mental unity being called into question. We can all feel attracted to and repelled by something or someone at the same time. In fact, often this apparent conflict does not involve a contradictory state of affairs, it being elliptical for a pair of non-mutually-exclusive wants. One can only speculate in this particular instance, but perhaps the man had two different reactions to two different factors relating to his wife, in which case the only difference between his attitudes and our own is the peculiar and extreme way in which his internal conflict was manifested.

I will now turn to another example, relating to a further aspect of simultaneous knowledge and absence of knowledge of certain information within the same person, which again applies to all of us, with normal brains. My example concerns an experience of my own. In fact, I can remember having had two such experiences, and, assuming that I am perfectly healthy, with no undetected neurological damage or peculiarity, I see no reason to doubt that such experiences are not unique occurences, but merely that they are rarely noticed or appreciated as having any significance. One morning I had to make a phonecall, and I was in a hurry. I located the number in the phone-book, closed the book, and began dialling. Once I had reached around the third digit of the number, I said to myself, 'Now what was that number again?' and found that I could not recollect it, or, more precisely, I found that I could not inwardly, sub-vocally, recite the number to myself. Nor could I form a mental picture of the number. Yet throughout my cogitations, my finger (my right index finger, incidentally) kept on dialling 'as if it knew what it was doing'. Over and above my puzzlement regarding this peculiar state of affairs, I felt confident that I had somehow dialled the correct number. I felt confident that I, in a sense, knew the number in question,

Although I could not manifest this knowledge by uttering the number, I could do so by the very act of dialling it - and it turned out that I had dialled the right number.

One obvious way of attempting to account for my experience is to dissolve the apparent contradiction by appealing to different types of knowledge, for example to the distinction between 'knowing how' and 'knowing that', where I knew the number in the former way, but not in the latter. While my example doesn't easily fit this distinction, it may well be that, given sufficiently careful rewording, a solution could be found by this device. It is all too tempting just to say 'my body knew but my mind didn't', but, no matter how phenomenologically accurate such a description seems, it is theoretically dubious, relying on a dualism that I don't want to encourage, as well as being devoid of any explanatory power, as it leaves the nature of these two forms of knowing totally opaque.

My example also reminds me of Wittgenstein's [1953] remarks on understanding and rule-following, to the effect that having a mental image before the mind is neither a necessary nor a sufficient condition of being able to follow a rule, or, in this case, a self-instruction. No doubt, given a fully satisfactory neurophysiology, we could describe in an impersonal way the processes by which both of my cognitive modes were realised, without reference to any single subject of experience, which would, of course, avoid any contradiction.

Having dealt with these couple of asides, I will now return to the main body of experimental findings regarding split-brain cases. One of the first things that any theorist must account for is the divergence between, on the one hand, the patients' seemingly perfectly integrated behaviour virtually all the time under normal conditions (which would seem to encourage the attribution of a single mind) with, on the other hand, the results of controlled experiments as I have described, which have been cited as grounds for saying that such a unity has been fundamentally split, and that we must therefore see such patients as having two separate minds.

7.5 PUCCETTI'S 'TWO PERSON' THEORY

There have been a wide range of responses to these experimental results. By far the most extreme has been from Roland Puccetti [1973], who sees the data as providing grounds for the claim that all human beings, including those of us with an intact cortex, have two minds, and, in fact, should be more accurately described as being two persons cohabiting a single body.

Puccetti agrees to a large extent with the arguments put forward by Sperry et al, to the effect that split-brain patients have two separate and distinct centres of consciousness, or 'minds'. (I will discuss Sperry's theory in detail shortly). However, he extends these arguments, saying that they support his position regarding all human beings. He offers several arguments to back up this truly bizarre claim, some of which are largely conceptual, and some of which are factually based. Firstly, he asks, *"How could commissurotomy create two minds or persons if there was just one before? Which mind - the left-based one or the right-based one - is brand new? And how are we to make a choice here? Both brains, as we have seen, were conscious and functioning in their rather specialised ways before the operation. It is just that they functioned more synchronously - because of the commissural connections - and no longer do so in test situations.... thus even in the normal cerebrally intact human being there must be two persons"(p351).* He is saying that the fact that our two 'minds' are connected by the corpus callosum will not significantly alter their functioning, except in making a little more information available to them.

He backs up his theory by pointing to well-documented cases of hemispherectomy, wherein *"the same personality, character traits, and long-term memory traces persist post-operatively. The only way I can see to explain this is to say that the same 'person' did not survive hemispherectomy at all. Because this former 'person' was never a unitary person at all. He or she was a compound of two persons who functioned in concert by trans-commissural exchange - what has survived is one of two very similar persons with roughly parallel memory traces, nearly synchronised emotional states, perceptual experiences, and so on, but different processing functions. (p352).*

Regarding his question of how two minds could be created out of one, apart from his claim, on neurophysiological grounds, that there

always were two minds present prior to the operation, he seems to be making a general conceptual point that one F cannot result in two Fs. This claim works in tandem with his usage of the word 'create', which suggests that he is making the mistake of reifying the mind (in a way that is incompatible with his wider stance), and asking how some thing (a new mind) can be made out of nothing. However, as Margolis [1975] says, *"Every analogy regarding the puzzles of numerical identity ..... makes it quite clear that one entity of a given kind may, under specifiable circumstances, yield two entities of that kind. There simply are no fixed rules regarding the individuation of things that would preclude such a possibility" (p277).*

Regarding hemispherectomy, this operation and its results does not point unequivocally to the conclusion that two minds existed prior to the operation. As Ritgerink [1980] says, *"Surely there is a big jump between saying that the two hemispheres are capable of operating as two minds and saying that this potential is being realised in either a normal human being or in a split-brain patient" (p442).*

In fact, as Marks [1986] argues, even if we grant for the moment that there exist two distinct minds following commissurotomy, this need not require there to have been two beforehand (whether we interpret Puccetti 'strongly' as saying that this state of affairs is implied, or 'Weakly' as offering the two minds hypothesis by inference to the best explanation), since all that is required is that there be the potential for two distinct minds within our one presently integrated mind. And given the high degree of redundancy in the neural structures embodying such a mind, we should not find such a result too surprising.

James Moor [1982] offers this analogy for hemispherectomy : imagine a large tree which forks into two equally large branches, which are thus fairly independent of each other. Clearly this is one tree, and if one of these branches were chopped off in full, one and the same tree would clearly remain. If one follows the analogy through, the branches, representing the two cerebral hemispheres, are connected to a common trunk, which represents the brain-stem, spinal cord, etc..... but I will not push this any further at the moment, because it would expose the crucial flaw in Puccetti's argument, a flaw that he shares with Sperry, and I want to withhold

this argument until I have discussed Sperry's case in detail.

But before I leave Puccetti and his claim that all of us possess two minds, I must say that his criterion for the individuation of minds is not clear to me. He would surely accept that information is directly exchanged between the hemispheres by means of the corpus callosum, and, in virtue of that fact, the cerebral cortex functions as one integrated system (within a larger system). He must then acknowledge that a person can have an overall integrative awareness of simultaneously having two mental contents (i.e. thinking about supper while listening to music) whose neural embodiments are spread out over areas including both hemispheres. Given this, it is hard to see how he can individuate two distinct minds without begging the question by reifying minds on a hemispherical model.

Another position that we can quickly dispose of is the false compromise that attempts to make divergent results compatible by suggesting that split-brain patients have one mind under normal conditions and two minds under experimental conditions. This is an entirely ad hoc move, with no explanatory force, and encouraging no independent grounds for support. It also has the untenable implication that the number of minds present is determined by something completely external to the patient, contravening what Brennan [1988] has called the 'only x and y' rule. Also, as Nagel says, *"there is nothing about the experimental situation that might be expected to produce a fundamental internal change in the patient. In fact it produces no anatomical changes and merely elicits a noteworthy set of symptoms. So unusual an event as a mind's popping in and out of existence would have to be explained by something more than its explanatory convenience" (p161).*

7.6 COGNITION IN THE RIGHT HEMISPHERE

Nagel is likely to be able to sustain a telling argument to the effect that the right hemisphere does not deserve to be regarded as a mind because the left is dominant. Such a position is held by Sir John Eccles [1977], who writes of *"the uniqueness and exclusiveness of the dominant hemisphere in respect of conscious experience".* Of course, we could sidestep such a challenge by restricting the argument to the previously mentioned cases where the language-related functions are shared by both hemispheres. But such a manoeuvre is unnecessary, as I can easily show that Eccles' argument

is/on very weak grounds. It is perfectly obvious that the inability to produce linguistic signs is no proof of the absence of conscious experience, as to make this claim would be to reduce pre-verbal children to the level of automata. Likewise, as Puccetti observes, *"if speech is a necessary condition of consciousness, then the aphasic who can play piano - as Ravel did - is playing unconsciously" (p342).* Also, one need only become acquainted with the modes of cognition associated with the right hemisphere, and their degree of complexity, to see the untenability of Eccles' view. As Nagel says, *"There seems no reason in principle to regard verbalisability as a necessary condition of consciousness... what the right hemisphere can do on its own is too elaborate, too intentionally directed and too psychologically intelligible to be regarded merely as a collection of unconscious responses... the right hemisphere follows instructions, integrates tactile, auditory and visual stimuli, and does most of the things a good mind should do" (p156).*

If one is still in any doubt over the legitimacy of calling the right hemisphere conscious, one need only be reminded of its ability to adapt and take over functions - including language related, to a degree - that had been in the province of the left hemisphere, in cases where this latter hemisphere has been removed or is largely incapacitated. It has been suggested that in normal conditions, the left hemisphere inhibits the right's latent linguistic abilities, but in cases of strokes or of left-hemispherectomy these effects are removed. The two cases of left-hemispherectomy that I am aware of showed a gradual marked increase in linguistic skills, but both patients died before the extent of the remaining hemisphere's full linguistic capacities could be ascertained.

There have been numerous studies published, both popular and academic, describing the different but complementary modes of cognition typical of the two hemispheres. For some light relief, I reproduce one of the most extensive, by Joseph Bogen (in Campbell [1982], a pioneer of split-brain research.

LEFT	RIGHT
intellect	intuition
convergent	divergent
intellectual	sensuous
deductive	imaginative
active	receptive

discrete	continuous
abstract	concrete
realistic	impulsive
propositional	imaginative
transformational	associative
lineal	non-lineal
historical	timeless
explicit	tacit
objective	subjective

Looking at this list of complementary-opposite pairs, I am inclined to suspect that Dr. Bogen received a Thesaurus for Christmas, as it repeats and elaborates on a couple of vague distinctions by means of near-synonyms and associated concepts. Rather than talking in such terms, it is perhaps safer to stick to descriptions of what the respective hemispheres can actually do.

So Eccles' claim that the right hemisphere is not a subject of conscious experience most certainly underestimates its abilities. However, in fairness to him, his view was once widely held, and seemed somewhat plausible, but in more recent times some experimental results have suggested that the right hemisphere has more language-related abilities than had been previously reckoned. In the 1960's, various tests by Gazzaniga, Bogen, and Sperry showed that in split-brain patients, the right hemisphere could not initiate a verbal response to questions regarding sensory stimuli that it alone was exposed to. This left it open whether the hemisphere merely lacked motor control to allow speech, or whether there was a more general lack of capacity for linguistic understanding. Later tests indicated that it had some albeit limited linguistic comprehension, for example it could match simple concrete nouns with pictures, but could not for complex nouns, or for verbs. Also, it seemed to be oblivious to synctactical distinctions. And when a verbal command was given, the left hand could retrieve the relevant object, whether it was asked for directly, or by means of clues, such as 'What do monkeys eat a lot of?', in response to which a banana was retrieved.

In the mid-seventies, Zaidel [1975] devised a set of experiments to see whether the right hemisphere's linguistic comprehension improved if images of words were presented for longer time intervals than tachistoscopic techniques allowed, by designing a 'Z-lens', a special type of contact lens which blocked light to the left-half

retinae. Given more time to scan the words, the right hemisphere displayed what Zaidel estimated to be the level of linguistic comprehension of a 10 year old - vast improvement on previous findings. This led theorists to consider whether the question of the lateralisation of linguistic functions was not more complex than was previously realised, with the right hemisphere being capable of significant contributions to at least some of the aspects concerned. However, as before, there is a frustrating lack of unity in experimental results. Zaidel's estimate was based on the performance of just two patients, with another four showing negligible right-hemispherical linguistic comprehension. Likewise, in Gazzaniga's earlier experiments, only three out of twenty eight showed any significant level of linguistic ability. All this indicates that research still has a long way to go before definitive theories can be presented.

Larry DeWitt [1975] attempts a softer version of Eccles' theory, in that he allows that the right hemisphere is conscious, and even that it qualifies as a mind, but that it lacks that particular reflexive consciousness that is required for it to count as a self. In other words, it lacks self-consciousness, the ability to think of oneself as oneself. As he elucidates this tripartite distinction : *"at the lowest level, we have 'consciousness', the basic phenomenal awareness that accompanies acts of perceptions, emotions, sensations: in order to possess a 'mind', the organism must possess a hierarchical ordering of behavioural priorities which are consciously utilised in intentional actions;.... by 'consciousness of self', I mean the ability to apprehend oneself as being distinct from other similar beings, to recognise one's actions and thoughts as belonging in some sense to oneself"(p42).*

DeWitt says that the right hemisphere lacks this self-consciousness, which he links with the ability to use language. He cites Gallup's experiments in presenting chimps with their mirror-images, whereupon they gradually cease to regard such images as being of other animals, but rather to be somehow associated with themselves, which would seem to require the application of a self-conception at a rudimentary level. DeWitt sees the significance of such a test in that it provides a means for checking for the existence of self-consciousness without the need for verbal communication, and he conjectures that a similar test would prove that the human right hemisphere lacks self-consciousness even to this paltry degree.

Puccetti [1975] replies that DeWitt, in saying that the right hemisphere lacks all linguistic abilities, is making the same mistake as did Eccles, in running together firstly the ability to evoke verbal responses, with secondly the ability to understand language, and argues that the right hemisphere does not lack this second ability, as is shown by Zaidel's tests, just described. Puccetti also cites Sperry's experiments whereby the right hemisphere responds to a verbal request, such as 'draw the figure you see with your left hand' where the image of a dollar sign is restricted to the right hemisphere and the image of a question mark is restricted to the left hemisphere. The patient is asked to draw what he sees, with his left hand (which, alongside the paper, is hidden from his view) and, as he draws, he is asked to say what he is drawing. The patient said that he was drawing a question mark whereas he was drawing a dollar sign.

Puccetti takes up DeWitt's experimental challenge regarding the right hemisphere's linguistic abilities, saying that an experiment could be devised wherein tachistoscopic images could be flashed exclusively to the right hemisphere, and the patient is asked to press a button with his left hand upon seeing images of his own face. Puccetti conjectures that the right hemisphere would actually do <u>better</u> at this than would the left hemisphere, since the right is known to be superior in facial recognition, and since failure to recognise faces, including one's own, is always found to be due to a lesion in the <u>right</u> parietal lobe. In fact, such experiments <u>were</u> performed by Sperry and Zaidel, and, while the results vindicated Puccetti to some extent, they were inconclusive. Basically, when the patient's own image was flashed to the right hemisphere, there was no immediate recognition - it only came gradually, i.e. 'I don't know..... yes... it might be me... it is me... yes, definately, it's me'. The researchers put this down to the assistance of verbal signals from the left hemisphere, and so the results indicate that while the right hemisphere has <u>some</u> sense of self, the extent cannot be assessed, nor how it compares to that of the left hemisphere. A more crucial reservation is that such results show merely that the right hemisphere recognises <u>a</u> self, and not that it recognises itself as itself as distinct from the left hemisphere.

7.7 SPERRY'S 'TWO MIND' THEORY

I will now turn to what I regard as a more credible view, associated originally with Sperry, namely that split-brain patients always have two distinct centres of consciousness, both in everyday life and under experimental conditions, in a way that the rest of us having an intact cerebral cortex do not. The most obvious issue that this theory must deal with is the apparently high degree of integrated behaviour exhibited by such patients under normal circumstances. In fact, this can be accounted for with no great difficulty. Take the case of vision, the sensory modality that most controlled experiments have focussed on. Normal saccadic eye movement causes an image to be registered on both halves of the retina within around 250 milliseconds. thus enabling both cerebral hemispheres to gain direct access to the same stimuli. This, plus normal head and neck movement, can explain how both hemispheres will gather virtually identical information about relatively stable (i.e. all but the instantaneous) features of the environment. Thus, under normal circumstances, split-brain patients will exhibit nothing that would indicate a split or a plurality in consciousness.

But this does not refute the claim that such people have a continuously divided consciousness, as the conclusion that follows more naturally from the above-mentioned facts is not that such patients have one unified stream of conscious experiences, but rather that they have two parallel streams, due to the independent duplication of sensory stimuli, with, to quote Sperry [1968], *"no direct causal connection between the corresponding neural representations - rather, there are separate independent causal pathways from the sensory receptors to each hemisphere" (p297).*

Sperry further points out that each hemisphere has indirect access to the other's contents by means of 'cross-cuing' strategies (which are sometimes referred to, rather question-beggingly, as 'self-cuing'). Cross-cuing is a means whereby one hemisphere uses information deriving from behavioural responses originating from the other hemisphere. The phenomenon was illustrated in experiments by Gazzaniga [1970] designed to test whether the right hemisphere could verbally identify colour stimuli. Under controlled conditions where the image was restricted to the right hemisphere, the patient was asked to guess whether red or green had been flashed, and, since it was the left hemisphere that was initiating the talking whilst being blind to

the stimuli, the patient had, it was assumed, a 50% chance of guessiing correctly. However, the researchers noted that a significantly higher rate of success was achieved if the initial guess was allowed to be revised, without the patient being told whether it was correct or not, and a second guess permitted. Yet they were satisfied that they had set up the experimental conditions to eliminate the possibility of the left hemisphere gaining direct access to the stimuli. The clue to what was going on came when they noticed that the patient seemed to know that his initial guess was wrong. Thus, if red was flashed, and the first guess was correct, the patient would be content to stick with this answer, whereas if he guessed that it was green, he would immediately react, commonly with a frown or a shake of the head, saying 'No, I mean't red'. What was happening was that since auditory impulses are transmitted to both hemispheres, and since both hemispheres can activate movements of the head and neck, the right hemisphere, having heard the left hemisphere answer 'green', and having recognised this as the wrong answer, would react with a disapproving gesture, which would be recognised and interpreted by the left hemisphere, which would infer that a mistake had been made, and suitably revise the answer, which would then be greeted with a smile.

Similarly, if an object is placed, out of sight, in a patient's left hand, and he is asked to identify it, he will more than likely guess wrongly, since there is only a small and primitive form of ipselateral recognition of tactile stimuli, and of course the dominant hemisphere has no direct access to the object. However, as in the previous case, the right hemisphere will hear the left's guess, and will make a frown, etc., which will act as a cue to the left hemisphere to revise the guess.

However, the implementation of information deriving from cross-cuing strategies does not constitute a proof of the existence of merely a single centre of consciousness, since the situation corresponds to the way in which we come to know the contents of another person's mind. That is, we have no direct access, but we come to it inferentially, albeit unconsciously, just as in the case of cross-cuing. In the case of split-brain patients, one hemisphere does not have access to the other's mental contents in the same way that it does to its own. It knows its own contents, or at least its own conscious contents, directly and non-inferentially, whereas its knowledge of the other hemisphere's contents is of the same epistemological status as is it's knowledge of the contents of other minds.

7.8 THE SUBJECTIVE VIEW

I want now to turn briefly from my main line of argument, and examine the 'subjective' aspect of split-brain surgery - the 'how it feels like from the inside', the 'what it's like'. The left hemisphere does not disown actions initiated by the right. Rather, it acts from a unitary perspective, confabulating, and integrating such an action into a comprehensible schema by means of plausible explanations, as if the intention and motive were its own. This tendency to gloss over paradox and the fundamental disruption of deep assumptions seems to be irresistable - the integrative impulse persists even in cases of patients who are relatively informed regarding the effects of commissurotomy. For example, Gazzaniga and Ledoux [1978] flashed the image of a snowy scene to the right hemisphere and simultaneously flashed the image of a chicken claw to the left hemisphere, and then showed each a sequence of pictures, asking each to select, by pointing, the picture that 'matched' the observed image best. The left hand indicated a shovel as linked with the snowy scene, and the right hand indicated a chicken head to go with the claw, justifying the choices as 'the checken claw goes with the chicken and you need a shovel to clean out the chicken shed'. Thus, the left hemisphere justifies the left-hand choice by integrating it into its own explanatory schema.

Gillett [1986] emphasises this integrative impulse, which he extends to circumstances in which the patient is conscious of a conflict which he then tries to overcome by various means (although I reckon that he is wrong in saying that such processes as cross-cuing and confabulation are conscious)... *"Because they do try to reintegrate their information, or make best use of their disrupted brain function in tackling the tasks they are set, they can properly be said to be struggling with certain confusions to which they find themselves subject rather than to have become two mutually independent streams of consciousness which are in a no more than contingent relation to each other" (p227).*

I disagree with Gillett here, for reasons that I will state when I consider Charles Marks' criticisms of Sperry. But for the moment, it is worth asking why it should be assumed that the patient's own subjective view of his situation has any special validity. It may just be a fact about the commissurotomised brain that it cannot help struggling towards greater integration of experience and action, and regarding itself as a single centre of consciousness, but may be

radically misled as to the real situation, like in the case of amputees who continue to suffer pains in 'phantom limbs'. Once we step outside the language of the neurosciences, we are at a loss to adequately describe what is going on, and there is no reason to assume that this confusion is less marked from the first-person perspective. Likewise, when Nagel points out that others meeting these patients find it natural to relate to them as single individuals, one is inclined to wonder just what the natural response would be if they regarded them as having two distinct minds. Anyhow, in place of 'natural' here, perhaps we should read 'entrenched', and, as I've said, the mere fact of its practice is no necessary indication of its correctness or appropriateness.

7.9 SPERRY CHALLENGED

Returning to Sperry's theory, I will now discuss his use of the central concept of 'unity of consciousness'. To Sperry, the criterion of a single unified consciousness is that the set of neural states coonstituting the physical realisation of this centre of consciousness be directly causally connected and integrated. This might be better expressed by saying that the physical basis must be such that it enables members of this set of neural events (which will, from the inside, realise token sensations and propositional attitudes) to be directly - that is, without some external intermediary - causally connected, so that one can effect the other, or so that they can be co-experienced. By this criterion, he argues, in cases of split-brain surgery, the left and right hemispheres cannot together constitute the physical basis of one unified consciousness, because the cutting of the corpus callosum removes the physical basis of the direct connecting system between them, and thus no direct causal relations can exist between two neural events where one is located in one hemisphere and one in the other - they can only be connected indirectly via inference from external cues.

This position is a strong one, but it is not conclusive. A few facts are worth considering :
Firstly, the nature of the communication between the hemispheres is still to some extent a matter of conjecture, as indeed is the very nature of information storage and retrieval in the brain. It is very tempting to think of the corpus callosum in terms of crude analogies such as with a form of telegraph network, relaying messages from place to place - but such models will more than likely be highly

misleading, and it may well be that at present we lack the appropriate concepts with which to give a more accurate picture.
Secondly, it is well-established that in infancy there is no traffic of information between the two hemispheres, as the corpus callosum only becomes functional at around two years of age, and is only fully operative at around ten years. It is assumed that this slowly evolving process allows the two hemispheres to develop their different modes of cognition with a fair degree of independence.
Thirdly, and even more strikingly, consider people born with asymptomatic agenesis of the corpus callosum - that is, who are born without a corpus callosum and never develop one, yet in whom no unusual behavioural nor cognitive responses are discovered, neither in normal circumstances nor, crucially, when exposed to the same tests that revealed the tell-tale responses in split-brain patients. They responded to such experiments exactly like someone with an intact and functional corpus callosum, and therefore we have absolutely no grounds for the view that such persons possess a 'divided consciousness' or 'two minds'.

But how can this be so if there is no physical basis for one unified consciousness? The answer is that it cannot be so - in other words, given that such people show every indication of having one unified consciousness, there must be some physical means whereby this is achieved - but obviously a different one. Neurologists reckon that the only viable explanation is that certain minor commissures have taken on the functions normally carried out by the corpus callosum. This is not considered to be peculiar, since, from the point of view of evolutionary fitness, it is to be expected that the structure of the brain should allow both for some degree of adaptability, and have a certain degree of redundancy, and we have already seen strong evidence for redundancy in brain structure with the independent bipolar duplication of sensory information, and, as I've just mentioned, the fairly late development of a fully-functioning corpus callosum.

In fact it is now believed that the transfer of information across the corpus callosum is highly redundant in the everyday functioning of normal brains, and that such information is largely filtered out by some information-integrating system within the cortex, and thus it plays little direct role in the integration of mental life. If this is the case, it appears that the integration of behaviour in normal brain functioning is, to a very large degree, carried out by the same integrating mechanisms as in the case of

split-brain patients, namely bipolar independent duplication of information. Marks [1986] takes this fact and argues that by Sperry's criterion of the unity of consciousness, we all, virtually all the time, fail to satisfy this condition, and therefore all of us have a disunified consciousness, a conclusion that Sperry would not want to accept.

This is a strong attack, but not conclusive. To remind ourselves, Sperry's criterion states that a 'mind' or centre of consciousness counts as single and unified if the neural processes constituting the physical basis for that consciousness can be directly causally related. The usual way that the hemispheres are connected is by the corpus callosum, which enables direct causal connections to hold between the contents of the cerebral cortex. The fact that this function is neither as crucial nor as extensive as was once thought does not negate the existence of the function itself. Even if most integration of behaviour is achieved via the independent bipolar duplication of sensory stimuli, it remains the case that some integration of behaviour is achieved by means of the corpus callosum. The crucial point of Sperry's criterion is that it is necessary that there be some physical basis that allows direct causal connections between processes constituting the 'centre of consciousness' in question. It is beside the point that the integration of behaviour and experience can also be achieved by other means.

It would also be wrong to suggest that the phenomenon of asymptomatic agenesis of the corpus callosum can provide a refutation of Sperry's theory. This would only be the case if possession of an intact and well-functioning corpus callosum was a necessary condition of having a unified consciousness , but it is not, given that people with agenesis of the corpus callosum display no indication of having a split in consciousness even under the most rigorous of tests, indeed under the paradigmatic tests used to reveal such a split. Sperry's criterion merely demands that there be some physical basis for the unity of consciousness - it doesn't matter what form this takes, as it is a functional requirement, and as such it cannot be reduced to any one physical specification.

Charles Marks uses the fact of the redundancy of much of the information transferred between the hemispheres via the corpus callosum to suggest that perhaps the independent duplication of sensory information can suffice to unite consciousness, and that

bilateral neural representation is a sufficient physical basis for such a unity. To back up his case, he suggests a weakening of Sperry's criterion, to the effect that it is not necessary that neural processes which together combine to realise a mental state be directly causally connected, but rather that it is sufficient that causally unrelated neural processes jointly, but separately, produce effects that realise a mental state.

My objections to this argument are firstly that it makes the fact of whether a consciousness is unified or divided, singular or plural, depend on factors outside itself and its physical basis, because it remains the case that experimental conditions can be set up whereby the independent duplication of sensory stimuli is prevented, thereby creating responses that are indicative of a major split in consciousness, where, in Nagel's words, *"there appear to be things happening simultaneously which cannot fit into a single mind : simultaneous attention to two incompatible tasks" (p160).* Nagel himself suggests a criterion for the unity of consciousness that is similar to, and compatible with Sperry's : *"Roughly, we assume that a single mind has sufficiently immediate access to its conscious states that, for elements of experience or other mental events occurring simultaneously or in close temporal proximity, the mind which is their subject can also experience the simpler relations between them if it attends to them. Thus we assume that when a single person has two visual impressions, he can usually also experience the sameness or difference of their coloration...(etc)"(p160).* Clearly a criterion for the unity of mind cannot be so vulnerable, so contingent on external circumstances, as I have shown Marks' to be.

My second objection to Marks' criterion is that it runs together two totally different concepts which need to be distinguished, namely (i) a unified single consciousness, and (ii) two qualitatively identical but numerically distinct centres of consciousness which, as it happens (given certain conditions, albeit almost always present) are running along parallel lines.

I will now return to the main body of experimental findings, and use it to construct an ideal experiment which goes to the heart of the matter regarding the issue of mental unity in split-brain patients, and which I regard as forming the strongest argument against the view that split-brain patients possess a single centre of consciousness. Contrary to Nagel, I do not believe that we have hard

and fast criteria to tell us when certain mental activities are mutually exclusive and which thus cannot be simultaneously entertained in a single mind.... but the following argument is designed to display a very strong contender for such a pair of activities. The experiment takes the usual form of a test to establish that, under the previously-described experimental conditions as devised by Sperry, Gazzaniga & Co., when patches of red and green are flashed simultaneously to opposite sides of the patient's visual field, then each hemisphere will be aware of one each of the colours, but not of the other. I will spell it out :

1. Take a situation as described above, where both hemispheres perceive and correctly identify their respective colour patches.
2. Red and green are logical contraries, so that nothing can be both all-red and all-green at the same time.
3. From this, we can say that one subject of experience cannot perceive only red and only green at the same time.
4. However, it is logically possible for a subject A to see only red at time t, and for a subject B to see only green at t, if A $\neq$ B.
5. Given that red and green were perceived simultaneously, it follows that A $\neq$ B. In other words, the left and right hemispheres must be separate and distinct subjects of experience at t.

7.10 DEGREES OF MENTAL UNITY

It may well appear from the past few pages as if I am advocating a position similar to that of Sperry in this debate, namely that split-brain patients always have a 'disunified consciousness' or 'two minds', whereas those of us with an intact cerebral cortex have a single unified consciousness. I do not hold this position. Rather, I maintain that all of the major positions that I have discussed, including those of Puccetti, Eccles, Sperry, and Marks. each share the common fundamental error of regarding the question of the 'unity of mind' and other equivalent expressions as being a determinate, 'all or nothing' matter. It is this mistake that allows them to regard 'minds' themselves as coming in discrete units or quanta.

As opposed to this, my view is that unity of mind is in all cases a matter of degree, and that firstly, our everyday concept of a 'single mind' can allow a substantial degree of disunity, and secondly that there is no all-or-nothing difference regarding mental unity

between ourselves and split-brain patients - our mental unity or disunity just holds to different degrees.

On this first point, I am in substantial agreement with Margolis [1975] when he says that *"the so-called unity of minds and persons is designed to accommodate all sorts of anomalies - for instance self-deception, contradictory beliefs, aphasias, loss of memory, compulsions, ignorance about one's motives and intentions, dreaming and sleep-walking, the subconscious, schizophrenia and dissociative personality" (p279).* I will discuss some of these conditions in the following chapter, where I will argue that Multiple Personality Disorder is the exception to the rule.

On my second point, let us go back to my description of the various forms of commissurotomy - complete, central and frontal (see ch.7.2) - and remember that the extent of the operation required is gradually becoming less and less due to our ability to locate the source of epileptic seizures. The important point here is that the degree of psychological disunity displayed by such patients is proportionate to the amount of surgery performed - so we do not have an all-or-nothing distinction between 'normal' and 'split' brains, but rather, in Moor's words, a 'split-brain spectrum', with varying degrees of mental disunity.

However, even in a 'far-end' case in this spectrum, involving complete commissurotomy, it would be a mistake to characterise the cerebral hemispheres as constituting two completely distinct and independently-functioning minds. To see why, let us return to the 'two-mind' theorists' account of why, if both hemispheres are conscious, and capable of independent thought and volition, there are not frequent and visible conflicts between them. Puccetti [1973] gives a clear answer to this question, and one with which Sperry would be in full agreement : not only do the two divided hemispheres have a history of almost identical experience and access to sensory stimuli, but also *"common sub-routines of learned behaviour stored in the still-intact cerebellum, and a common internal milieu (blood sugar, hormones, etc)... same autonomic, humoral and muscular reactions are shared, either by peripheral sensory feedback or via the intact brain stem, the shared vascular system, cerebrospinal fluid, and so on. Since primary drives are mediated at subcortical levels, it is no surprise that both hemispheres feel hungry, thirsty, lustful, or what have you, at the same time. Even in test conditions an emotional*

reaction gets into both hemispheres"(p343).

Quite right, but surely all of this can be used against a 'two-minds' theory. As Ritgerink [1980] says, *"Surely motivational drives and emotive responses are an integral part of any human mind. If these experiences cannot occur without the operations of the sub-cortical areas of the brain, then these regions should be included as parts of the same mind"*(p442). So, since the brain stem and other sub-cortical structures are not sectioned, then both alleged 'minds' share common parts, and therefore are not distinct and separate. Thus we see the strength of Moor's analogy of the tree which forked into two main branches.

It was forgotten by the 'two-mind' theorists that despite the fact that the cerebral cortex can, with a fair degree of accuracy, be called the 'seat' of conscious experience, the brain stem constitutes a necessary requirement for such experience to take place. To quote Pallis [1983], *"The reticular formation forms the central core of the brain stem and projects to wide areas of the limbic system and neo-cortex. Projections from the upper part of the brain stem are responsible for alerting mechanisms. These can be thought of as the capacity for generating consciousness. The content of consciousness (what a person knows, thinks, feels) is a function of activated cerebral hemispheres. But unless there is a functioning brain stem 'switching on' the hemispheres, one cannot speak of such a content" (p7).* Penfield [in Eccles 1970] is equally explicit : *"Consciousness continues, regardless of what area of cerebral cortex is removed. On the other hand, consciousness is inevitably lost when the function of the higher brain stem (diencephalon) is interrupted by injury, pressure, disease, or local epileptic discharge"(p234).*

Let us now return to my 'side-issues' of ch.7.4, which now turn out to be more relevant than at first they appeared. You will recall that I described my state of simultaneously both knowing and not knowing a phone number. This was acknowledged as a form of mental disunity that does not prompt us to ascribe a plurality of minds. However, the issue that I want to take up now derives from my remark then, almost in passing, that an account of this state of affairs could possibly be given in the language of neurology in such a way as to avoid the contradiction and incomprehensibility resulting from the mentalistic, intentional description. My point is that this can also be done regarding descriptions of the results of split-brain

research.

When these results were described in mentalistic terms (i.e. the person both knew and didn't know that p, both saw and didn't see q; saw both q and r simultaneously when q and r form a mutually exclusive pair, etc.,), it seemed that we could only avoid paradox at the expense of giving up the notion of a single unified subject of experience, and this, of course, was the source of all the philosophical perplexity. However, when we restrict our description to the language of the neurosciences, the paradox disappears, along with the philosophical problems. Problems only re-emerge when we attempt to translate this description into the commonsense vocabulary of folk psychology. But (as I discuss in the following chapter), like cases of self-deception and Multiple Personality Disorder, such phenomena associated with split-brain patients are unintelligible within this framework, as, to quote Wilkes [1988], *"We are, hopelessly, trying to show how irrational or non-rational behaviour is, after all, rational. This is all we can do with the commonsense vocabulary of psychological terms, since that is what it is for" (p159).* In other words, the folk-psychological framework presupposes a form of unity of mind that is absent in these cases.

Wilkes compares split-brain cases with those involving pure alexia where, for example, someone who can write clearly and fluently cannot then read what he has written. Such a situation is not intelligible within the folk-psychological framework, which takes it for granted that the two abilities are always correlated. However, again the situation is explicable at the neurological level. (Very roughly, there is damage to the left visual cortex and its associated area of the corpus callosum, so that there is no direct communication between the language centres in the left hemisphere and the intact right visual cortex). As Gardner [1975] says, *Here the study of brain damage has helped us dissect a skill into its component parts; a capacity thought to be unitary - the reading of symbols - is shown to be divisible into separate functions"(p16).* It follows that, to quote Wilkes again, *"pure alexia is no longer philosophically puzzling, as the 'rational' explanation has been superceded.... we cannot and should not expect an explanatory apparatus, developed precisely to explain what is rational and sensible about human purposive action, to cope with the non-rational, the irrational, the not-sensible. It must be supplemented or superceded"(p161).* Such brain disorders are legion. I refer the reader to Sacks [1986] for a selection.

So what do we say in conclusion? We accept that some disunity of mind is inevitable, and that this holds to varying degrees, beyond which our folk-psychological framework leads to paradox. When the disunity takes place against a backdrop of general integration, such as in cases of split-brain surgery, we should maintain that there is one person present, with one mind, albeit a significantly disunited one. Since mental unity/disunity is not an all-or-nothing matter, we cannot derive a plurality of minds from the existence of disunity. James Moor again offers an excellent analogy : *"I believe that the concept of a single person with a disunified consciousness will become less strange as science advances. There is a rough analogy with our notion of an atom. For a long time it was commonsense that atoms were indivisible. Now we regard them as complex systems.... In a similar way, I think we should understand persons as complex systems - even systems that can be somewhat disconnected"(p104).*

7.11 Appendix : 'MY PHYSICS EXAM'

I will end this chapter with some reflections on Parfit's [1984] thought-experiment entitled 'My Physics Exam', an imaginative application of having a divided consciousness like a split-brain patient. I will follow Parfit in telling the story in the first person :

I am one of the small percentage of the population whose linguistic functions are shared more or less equally by both hemispheres. I am also in possession of a device whereby, by activating it, I can block all direct communication between my hemispheres by raising my eyebrows, and reunite them again by lowering each eyebrow. I am sitting a physics exam, and realise that I only have fifteen minutes left with which to answer a question, but there are two possible strategies to adopt in answering it, and I don't know which one is the more appropriate, and I don't have time to attempt both. So instead I 'divide my mind', assigning one strategy to each hemisphere, reuniting them in time to write up the best answer.

Parfit reckons that I would not experience any sensation of division, as each stream of consciousness would seem to be straightforwardly continuous with the pre-division single stream, the only difference for each 'subject' being the disappearence of half the visual field, and of the sensation and control of half the body. As

one subject works on one strategy, he can see the hand controlled by the other subject working on the other strategy, but he is unaware of the corresponding mental processes underlying the writing - in looking at it, it is just as if he is 'peeking' at a neighbour's work (without the risk of being disqualified for cheating!). Once both sets of calculations are done, there comes a time where 'I am now about to unite my mind'. Once this reunion is achieved, Parfit expects that I will remember having worked on both strategies, and that both will be genuine memories.

The first point I have to make here is that Parfit's thought-experiment is invalid because he has made errors concerning the facts on which it is based. If some such device could cut off all direct communication between my hemispheres, each half-brain would not therefore lose half their visual field in such a situation, as head and eye movements would compensate perfectly. Parfit might reply that this is a minor error which doesn't affect his major point. Perhaps ..... but he faces other difficulties. One concerns the above-mentioned quote 'I am now about to unite my mind'. This is unfortunately expressed, as it implies the existence of some 'I', some conscious subject of experience, over and above the two disconnected hemispheres. Certainly Parfit would not want to admit such a subject, and neither would I. Still, this action of reuniting the mind would seem to require a high degree of cooperation between the two disconnected hemispheres, thus posing a problem as to how this is achieved. Perhaps one could signal to the other 'I've finished' and the other could reply 'so have I' by a pre-arranged code. If not, the difficulty can be overcome by recasting the story while retaining it's central point, for example by dropping Parfit's eyebrow device, and replacing it with some pre-set timing device, so that one could set it to divide the mind and to reunite it after a given time, thus relieving the hemispheres of the task of having to arrange it by themselves.

In fact, if we revise the thought-experiment as I suggest, it has the added advantage of improving its credibility by bringing it a little closer to what is practically possible at present. Wada (see Patricia Churchland [1986]) has shown that it is possible to anaesthetise a single cerebral hemisphere by injecting sodium amytal into the ipselateral carotid artery, thus putting it out of action for as long as the drug is operative. Let us imagine that a drug has been invented that can similarly block the activity of the corpus

callosum, while leaving the hemispheres themselves unaffected, apart from their functional separation. Let us also imagine that the process has been refined to such an extent that one can be so specific as to put the corpus callosum out of action for ten minutes, give or take a few seconds. We now have the means for reconstrucing Parfit's story in such a way that the problem of re-uniting is avoided.

However, my second problem isn't so easy. It goes back to the question of whether the hemispheres themselves have the self-consciousness necessary to earn the status of 'selves' (see the debate between DeWitt and Puccetti, ch.7.6). I have already noted that the experimental results have proved inconclusive on this question. I propose a means whereby this matter can be settled. Quite simply, the experimental set-up remains that as devised by Sperry and Zaidel, incorporating a Z-lens, whereby an image of a printed sentence is flashed to each hemisphere separately, namely 'which are you, the left hemisphere or the right?' Patients would be told to reply by means of previously arranged hand signals.

Let us imagine for the moment that the hemispheres fail the test, and return to Parfit's thought-experiment. If such a self-consciousness was beyond them, then a self-identification would be equally impossible, and the thought-experiment could not get started. The plan that 'the left hemisphere will attempt strategy A' is only possible if the subject knows that he is the left hemisphere. It will clearly not do to peek at what the other subject is doing - i.e. to see that he's doing strategy B, and that therefore he must be the right hemisphere, and therefore I must be the left. This will not work, because obviously the problem of self-identificatoin will affect this other subject as well, so it will only succeed if the other's identity could be established independently of his attempting strategy B. So if neither he nor I know which one he is, then I cannot know whether the strategy he is working on is his originally allocated one, and therefore I cannot deduce my correct strategy, nor my identity, from it.

Now the obvious solution to this problem is to suggest that each subject could identify himself by reference to his relationship to the body - i.e. given the established facts regarding contralateral control of limbs, etc. for example, a subject could attempt to write with both hands, and if he could only do so with the right, then he can deduce that he is the left hemisphere. Now I accept this, but

strange consequences arise from it. It means that for 'subjects' in a split-brain situation, correct self-identification is contingent on knowing certain neurological facts which, while they are common knowledge within the appropriate scientific circles, are still unknown to most people. This state of affairs is fundamentally different from the position of subjects embodied by undivided brains - I do not deduce my identity by recourse to any tests. It is also worth noting that such a test is not applicable to all situations - for example in an isolation tank, where one is deprived of all external sensory stimuli by being within a light-proof, sound-proof tank, floating in a saline solution at body temperature. If a split-brain patient, as depicted by Parfit, was placed in such an environment, his means of self-identification would be removed.

CHAPTER 8 DEGREES OF PSYCHOLOGICAL UNITY

8.1 MULTIPLE PERSONALITY DISORDER : HISTORICAL BACKGROUND

Having cleared up some misconceptions regarding the most-discussed 'problem case' regarding personal identity and the concept of the unity of mind, I wish to turn now to another prima facie paradoxical condition, namely Multiple Personality Disorder (MPD), which seems to call into question the fundamental notion of mental unity that we take for granted.

MPD was originally categorised as an extreme form of dissociative disorder, and, after numerous changes in status, this is how it is regarded at present. As Fahy [1988] says, *"The nosological status of MPD was altered when it was listed as a diagnosis among the dissociative stated in DSM-III (American Psychiatric Association 1980) having been included as a symptom in the hysteria section in DSM-II (APA 1968). MPD will also be accorded special recognition in ICD-10, where it will become a special diagnosis classified under the dissociative disorders of memory, awareness and identity (World Health Organisation 1987)" (p598).*

The concept of dissociation originally derives from the theory of 'association', that memories were brought to conscious awareness by the mechanism of 'association of ideas', so that memories that are not available by connection to such a network are 'dissociated'. More recently, Taylor and Martin [1944] define MPD as a generic term specifying psychological conditions wherein *"two or more personalities, each of which is so well-developed and integrated as to have a relatively coordinated, rich, unified and stable life of its own"(p282)* appear to occupy one and the same physical body. Each personality is equipped with a complex integrated structure of memories, behaviour-patterns, opinions, values, etc., with each such structure having autonomy and independence from its other 'cohabitees', often to the extent of complete ignorance as to the others' existence, except through indirect, external channels.

It must be noted that MPD is a very rare condition. It seems to have reached a peak between 1840 and 1910. For the next few decades there was scarcely a book or article to be found on the subject (one notable exception, to be discussed in the following pages, was the case

of Eve White/Eve Black) and the few that did appear tended to rely on historical cases from the Victorian heyday. As Nemiah [1981] remarks, on the subject of dissociative disorders in general, by the early 20th century, *for reasons that are not entirely clear, there was a sudden loss of interest in these clinical syndromes... (but)... such clinical apathy cannot be blamed on a disappearence of patients suffering from the disorders, for there is no evidence to suggest that the incidence of psychogenic amnesia and most of its related clinical states has in any way diminished since that time"(p1545).* However, it should be admitted that Nemiah regards MPD as the exception, having all but disappeared. On the other hand, at a recent symposium on MPD by the Psychiatric Clinics of North America, the considered opinion was that there are around one thousand cases at present in the USA alone.

In the light of Nemiah's observations, it seems most likely that, given that the conditions described as 'dissociative' have been with us throughout this century, their low profile was a consequence of major shifts taking place in psychological and psychiatric theory, firstly under the impact of Freud (around the time that the syndrome 'disappeared'), and later as a result of the shift of attention from 'mind' and 'consciousness' by the behaviourists. As Kuhn made very clear, such a change in perspective is accompanied by a change in interests and priorities, so that what was once a central research project or field of inquiry suddenly becomes an unfashionable, marginal issue, and is subsequently ignored.

It has been argued that the present relative scarcity of cases of MPD, and its virtual confinement to one socio-historical period, is a strong indication that the condition may have been an artefact of one particular social environment, and, more particularly, the self-fulfilling creation of a specific psychiatric theory - in other words, the suggestion is that MPD was what we would now call an iatrogenic illness. For instance, in a supposedly authoritative article, Charles Rycroft [1987] argued that two necessary conditions for the development of the condition are *"(1) Prevailing views on the nature of personality make it conceivable that two personalities can occupy the same bodily frame, and (2) The potential case of multiple or split personality encounters a psychiatrist who believes in, or is already interested in, dissociation of the personality" (p197).*

This first alleged condition looks confused. It is certainly the

case that such a psychiatric condition would not, nay could not be characterised, nor even recognised, except within a theoretical framework that regarded it as conceivable. But this is a general point about the conditions of perception and of understanding, and as such cannot carry the critical weight that Rycroft intends for it. Also, it seems to be the wrong way round - if there had not been cases presented to psychiatrists that exhibited the symptomatology that could possibly be schematized as falling under the category of MPD, it is hard to see how any theories put forward to account for such phenomena could ever have been developed (or indeed why they would ever have been formulated). Furthermore, many case histories clearly state that the symptoms were well-developed before the patient was presented to any psychiatrist. So it seems to be a very strong bet that the symptoms that had previously been put down to demonic possession and suchlike, and were later put down to repression and the unconscious, had at least very strong similarities to those which were diagnosed as MPD.

Rycroft's second condition can be interpreted in two ways : firstly, as with his first condition, it can point to the fact that a theory is a lens through which we come to interpret experience, or, more significantly, he can be interpreted as pointing to the phenomenon of transference, wherein patients will unconsciously produce the symptoms that the doctor wants, in order to please him, for praise, etc. This process was not recognised in the last century. It is certainly a factor that must be seriously considered, but, since Freud, it is recognised as a hazard that can affect all forms of psychotherapy, and be present in any psychiatric condition, not just MPD, and thus it cannot be employed to dismiss the condition.

It has also been suggested, in conjunction with Rycroft's first condition, that, given that the vast majority of documented cases of MPD exhibited a pronounced split between, on the one hand, the original reserved, straight-laced personality, and a second emergent outgoing, fun-loving personality, the dissociation of the two was explained as a result of the repression of life-affirming impulses under the weight of convention and 'Victorian values', which, in tandem with the second condition, would be reflected in the formality and reticence of the doctor/patient relationship, so that the 'new' personality would only be allowed to come out into the open under the aegis of specially-designed contrivances such as hypnosis.

However, we can accept that these abovementioned social restraints

were responsible for the typical form that the split took, while maintaining that there is a general underlying tendency in the psyche to fragment under pressure, which is common to all social settings. There are also counterexamples to this typical form of split, for instance the Rev. Thomas Hanna, who, after sustaining a head injury, began to alternate between two personae, those personae while being mutually amnesic, were very similar in character, on the other hand, his general character - likes/dislikes, opinions, tastes, values - seems not to have been affected at all.

Anyhow, Rycroft's arguments are totally beside the point. Even if he is correct in saying that MPD was a purely iatrogenic condition and restricted to certain social environments, this does not remove the fact that the phenomenon existed (or exists) - it merely offers an account of how it came to exist, and whatever the cause, the effect is real enough. Furthermore, there now exist various independent techniques with which to detect splits in personality, based on the recognition that such psychological conditions have certain physical correlates or accompaniments. For example, as Wilkes [1988] reports, it has been demonstrated that different 'personalities' display different responses to EEG tests; in galvanic skin response to emotionally loaded words and phrases, and in visually evoked responses to light flashes. THere are also cases in which one personality is subject to specific allergies that the others do not share.

8.2 MINDS, PERSONS & PERSONALITIES

But what are we to say about cases of MPD? That there are two (or however many) persons coexisting in one body? That there is one person with two minds? One person with one divided mind? Note that the range of possibilities is similar to that characterising the debate concerning split-brain patients. Incidentally, it should be noted that in non-philosophical discussion of these issues, the terms 'mind' and 'self' (and others, such as 'centre of consciousness') are used fairly interchangeably, with, perhaps, in the case of 'mind', the more generally cognitive aspects stressed, and in the case of 'self', the more reflexive aspects emphasised. I hope that my sometimes informal usage will not lead to confusion. I will also try and show why some degree of looseness is inevitable in using such basic psychological terms.

But to return to the question raised in the previous paragraph, it is certainly the case that due to fundamental, deeply entrenched assumptions regarding persons - namely that the mind is unitary and indivisible, with one to each body, embodied in the brain, and with the person essentially identified with this mind-brain - then any of the above speculative suggestions will be perceived as having a substantial degree of oddity, since they diverge so strongly from common usage. But, as is so often the case, such a conflict can be an indication of an inadequate theory being stretched out of shape by the pressure of unwieldy phenomena.

My conclusion will be that to demand a decision between the above formulations, to ask which one is correct, is to ask an empty question, as they are merely equivalent descriptions of the same condition, all struggling to deal with the fact that MPD poses an immense challenge for the commonsense view that a person necessarily possesses a single indivisible conscious mind. I will also argue that all of the above formulations are equally inadequate to account for what is going on in cases of MPD, due to their being couched in folk-psychological terms. I reckon that we are fooled into thinking that there is a real disagreement between saying that it involves 'one person with two minds' or 'one person with a single divided mind' due to our deeply-rooted tendency to reify the mind (and the self) - to think that minds are <u>things</u> that come in discrete pristine units. In other words, we are still plagued by the ghost of the Cartesian Ego. Hopefully, elucidation will lead to exorcism.

I will also claim that there is no clear-cut black and white difference between someone with MPD and the rest of us, but merely different shades of grey. The integrity of the mind is a matter of degree, with each individual being placed at some point on a continuum, with, at one end, the (unattainable) ideal of a perfectly integrated psyche that fully exemplifies the Socratic maxim of 'Man, know thyself', and, at the other extreme, a state of complete fragmentation that is exemplified not by cases of MPD, but by someone like the Korsakoff's Syndrome victim William Thompson, discusses in Chapter 6, and who, as I have already suggested, cannot in any substantial sense be said to have a self at all.

At this point, I should perhaps be a little more specific regarding the central notion of 'unity of mind', and what constitutes a single

mind (although, as I have already hinted, we can only be precise up to a point). A single mind, and our network of folk-psychological concepts, primarily belief and desire, form the basis of a deeply entrenched though rarely explicit theory of human behaviour. Forgetting for the moment any question of mind/brain identity or any physical criterion for the individuation of minds, one highly plausible suggestion is that the mind is put forward not as a theoretical entity proposed in order to explain behaviour, but as a descriptive term to mark the coherence of one's behaviour through time. Such accounts will inevitably be holistic, invoking a network of propositional attitudes spreading out to account for a piece of behaviour. A simple model would be along the lines of 'Why did you do x?' - 'Because I wanted y, and thought that doing x would lead to y'. Underlying this basic explanation will be a complex system of beliefs, desires, values and interests, and the whole memory system underlying knowledge. So when any piece of behaviour can be viewed as rational in the light of this network, the subject can be said to have a mind.

We only feel drawn to make this theory explicit when we are presented with anomalous cases such as MPD that seem to be outwith its range. Due to the inherent vagueness of folk-psychological terms, it is not possible to give a clear demarcation line (except in a purely stipulative way) that would allow us to decide up to exactly what degree of disunity or fragmentation we can regard a mind as single - but clearly there are some cases that look to be definately beyond the pale. It is to these that I shall now turn.

8.3 THE CASE OF MARY REYNOLDS

Let us consider the famous case of Mary Reynolds (1793-1854) (see McDougal [1926]). At 18 years of age, this young woman 'of dull and melancholy temperament' awoke from a long sleep with a severe and wide-ranging amnesia regarding episodic, semantic and procedural memory - for example she remembered nothing of her past life, she was unable to recognise her family, and no longer knew how to read or write. Apart from this, the most noticeable feature of her condition was that she displayed a strikingly different personality, now being 'friendly, merry and adventurous, with a new interest in the outdoors'. I will call her original personality MR1 and her newly acquired one MR2, and when I want to refer to her without implying any split, I will simply call her Mary. After five weeks, and following another

prolonged sleep, she awoke as MR1, with all her pre-MR2 memories now restored, but with total amnesia regarding the previous five weeks as MR2. She alternated between these two mutually amnesic modes for another sixteen years before stabilising as a 'modified' MR2, where she remained for the rest of her life.

The argument that MR1 and MR2 should be regarded as being or as having two distinct minds focusses on the self-containment and internal integrity of structure of each personality. In order to account for any piece of behaviour of Mary as MR1, we would attempt to rationalise it in terms of the beliefs and desires of MR1 alone. Those of MR2 would be irrelevant, since they are as causally separated from those of MR1 as are those of any other person.

Another closely related reason for regarding MR1 and MR2 as being two separate minds is that any attempt to combine their two respective networks of propositional attitudes leads to paradox. Take some piece of information, 'p', that is discovered (and is both known and believed) by MR1, but not by MR2. If we want to deny the distinctness of the minds of MR1 and MR2, we are forced to say that Mary, at time t, both knows and doesn't know (and both believes and doesn't believe) that p.

However, as we have seen from my account of my experience of both knowing and not knowing a telephone number (in chapter 7), such a prima facie paradoxical result does not force us into positing a plurality of minds. However, unlike this case, in which it seemed that there was a lack of integration between different modes of information acquisition or recall, in such a case of MPD we are presented with the appearence of two fully formed subjects manifesting (or not) the same <u>form</u> of knowing (This is even more clearly illustrated in cases of intraconsciousness, to be discussed shortly). There is no correlated structural breakdown, such as between speech centres and the visual cortex in the case of pure alexia, or lack of integration of sensory stimuli as in the case of split-brain patients. And regarding this latter condition, any attempt to account for MPD on the basis of a split-brain model (by Fingarette [1969], for example) is completely unwarranted by the facts, as well as being clearly inappropriate to cases involving several co-existing personalities, such as with Christine Beauchamp, where such a model breaks down as there just aren't enough cerebral hemispheres to go round!

Returning to Mary's state of knowing-and-not-knowing that p, we can, following Haight [1980] construct an argument to show that the peculiarly selective and recurring qualities of the mutual amnesia and the non-cooperation between the two personae is strongly counter-indicative of regarding them as different aspects of one single mind. Imagine that MR1 discovers some fact, p again, that she reckons that MR2 could not be trusted to keep secret, so she destroys or alters any material evidence suggestive of p, making it look, if anything, that not p is the true state of affairs. This external rearrangement would be sufficient to prevent MR2 from acquiring the knowledge or belief that p. MR1 acted in the knowledge that mutual amnesia between MR2 and herself would bar MR2 from having direct access to MR1's knowledge that p. She wouldn't need to suppress, or to perform mental manipulations on any of her mental contents in order for the deception to go through.

Now it is certain that there is a paradox in the idea of a complete and conscious self-deception - roughly, If I know enough to devise such a deception, and to carry it off, then I'll know too much to be taken in by it. So, given that the deception by MR1 on MR2 succeeds, and is complete and consciously adopted, it cannot be a case of self-deception, and thus is incompatible with any attempt to combine MR1 and MR2 into one single mind, and thus we are forced to regard them as distinct and separate minds.

8.4 ALL ABOUT EVE

I will now turn to a second, equally renowned case of MPD, namely Ms. Chris Sizemore, who achieved fame (or notoriety) via the Hollywoodesque treatment of her case by Drs. Thigpen and Cleckley [1957],wherein Ms. Sizemore's extreme dissociative split was characterised as being between two highly distinct personalities, the outgoing, direct 'Eve Black' and the timid, constrained 'Eve White' (These two surnames were devised by Ms. Sizemore herself). Unlike the previous case of Mary Reynolds, this is not a case of two personae alternating in a relationship of complete mutual amnesia. Rather, their relationship exhibited one-sided 'intraconsciousness', where one personality is aware of the other and of her thoughts throughout, even when this other is dominant or 'manifested'. Eve Black (EB) was intraconscious of Eve White (EW) while EW was totally ignorant of the existence of EB. As EB said, '*I know her thoughts like she knows them herself. I*

don't think 'em, of course. But I can nearly always tell what's on her mind". It follows from this testimony that EB could somehow clearly distinguish EW's thoughts from her own. So we have the peculiar situation that EW was 'transparent' to EB, like with one's own self-awareness (or perhaps more like a case of complete telepathy), yet EB was a complete stranger to EW - she had no direct access to EB's thoughts, any more than to those of any other person, and had no direct evidence of the existence of her omniscient shadow.

EB's dominance over EW had a further striking aspect - by some peculiar form of intense concentration, EB could 'erase' certain of EW's memories, should she decide that EW should not be allowed to have them; yet she herself, EB, would still have access to these memories. Alternatively, on a Freudian model, we could say that these memories were buried to the EW mode, trapped in the unconscious, whilst remaining available to EB. But, looking back at the analysis of deception in the case of Mary Reynolds, again, on either way of looking at it, it is hard to accommodate it into any theory claiming that EB and EW constitute a single mind.

Wilkes [1988] examines the famous Beauchamp case in terms of Dennett's [1976] six 'Conditions of Personhood' to see if these encourage a singular or plural view of the patient. I will do the same for Chris Sizemore/Eve Black/Eve White. Now none of these conditions are unproblematic - but for my purposes, all I need to demonstrate in order to show that the plural view is the more appropriate is that there are no extra problems in the application of these conditions to a 'personality' to those that arise in their application to the rest of us.

Firstly then, both EW and EB are clearly capable of rational thought to the degree required of persons. Secondly, states of consciousness can be ascribed to them, as can mental/intentional predicates. Thirdly, both are capable of verbal communication. Fourthly, both can be ascribed self-consciousness, as can be seen from EB's ability to distinguish 'her' thoughts from those of EW. As Wilkes notes, the next condition of personhood is slightly more problematic in its application to personalities : Regarding the condition that whether x counts as a person depends in part on whether we treat her as an object of moral concern, then, as Wilkes admits, *"it is just a plain fact that the doctors in charge treat these patients as single individuals to be cured" (p121)*....yet she

admits the paradoxical corollary that at this point in treatment, involving distinct dissociated personalities, the doctor may not have a single person x there to treat, as the creation, or the recovery of such an integrated individual is the whole point of the therapy - i.e. one is presented with either EW or EB, so who, or where is Chris Sizemore? But to this it can be added that the treatment also involves the doctor taking the intentional stance towards individual personalities, for example in his attitude towards one personality regarding her attitudes and actions towards another personality. This last point shows that the sixth condition, regarding the ability to reciprocate the intentional stance, to see others as persons, is also satisfied by such personalities, again pointing to a plural view of the dissociated patient. These are clearly capable of what Frankfurt [1971] calls 'second-order' beliefs and desires, as is shown by their strategies towards other 'cohabitees'.

8.5 DISSOCIATION AND HYPNOSIS

Such cases of MPD exhibiting intraconsciousness have distinct parallels with certain states of mind that can be induced in at least the majority of us by means of hypnosis. Such parallels focus on the much-documented but little-understood phenomenon of post-hypnotic suggestion, wherein commands that are given while the person is in the hypnotic state will be acted upon later, once the person is brought out of the hypnotic trance, by means of some pre-planted cue, without the person remembering having received these instructions, and thus being blind to the true cause of his having performed the action relating to the command. Afterwards, if the action is innocuous or unremarkable (e.g. lighting a cigarette after the hypnotist scratches his head), he will rationalise his action, for example by claiming that he just felt a sudden urge for a cigarette. And, of course, he will not be lying - this is what it will feel like from the inside. However, if the programmed action is bizarre - for example, one incident that I witnessed where an unfortunate victim of a stage hypnotist suddenly jumped bolt upright in the middle of the audience, screaming 'The Russians are coming!! ' - such an action, totally incongruous in terms of his ordinary beliefs and desires, would soon be seen for what it was.

We can see the similarities to the case of EW/EB, in terms of a split in consciousness with one-way amnesia, and also in terms of inaccessible information surface to directly affect behaviour. But

the strongest parallel with hypnotic phenomena is with intraconsciousness. If we take a person 'A', and call him 'AH' when he is undergoing hypnosis, and 'AW' when he has awoken from the hypnotic trance, but still vulnerable to post-hypnotic suggestion, we can say that AW performs an action which has roots that are completely unknown to him, and since observation of the cue is a necessary condition of such an action (and since such cues can be subtle and undemonstrative, which implies that anyone acting upon it must, on some level, be on the lookout for it), and given that AW is unaware of the cue's significance, all this suggests that some residue of AH remains while AW is dominant. This suggests that A's condition is not like having two alternating personalities, but more clearly resembles a one-way intraconscious pair like EW and EB. Of course, any split in A is much more limited, as there is no way that AH is complex or developed or autonomous enough to be regarded as a personality, as his sphere of operation is ... that of acting upon a single cue. So AH is better described as an intraconscious dissociated response mechanism.

While acknowledging the similarities between hypnotic states and MPD, the actual relationship between them is not clear, as I mentioned in ch.6.9. As Fahy [1988] says, *"Since hypnosis has been used to treat MPD it has been important to clarify the extent to which it is responsible for the disorder. Under hypnosis, alternate personalities may reveal themselves when required by the therapist. That there is a relationship between hypnosis and MPD is suggested by the high hypnotisability scores of mostMPD patients"*. Subjects under hypnosis have some features in common with MPD patients, i.e. production of alternates with different behaviour patterns and amnesia, the appearence of such phenomena as automatic writing and post-hypnotic suggestion, and also *"7% of normal individuals were able to respond to suggestions to create a secondary personality using an age regression test..... However, the age regression model, like the automatic writing model, does not provide an entirely satisfactory comparison with MPD, lacking the complexity and chronicity of the clinical syndrome. There is scanty evidence that short-term exposure to hypnosis can induce well-developed alternates through the use of hypnosis alone.... That patients are excellent hypnotic subjects and prone to self-hypnosis does not prove that the relationship between hypnosis and MPD is causal" (p601).*

Hypnosis gives us clear examples of seemingly paradoxical states of 'knowing and not knowing', 'seeing and not seeing' that resemble

results of experiments on split-brain patients that I have already discussed. For example, as Wilkes tells us, one good way of finding out if someone is really hypnotised or is just faking it is to place an object, e.g. a chair, in his path, whereupon the faker will walk into it whereas the genuinely hypnotised person will walk round it. The peculiarity of this response lies in the fact that the person clearly sees the chair (as is shown by his ability to avoid collision with it), yet it is also as if he never saw it, as is suggested by the fact that he will not mention the chair if he is asked to give an inventory of the contents of the room, and also if he is asked why he diverted from his straight path, he will rationalise his action without reference to the chair.

Another split that is strikingly parallel to those shown in split-brain patients is hypnotic anaesthesia. Hilgard [1977] describes tests wherein a person is hypnotised and is told that he will feel no pain. His hand and arm are placed in a container filled with icy water, whereupon he continues to appear to be without distress, and, if asked, will confirm that he feels nothing. However, if he is given a pen in his other hand, and asked to write down on paper (positioned out of his sight) how he is feeling, he will complain bitterly about the pain. (The 'writer' is an example of a phenomenon that is commonly referred to as the 'hidden observer'). So here we seem to have a case of someone who both feels and does not feel pain, or, who both is and is not in pain.

8.6 THE SELF

Discussion of 'the self' in contemporary clinical psychotherapy does not presuppose the existence of the self as some separately-existing entity of that name, over and above the facts of physical and psychological continuity. As Toulmin [1977] says, we must distinguish this project from that of speculative philosophical psychology, where a major focus of attention was the purported existence of such an entity, which was postulated as a theoretical entity to explain the order and regularity among mental contents - an enterprise doomed to failure, as Hume recognised. To quote Toulmin : *"In clinical contexts, the point of invoking the self is not to speculate about concealed mechanisms... Theories of the self 'cash in for' a special class of empirical relations within the whole spectrum of 'self-' phenomena : self-esteem, self-control, self-understanding,*

etc." (p308). So to talk of a self has no hypothetical or explanotory overtones, but is rather descriptive and diagnostic, and rests within a context that provides both a grounding and an extension of everyday talk on reflexive conduct.

Rather than dwelling on the question of the existence of the self, the central issues tackled by a clinical theory of the self concern the dual concepts of integration and fragmentation of mental contents - beliefs, desires, etc. Toulmin again ; *"To be mature (or free from psychological troubles) is to have a 'cohesive, well-integrated self'; to suffer from psychological immaturities (or difficulties in the area of self-knowledge) is to have a 'fragile, fragmented, and/or incompletely-cohesive self'" (p309).* Such a schema is reflected in such common expressions as 'cracking up', 'falling apart' or, conversely of 'being together'.

Toulmin's position fits well with my depiction of what is involved in the 'unity of mind' as I have been building up over the last three chapters : *"The integration of the cohesive self is the compendious label for the coherence of feelings and motives, intentiona and actions, typical of a psychologically free agent....(consisting in).. a rational coherence between the different components of his self-knowledge.... Conversely, fragmentation of the self (consists in).... the absence of intelligible coherence between these different reflexive characteristics of the agent's conduct" (p309).*

8.7 A MATTER OF DEGREE

Perhaps one day, once a fully satisfactory neurophysiology or neuropsychology has been established, MPD will be given an explanation that dissolves its apparently paradoxical nature, where, instead of positing two or more 'minds', we could describe the physical processes and subsystems that fail to be integrated with each other. Perhaps we will be able to specify the brain processes underlying such a condition, and be able to cure or to pre-empt it - perhaps by chemical means, perhaps by electrical means, or by something completely different (e.g. acupuncture). However, for the moment, we must accept that, in Nemiah's words, *"the information currently available concerning neurophysiological processes is not sufficiently detailed to provide clinically useful concepts and explanations" (p1545).*

So, for the moment, explanations are confined to the psychological domain, and these tend to consist of hypotheses regarding the social conditions that could contribute to the development of MPD. Any 'philosophical' problems regarding the condition are left virtually untouched. I want now to further develop my position that we 'normal' people and those with MPD are not distinguished by having any different numbers of entities called 'minds' or 'selves', but that we occupy different positions on the integration-fragmentation continuum with regard to the relations between our beliefs, desires, etc. I will do so by examining a range of mental phenomena experienced by people in general (i.e. not clinical cases) to gain more insight into the issue of unity of mind, and to emphasise just how theoretically loose a concept it is. As an entry to this survey, I want to turn to a causative account of MPD as suggested by Glover [1988]. He notes that, as far as contemporary documented cases indicate, MPD is often linked with a history of child abuse. He argues plausibly that *"it may be very hard to form a single picture of yourself that includes both normal relationships..... and your role as the victim of these assaults. One way of coping with this would be to compartmentalise your life, behaving as different people in different contexts". (p23).*

By comparison, Glover draws our attention to well-documented cases of those who worked in Nazi concentration camps, as doctors, guards, etc., who participated in, or turned a blind eye to the most hideously evil deeds, yet returned home each night to resume life as, say, a loving husband and father. Here, one would not want to say that there were two minds or selves present, since, unlike in cases of MPD, there is no mutual amnesia, in any literal sense, between the two roles. However, even against a background of some integration, there is clearly a high degree of dissociation involved, and indeed this should not surprise us - how else would one cope with the contradictions of such a life? And let us not be complacent or superior - no-one who has studied the work of Stanley Milgram [1974] can assume that they would be any different. His subjects, who were prepared to torture others on command, were not essentially different or evil people, but ordinary people like you or I, responding to extraordinary conditions.

Glover quotes Robert Jay Lifton [1987], who describes the condition of the concentration camp worker as involving 'doubling' - a term he uses to describe a condition involving a psychological 'barrier', which is set up in order that the two highly conflicting modes of being, as the sadistic torturer and as the family man, are

felt as having nothing to do with each other, are turned into two mutually exclusive modes of living, where such a person leaves his humane persona at the gates as he 'clocks in' to the guard mode, and collects it again, intact, as he 'clocks out' at the end of the shift.

Lifton uses this term 'doubling' rather than mere 'splitting' to draw attention to the fact that such a process involves 'the creation of two autonomous selves'. In other words, there is a high degree of internal integration between the thoughts, feelings, etc., within each persona, but little interaction between them; unlike the hermetic aspect of non-intraconscious MPD. However, there are strong enough similarities to warrant the suggestion that MPD is basically a more extreme development of a general tendency of the psyche to fragment under pressure, whether internal or external.

Moving along the continuum towards the ideal of integration, we come to ourselves. We all know the man who is a subservient doormat at the office and a tyrant at home. Isn't this just a less extreme version of what Lifton is talking about? It is always easier to see such compartmentalisation in others! But it is now a commonplace that people adopt different roles and behaviour patterns in different situations and conditions, with little self-awareness or appreciation of doing so. We <u>all</u> express fragmentation of this nature to <u>some</u> degree. One of the most penetrating observations of this everyday condition comes not from the world of professional philosophy, psychology or psychotherapy, but from the Armenian mystic G.I.Gurdjieff (see Ouspensky [1950]). This is how he describes the average person's state of psychological integration. I include it as it makes an interesting comparison with Lifton, not least because of the similarity between their metaphors, of 'barriers' and 'buffers'.....

"'Buffer' is a term which requires special explanation. We know what buffers on railway carriages are, They are contrivances which lessen the shock when carriages or trucks strike one another. If there were no buffers the shock of one carriage against another would be very unpleasant and dangerous. Buffers soften the results of these shocks and render them unnoticeable and imperceptible. Exactly the same appliances are to be found in man..... The cause of their appearence is the existence in man of many contradictions.... of opinions, feelings, sympathies, words and actions. If a man throughout the whole of his life were to feel all the contradictions that are within him he could not live and act as calmly as he lives and acts now. He would have constant friction, constant unrest.... but if buffers are created

in him he can cease to feel them, and he will not feel the impact from the clash of contradictory views, contradictory emotions, contradictory words" (p154).

These groups of thoughts and feelings separated by 'buffers' are not substantial enough to count as 'selves' or 'minds', and are best described as roles or sub-personae, and represent the far less drastic degree and form of fragmentation found in people in everyday life, without the greater pressures of the concentration camp worker. Such 'buffers' are not depicted as static, but are dynamic, adapting to new experiences. Nor are they indestructable, although their 'dismantling' is a tricky business. However, a correct dismantling of buffers will allow one to achieve a greater integrity of function, free from the internal coercive agencies that such a departmentalisation of experience brings, and leads us up the continuum towards what Toulmin calls 'maturity'. This is, of course, the aim of psychoanalysis and of all forms of psychotherapy. (Of course, 'buffers' do not exist any more than do 'selves', over and above the beliefs, desires etc. themselves. The utility of such a concept is to mark a form of fragmentation between these).

Returning now to the issue of self-deception, this offers us another view into the relative fragmentation of the psyche in everyday life. As we have seen, there is a paradox in the idea that one mind can consciously and completely achieve a self-deception, and it was the factors of a consciously applied strategy and the complete success of the deception that made it impossible to characterise the hypothetical deception of MR1 on MR2 as <u>self</u>-deception, and, therefore, why we had to regard MR1 and MR2 as separate minds. Howver, despite the air of paradox, it remains true that we regard self-deception as a very common occurrence. I agree with this belief that self-deception does take place, and maintain that it is another indication of the relative fragmentation of the psyche, albeit to a significantly lesser extent than in the case of MR1 and MR2. I agree with Glover that the plausibility of the idea of a complete, conscious self-deception derives from an unconscious amalgamation of two different, lesser forms of self-deception.

Firstly, we have a self-deception that is conscious but incomplete, where one suspects that something is the case, but deliberately don't investigate the matter thoroughly or spell out to oneself the consequences of its being so. For example, take a self-deception I

played on myself not long ago : I noticed that I had developed a small but painful lump on my body. After an immediate panic of 'Oh no, I've got cancer!', I found myself deliberately ignoring it, not going to the doctor, immediately blocking the thought should I remember about it. This situation went on for weeks until I finally plucked up the courage to find out what the lump was (to my relief).

Secondly, we have a self-deception that is complete, but is unconsciously performed, where we unconsciously ignore evidence which, with hindsight, was staring us in the face, but which, if it was suggested to you at the time, you would deny it. For example, after a relationship has ended, you can often recognise certain events as signifying the beginning of the end, or as having an import that wasn't admitted at the time, through unconsciously choosing to go against an unpleasant truth.

8.8 IN CONCLUSION

I will finish this chapter by relating its conclusions to those of chapter 7 regarding split-brain patients, and thus expand on my theory of unity of mind. Shoemaker [1984] is in some agreement with my general position, in that he accepts that 'unity of mind' and the 'integration of mental states' is one and the same thing, and it applies not in an all-or-nothing way, but to varying degrees. He accepts that a mind can be 'compartmentalised', having various subsystems of beliefs, desires, values, etc., and that, while internally coherent, these subsystems have little or less coherence with other subsystems. He also contrasts this state with that of an 'integrated' mind in which all its' mental states form a coherent consistent set, so that one's actions can be seen as rational in the light of all one's mental states (although it is not clear whether he regards such a condition as practically possible, or, as I do, as an unattainable ideal or limit). But he accepts that most of us, having a modest level of fragmentation, count as having a single mind. In this context, he discusses the 'unity relation', this being the relation holding between mental states just in case there is the possibility of their being integrated into a single set.

But how are we to understand this cited 'possibility'? Clearly it cannot be 'logical possibility', as there is no formal contradiction involved in the claim that your mental state a, and my mental state b, can directly combine to produce an action in another body. So such a

type of'possibility' is far too lax for our purposes. In contrast, far too strict is something like 'practically or technically possible'. It may well be impossible, given the present state of play in medical science, to enable the unity relation to hold between certain mental states embodied in a single brain - for example in a particularly intransigent case of psychogenic amnesia, yet scientific research continues on the assumption that many things that are not realisable at present will be so in the future. So a case where the unity relation would not hold in 1900 may well do in 2001, by this reading of 'possibility'.

Instead, we need what I have called 'theoretical possibility', which is determined by locating the phenomenon under scrutiny within a theory that locates the limits of what is possible for an entity of the kind concerned. However, this cannot be done for mental states per se, and thus the unity relation cannot be given precise conditions of applicability. Folk-psychological concepts such as 'Mental ststes', 'consciousness', 'will', etc., are not natural kind terms. Crucially, they are inherently too theoretically vague to allow any clear set of entailments and implications that are required to be able to locate such items within a framework of natural laws.

So we can see now why MPD is still regarded as a source of philosophical perplexity. Like split-brain phenomena, pure alexia, visual agnosia, it is inexplicable in terms of folk psychology with its assumption of a single conscious agent, but, unlike these cases, the commonsense model has not been supplanted by another, more scientific, systematic model of explanation. So far, the neurosciences have shed little light on this condition, as there seems to be no associated physical/structural damage of an obvious form [illegible] although, as I have said, we now have the diagnostic tools, in the form of various scanning devices, that could easily locate such a lesion). It seems more likely that the underlying disorder (of which there obviously is one) is located on the biochemical level, i.e. a mutant gene, but, at the moment, there is no substantial theory of such a kind that can provide a scientific account of MPD.

FINAL OVERVIEW

The crucial point, from which all my other conclusions follow., involves the dismantling of the very 'Problem of Personal Identity', by showing how the traditionally problematic issues stem from focussing on 'puzzle cases', created by placing certain important aspects of human life, namely bodily and mental continuity, in conflict with each other, and thereby creating opposing 'criteria of personal identity', only one of which could be correct. SUch an error was associated with the employment of far too lax a conception of 'possibility', as regards whether these counterfactual states of affairs could really occur, which in turn led to a false estimation of the relevance of these hypothetical cases to the enquiries.

Once these criteria cease to be placed in opposition to each other, then, firstly, regarding the Physical Criterion, we can accept the commonsense view that we are essentially embodied beings, where there is typically continuity over time of a single body. Within this, we acknowledge the unique significance of the brain. I remain agnostic over the possibility of brain transplants. Should they occur, then I accept the judgement that I go where my brain goes. Pushing my liberalism regarding thought-experiments to the limit, I would say that should 'brain division' and hemisphere transplants ever occur, so that two surviving persons have portions of my brain, then Parfit's analysis is vindicated.

Regarding the Psychological Criterion, as we have seen, the issues that remain centre around the concepts of psychological unity and psychological continuity, concerning, respectively, the co-personality of mental events at-a-time, and over time. These issues are firmly located within the fields of the sciences, for instance the neurosciences in the cases of memory and split-brain research, and psychiatry as regards Multiple Personality Disorder. In order to make a useful contribution to the clarification of these issues, philosophers will have to become far more familiar with the relevant areas of scientific research than they have (with notable exceptions) shown themselves to be.

This contribution will involve an investigation of the relationships between firstly our commonsense theories and concepts; secondly our inherited philosophical theories; and thirdly the

results of scientific research, and their interpretation within scientific theories and concepts. Within this, the role of the philosopher is akin to that of the industrial service ACAS, i.e. advice, conciliation, arbitration : determining errors, agreements, misunderstandings, and, in cases of irreconcilable differences, delimiting the domain of application of each party. This will be done with the crucial acknowledgement that the philosopher is not thereby making judgements from 'on high', from some perspective above, or prior to that of the empirical sciences, but from the ground floor level, working with the scientists within the given field of enquiry, with each being engaged in different, but complementary aspects of the same enterprise.

BIBLIOGRAPHY

BARTLETT,F.C. [1932] Remembering Cambridge

BRAUDE, Stephen [1988] PSI and our Picture of the World;(Inquiry Vol30)

BRENNAN, Andrew [1988] Conditions of Identity Clarendon Press

BROAD, C.D. The Relevance of Psychical Research to Philosophy (in Wheatley and Edge)

BUTLER, Joseph [1736] Of Personal Identity (in Perry)

CAMPBELL, Jeremy [1982] Grammatical Man Simon & Schuster

CHURCHLAND, Patricia [1986] Neurophilosophy Bradford MIT

CHURCHLAND, Paul [1988] How Parapsychology could become a Science Inquiry Vol.30

DENNETT, Daniel [1976] Conditions of Personhood (in his 'Brainstorms' Harvester 1979)

DEWITT Larry [1975] Consciousness, Mind, and Self (in British Journal of the Philosophy of Science 26)

DUCASSE,C.J. [1961] How Stands the Case for the Reality of Survival (in Wheatley & Edge)

ECCLES, John [1970] Brain and Conscious Experience (in 'Facing Reality: Philosophical Adventures by a Brain Scientist' Springer Verlag, New York)

ECCLES, John,& POPPER, Karl [1977] The Self and its Brain Springer

EVANS, Gareth [1982] The Varieties of Reference Clarendon Press

FAHY, T.A. [1988] The Diagnosis of MPD : A Critical Review (in British Journal of Psychiatry 153)

FLEW, Antony [1951] Locke and the Problem of Personal Identity (in Philosophy 1951)

FOSTER, John [1979] In Self-Defence (in G.F.McDonald : 'Perception and Identity' Clarendon Press)

FRANKFURT, Harry [1971] Freedom of the Will and the Concept of a Person (Journal of Philosophy LXVII)

GARDNER, Howard [1975] The Shattered MInd Alfred E Knopf, N.Y.

GAZZANIGA, M. [1970] The Bisected Brain Plenum N.Y.

GAZZANIGA,M.& LEDOUX,J. [1978] The Integrated Mind Plenum, N.Y.

GILLETT, Grant [1986] Brain Bisection and Personal Identity (Mind 1986)

GLOVER, Jonathan [1988] 'I', The Philosophy and Psychology of Personal Identity Allen Lane

GREGORY, R.L. [1987] (ed) Oxford Companion to the Mind Oxford

HAIGHT, M.R. [1980] A Study in Self-Deception Harvester

HILGARD,E. [1977] Divided Consciousness : Multiple Controls in Human Thought and Action Harcourt

HUME, David [1739] A Treatise on Human Nature (Ed. L.A.Selby-Bigge, 1965) Clarendon Press

JAMES, William [1890] Principles of Psychology Dover Press, N.Y.

KUHN, Thomas [1962] The Structure of Scientific Revolutions University of Chicago Press

LEVY,J.,TREVARTHAN,C.,SPERRY,R.,[1972] Perception of Bilateral

LEWIS,D.K. [1983] Survival and Identity (& Postscript) (in his 'Philosophical Papers Vol.1, Clarendon Press)

LIFTON, Robert Jay [1987] The Future of Immorality New York

LOCKE, John [1694] An Essay Concerning Human Understanding 2nd Edition Ed. P. Nidditch Clarendon Press

MACKIE,J.L. [1976] Problems From Locke Clarendon Press

MADELL, Geoffrey [1981] The Identity of the Self Edinburgh U.P.

MADELL, Geoffrey [1985] Derek Parfit and Greta Garbo (Analysis 1985)

MARGOLIS, Joseph [1975] Puccetti on Minds, Brains and Persons (British Journal of the Philosophy of Science)

MARKS, Charles [1986] Commissurotomy, Consciousness & Unity of Mind Bradford MIT

MCDOUGAL,W. [1926] Outline of Abnormal Psychology Methuen

MILGRAM,S. [1974] Obedience to Authority New York & London

MISCHEL,T. [1977] The Self Blackwell

MOOR, James [1982] Split Brains and Atomic Persons (Philosophy of Science 1982)

NAGEL, Thomas [1971] Brain Bisection and the Unity of Consciousness (in his 'Mortal Questions, C.U.P. 1979)

NEMIAH,J. [1980] Dissociative Disorders (in Kaplan, Freedman & Siddick Comprehensive Textbook of Psychiatry,3rd ed. Williams)

NOZICK, Robert [1981] Philosophical Explanations Clarendon Press

OUSPENSKY,P.D. [1950] In Search of the Miraculous RKP

PALLIS,C. [1983] ABC of Brain Stem Death (British Journal of Medicine 1983)

PARFIT, Derek [1971] Personal Identity (in Perry)

PARFIT, Derek [1976] Lewis, Perry and What Matters (in Rorty)

PARFIT, Derek [1984] Reasons and Persons Clarendon Press

PENFIELD,W.& RASMUSSEN,T. [1957] The Cerebral Cortex of Man. MacMillan

PERRY, John [1975] (Ed.) Personal Identity University of California Press)

PIAGET,J. [1971] Biology and Knowledge. University of Chicago Press

PRICE, Marjorie [1974] Identity Through Time (Journal of Philosophy 1974)

PSYCHIATRIC CLINICS OF AMERICA [1984] Symposium on Multiple Personality

PUCCETTI, Roland [1973] Brain Bisection and Personal Identity (Britich Journal of the Philosophy of Science 24)

PUCCETTI, Roland [1975] The Mute Self (British Journal of the Philosophy of Science 26)

PUTNAM, Hilary [1970] Is Semantics Possible? (in his 'Mind, Language and Reality' C.U.P. 1975)

REID, Thomas [1785] On Mr. Locke's Account of Personal Identity (in Perry)

RITGERINK, Roger [1980] An Attempt at Mental Division (Canadian Journal of Philosophy 1980)

ROBINSON, John [1988] Personal Identity and Survival (Journal of Philosophy 1988)

RORTY,Amelie [1976] The Identities of Persons University of California Press

RYCROFT,C. [1987] Dissociation of the Personality (in Gregory)

SACKS, Oliver [1986] The Man Who Mistook His Wife For A Hat Picador

SHOEMAKER,S.& SWINBURNE,R. [1984] Personal Identity Blackwell

SHOEMAKER,S. [1985] Critical Notice of 'Reasons and Persons' Mind 1985

SIDIS,B.& GOODHART,S [1905] Multiple Personality : An Experimental Investigation into the Nature of Human Personality Appleton, N.Y.

STEVENSON,I. [1966] 20 Cases Suggestive of Reincarnation (see Watson)

STRAWSON,P.F. [1959] Individuals Methuen

SWINBURNE,R, [1974] Personal Identity (Proceedings of the Aristotelian Society 1974)

TAYLOR,W.& MARTIN,M W1944] Multiple Personality (Journal of Abnormal and Social Psychiatry 1944)

THIGPEN,C.H.& CLECKLEY,H. [1957] The Three Faces of Eve N.Y.

TOULMIN,S. [1977] Self-Knowledge and Knowledge of the Self (in Mischel)

WATSON,L. [1974] The Romeo Error Hodder & Stoughton

WHEATLEY,J.M.O & EDGE,H.L. [1976] Philosophical Dimensions of Para-Psychology Charles Thomas

WIGGINS, David [1980] Sameness and Substance Blackwell

WILKES,K.V. [1978] Consciousness and Commissurotomy (Philosophy 53)

WILKES,K.V. [1988] Real People Clarendon Press

WILLIAMS, Bernard [1957] Personal Identity and Individuation (in WIlliams [1973])

WILLIAMS, Bernard [1970] The Self and the Future (in Williams W1973U

WILLIAMS, Bernard [1973] Problems of the Self C.U.P.

WITTGENSTEIN,L. [1953] Philosophical Investigations Blackwell

WITTGENSTEIN,L. [1958] The Blue and Brown Books Blackwell

YOUNG,J.Z. [1987] Philosophy and the Brain O.U.P.

ZAIDEL,E. [1975] A Technique for presenting Lateralised Vision Input with prolonged exposure (Vision Research 15)

NAME INDEX

SUBJECT INDEX