Psychological Reductionism

About Persons

A Critical Development

Julian Guiseppe Baggini

Thesis submitted for PhD Examination

1996

University College London

Contents

Abstract 1

Introduction: Philosophy and Persons 2

Part One

Chapter One: The Kierkegaardian Requirement 8

1. Why Kierkegaard? 9
2. Kierkegaard’s Stages 11
3. Two Elements to be Resolved: Some Familiar Analogues 21
4. Kierkegaard’s Solution 24
5. Creating a Self 30
6. Conclusion 33

Chapter Two: The Relevance Requirement 35

1.The Issue 37

2. Identity Over Time 44
3. Fission and the Impotence of Identity 51
4. Teletransportation and the Limits of Identity 56
5. Conclusion 59

Chapter Three: Persons and Thought Experiments 62

1. The Thought Experiments 63
2. Two Types of Objection to Thought Experiments 67
3. The Method of Cases 78
4. Persons, Animals and Humans 91

Part Two

Introduction 102

Chapter Four: Parfit and “I” 104

1. What is Reductionism? 105
2. Kant: No Thoughts Without a Subject 111
3. Cassam: No I-thoughts Without a Subject 118
4. Parfit and Cassam 120

5. Parfit and Kant 124
6. Conclusion 126

Chapter Five: Parfit's Relation R 128
1. Psychological Reductionism and Diachronic Unity 128
2. Problems With Relation R 141

Chapter Six: The Parfitian Conception Revised 152
1. Direct Phenomenological Connectedness 155
2. Functional Indexicality 160
3. The Internal Narrative 163
4. The I* Relation v. Relation R 167
5. "What Matters in Survival" 172
6. Conclusion 177

Part Three

Chapter Seven: Appraisal 178
1. What is a Person? 180
2. Why Is the First Person Point of View Important? 184
3. The Self and Society 189
4. The Relevance Requirement 199
5. The Kierkegaardian Requirement 201
6. Conclusion 206

Bibliography 208
Acknowledgements 214

Abstract

There is a need to distinguish two questions in the philosophy of persons. One of these is the *factual question of identity*. This is the question of the conditions of personal identity over time. The other is the *first person question of survival*. This can be expressed as, "Under which circumstances should I consider a person at another time to be my survivor, who I have reason to care about just as much if he were me?" This second question does not presuppose that the survivor is numerically identical with her predecessor and is the question considered in this thesis.

Answering this question requires us to resolve the tension in our concept of a person between, on the one hand, the view of persons as purely physical beings, no more than the sum of their particular parts, bound to the here and now, and on the other hand, as somehow transcendent, beings who exist beyond the here and now.

The conception built upon is that offered by Derek Parfit in *Reasons and Persons*. Two errors in Parfit's account are explained and amendments suggested. The first is Parfit's explanation of the unity of a mental life over time in terms of connectedness and continuity between individual, independent thoughts, and secondly his account of connectedness and continuity itself. I suggest that psychological connectedness and continuity must be between persons-at-a-time, not individual thoughts, and that a unified mental life over time is not just a product of enough connections, as Parfit argues, but is determined by the kind of connectedness there is.

Introduction

Philosophy and Persons

It seems odd that although we would all of us agree that we are all persons, it is notoriously difficult to say just what a person is. Currently, there are two general types of accounts of what persons are which attract significant support among philosophers. One is that a person is simply a human being. On this view something qualifies as a person on the grounds of its biological genus, i.e. what specie it is. The other is that a person is any rational, self conscious being. On this view, something qualifies as a person on the grounds of certain attributes and capacities it has. These two views have very different implications for what constitutes personal identity over time. If a person is simply a human being, then the conditions for the identity over time of a person are the same as the conditions for identity over time of the particular *homo sapiens* that person is. If a person is conceived as a being with certain mental attributes and capacities, then there are grounds for believing that psychological, not physical, continuity is the basis of personal identity over time. If these were the only disagreement concerning the term ‘person’ we would have arguments enough. But the waters of this debate have been further clouded in recent years by Derek Parfit’s *Reasons and Persons*. In this book, Parfit claims that the debate concerning what persons are and what their conditions for survival over time are, is not only a debate concerning facts, but also about “what matters”. This phrase, “what matters” is far from unambiguous. But it at least suggests that philosophical questions concerning persons are questions as much of value as of fact. If in her treatment of the subject of persons the philosopher has to pay due regard to questions of fact and value, which also requires identifying just what that ‘value’ is, her task is yet more difficult.

Considering these questions myself, I became increasingly convinced that the centre of gravity for the debate had become obscured by the rapid developments in the subject. The emphasis on the concept of mattering by Parfit had left a real confusion as to just what was at issue in the area of what I call philosophical anthropology – the philosophy of persons. What I felt was needed was a re-examination of just what questions this subject is trying to answer. Without such an examination, the suspicion would remain that philosophers who were ostensibly asking the same questions in fact understood these questions in quite different ways.

I believe that there are in fact two main, distinct questions in philosophical anthropology. One of these is what I call the *factual question of identity*. This is the question of the conditions of personal identity over time, traditionally expressed in the form, “What are the necessary and sufficient conditions for a person at one time and a person at another time being the same person?” However, I believe there is another, distinct question which is equally pertinent, namely the *first person question of survival*. This can be expressed as, “Under which circumstances should I consider a person at another time to be my survivor, who I have reason to care about just as much if he were me?” This question does not presuppose that the survivor is numerically identical with her predecessor. “Survival” here has a special meaning which does not entail identity. In chapter two I argue that the factual question of identity and the first person question of survival can be distinguished and that the latter question is not only a pertinent one, but one which is at least as central to the philosophy of persons as the question of identity. It is also the central question of this thesis.

One challenge for the philosophy of the persons is therefore to answer the first person question of survival as a question distinct from the factual question of identity. In chapter one I set out a further challenge. There is a tension in our concept of a person between, on the one hand, the view of persons as purely physical beings, no more than the sum of their particular parts, bound to the here and now, and on the other hand, as somehow transcendent, beings who exist beyond the here and now. In the work of Søren Kierkegaard, we can see how important this tension is and how difficult it is to resolve. Kierkegaard not only formulates this tension very well, but he also shows why it is necessary to resolve the tension, if our understanding of persons is to be complete. In chapter one, I explain in more detail what this tension is, and why the challenge to resolve it is one we cannot ignore.

I believe that the philosophical approach to persons best equipped to meet the challenges set out in the first two chapters is psychological reductionism, and in particular, the version set out by Derek Parfit in *Reasons and Persons*. In order to even get off the ground with Parfit's position, we have to consider two foundations upon which his arguments rest. The first is the use of thought experiments in argument. Parfit's discussions of teletransportation and fission have helped make his work one of the most engaging, colourful and fascinating works of contemporary philosophy. But it has also lead to the suspicion that somehow these thought experiments have made the philosophy of persons too speculative, or that they have lead us to rely too much on our untrustworthy intuitions. In chapter three I consider the main objections against thought experiments in the philosophy of persons and defend their use in the debate.

Also in Chapter three, I consider objections against the use of the term "person" at all. Some see this as a concept so ill-defined and vague that it is not a suitable object of study. Many argue that "human being" would be a much more favourable substitute. I resist these claims and argue that we must talk of persons if we are to answer the first person question of survival.

Part one establishes the challenges of philosophical anthropology and defends a certain approach to the subject. In part two, I turn to Parfit's particular thesis to see if it can meet the challenges set out in part one. Parfit's position includes many claims. Which of these are central to his conception is clearly a subject of debate. However, I see three basic tenets of Parfitianism which together mark it out as a distinctive conception of persons. The first of these is an explanation of what is required for a unified mental life over time in terms of psychological connectedness and continuity. This explanation consists of a relation, which Parfit calls relation R. The second feature of Parfitianism is that this relation is distinct from the identity relation. The third feature is that it is this relation, not the identity relation, which is required for our survival. The way I define "Parfitian", if these three features can be retained, then no matter how much we change Parfit's account, we will still have a distinctly Parfitian conception of personal survival. This thesis, although a critical development of psychological reductionism, still aims to retain such a Parfitian conception.

However, we do need to make some important changes to Parfit's account. Parfit explains the unity of a mental life over time in terms of connectedness and continuity between individual, independent thoughts. This is, I believe, his greatest error. Firstly, as I explain in chapter four, thoughts cannot be entirely independent of the subjects that think them. And secondly, as I argue in chapter five, Parfit's account of connectedness and continuity is inadequate. I suggest in these two chapters that psychological connectedness and continuity must be between persons-at-a-time, not individual thoughts, and that a unified mental life over time is not just a product of enough connections, as Parfit argues, but is determined by the kind of connectedness there is.

In chapter six I show how an alternative to the Parfitian conception can be formulated that follows these suggestions and thus avoids the errors of Parfit. I do this by replacing Relation R with what I call the *I* relation*, which sets out the ways in which persons-at-a-time must be related in order for there to be a unified mental life which joins them. The formulation of the I* relation does not pretend to be definitive or final, but what it does show is how it is possible to formulate a relation that can do the work of Parfit's Relation R without committing Parfit's errors, whilst retaining the distinctive features of Parfitianism.

I conclude in chapter seven by assessing the revised Parfitian view and seeing if it meets the challenges set out in part one.

There is one perhaps surprising upshot of my argument. In attempting to develop psychological reductionism, it is possible that I have actually eliminated the more reductive elements of the Parfitian account. I have tried to avoid the question of whether or not my final position is a reductive one, because I have come to the conclusion that this is not an important issue. I believe that my revision of Parfit is still close enough to his view to be considered Parfitian, but the question of its reductiveness I leave to one side.

The core of my argument can be expressed briefly. There are various ways in which persons at different times can be related. One relation is that of identity. Another is biological relatedness. And another is that the persons are psychologically connected and continuous with each other. In the latter case, the persons in question need not be identical. Even so, the person at the earlier time can view the latter person as their survivor. Furthermore, when I hope that I continue to exist, what is important in this hope is fulfilled by my survivor. Such a position is tenable even if we overthrow the particulars of Parfit's reductionist account of psychological connectedness and continuity. How and why this could be true is the subject of this thesis.[1]

[1] It is unfortunate that at the time of writing, I have not been able to see Dancy's forthcoming volume, *Reading Parfit*, which contains Parfit's first major set of responses to criticisms of his position in *Reasons and Persons*, together with some new essays on Parfit. The opportunity to have seen this book would have been extremely useful in assessing how far Parfit has already gone to accommodate some of the criticisms of his position discussed in this thesis

Chapter One

The Kierkegaardian Requirement

The worst readers are those who behave like plundering troops: they take away a few things they can use, dirty and confound the remainder and revile the whole.

Nietzsche, *Assorted Opinions and Maxims*

The first two chapters of this thesis argue for two requirements that any philosophical account of persons, including the one put forward in this thesis, should meet, if that account is to be a convincing one. I have developed the first of these requirements from the work of Søren Kierkegaard (1813-1855). Although Kierkegaard does not belong to the analytic tradition in philosophy, this thesis does. By looking at his work I aim not to abandon but to to enrich the analytic debate. The structure of this section is as follows: Firstly, because the fairly lengthy inclusion of the thought of a religious 'existentialist' thinker in a work of this kind may seem incongruous, I offer a brief explanation of the role I wish it to play. There follows a short exegesis of the central idea of Kierkegaard's on which I wish to concentrate, that of his concept of Stages. This leads to the main dilemma I am interested in; a dilemma that an account of personhood must solve. I follow this by looking at some more familiar philosophical ideas that are very closely related to those of Kierkegaard, in order to make the relevance of his ideas clearer. I then return to Kierkegaard's work to see how he answers the challenge laid down to philosophies of persons. Although I will not be recommending Kierkegaard's own solution to the problem he sets, I will suggest that there is plenty we can learn from his response. My hope is that when we then turn back in the next section to more familiar arguments and positions, we will have gained valuable new insights and tools.

1. Why Kierkegaard?

Like it or not, be it desirable or not, the post-Kantian split between ‘continental’ and ‘analytic’ philosophy is a real one, that is say, intellectually as well as politically. Without justification, one can no more easily draw on a continental philosopher as a source than one can the work of a sociologist or anthropologist. But, of course, philosophy does draw upon other disciplines, even technical ones such as quantum and chaos theory, and this is perfectly acceptable just so long as we can clearly say how they connect. In the case of Kierkegaard, there are at least two reasons why I think what he wrote can help illuminate our inquiry.

Firstly, whatever method of philosophy is employed, most of the great works of philosophy contain what I will call “insight”, by which I do not mean to imply the necessity of any blinding flash of inspiration. By insight I mean a central tenet or maxim which can be understood self-sufficiently, independently of any system of thought to which it may belong. A good example of this comes from one of the great system builders, Kant. Kant’s insight was that it only makes sense to think of the world as it conforms to our perceptions: his ‘Copernican revolution’. We can understand and debate this maxim without ever needing to talk of categories, antinomies or paralogisms.[2] Hume’s famous discussion of causation contains the insight that no empirical observation can be used as grounds for logically deducing that causation took place. This is not, of course, to say that Kant and Hume’s works were nothing but insights, or were simply the working out of consequences that followed from such insights. It is simply the case that these insights are not necessarily part of a wider ‘take it or leave it’ package. Phenomenologists could have, and indeed have, taken the same starting points

[2] By doing so, we may move away from Kant’s particular position, but I hope the reader can see that the idea of the Copernican revolution can have application outside of Kant’s system too.

and developed their philosophies in quite different ways. Analytic philosophers can build upon the insights of Continental philosophers just as continental philosophers can build upon the insights of analytic philosophers. This is indeed why I have found Kierkegaard so useful. It is not so much the whole picture his philosophy paints that interests me, it is some particularly brilliant brushstrokes. His philosophy provides insights that I believe are acute and which can play a part if transferred over to the analytic debate.

Secondly, we would be wrong to automatically dismiss all continental philosophy as being totally ‘other’. Much of Kierkegaard’s work is taken up with a rejection of Hegel, a rejection that depends upon highlighting contradiction and inconsistency. In this respect, a certain amount of what is actually done with the insights is compatible with the analytic method. So I aim to do rather more than simply borrow an insight or two. I am also going to examine at least some of the arguments Kierkegaard utilises.

The next question must therefore be: is what I say meant to represent Kierkegaard’s thought faithfully or am I simply acting like Nietzsche’s plundering troops, cutting, pasting and interpreting as suits my purpose? The answer is that there shall be a little of both. When, in parts two and four of this section I discuss Kierkegaard’s own texts, I hope I am being faithful to what he himself argued for. But in the remainder I shall assess what we can take away from it and from there on our route shall diverge from Kierkegaard’s. I hope that any exceptions to this schema will be sufficiently sign-posted to avoid any confusion between Kierkegaard’s position and my own.

In summary then, the answer to the question, “what is the place of Kierkegaard in this thesis?”, the answer is that we are seizing upon some of his insights and arguments in order to shed new light on our enterprise, although in uprooting them from the soil of Kierkegaard’s own thought, they shall grow in a different manner and bear different fruit.

2. Kierkegaard’s Stages.

There are various factors that make a brief exposition of Kierkegaard’s thought difficult. Much of his work is written under pseudonyms, each of which has a distinct personality, based on their metaphysical world views. The epistemological reason for this is that Kierkegaard rejected the idea of objective rationality, and therefore held that theories can only be looked at from inside, from particular points of views. There is also a methodological reason in that, like his hero Socrates, Kierkegaard believes that one only begins to philosophise when one enters a debate oneself, and holds that the writing of tracts is a barrier to such participation. Thus each individual work by Kierkegaard does not provide a self-contained exposition of one area of his thought. Each work needs to be read in the light of other works if we are to fully understand what each work really says. It is only when the works are brought together that one can see his philosophical position truly emerge. What follows therefore has an ordering and explicitness not found in Kierkegaard’s own writing, although I hope to have remained true to his arguments.

Kierkegaard is concerned are as much, if not more, with how we should live as with what we are, though both issues feature in his work and are strongly connected. It is easier to approach his work through what he says about how we should live. We must begin with a look at Kierkegaard’s conception of the different Stages within which we can live: the aesthetic, the ethical and the religious, the last of which I shall deal with in section four. There is a minor disagreement as to whether the term ‘sphere’ or ‘stage’ is the best to describe these three realms.[3] ‘Sphere’ has the advantage of implying neither a need to pass through all three nor total exclusivity between them. However, ‘Stages’ importantly implies a hierarchy of importance and for this reason is the favoured translation. The first of the three stages is the aesthetic. The aesthetic is concerned with the immediate and the instant, with what is finite in the human condition. The paradigm of the aesthetic life is described in *Stages on Life’s Way*, in the section entitled *in vino veritas*. It describes a banquet to which a variety of characters are invited. No-one is told of it in advance, everything is prepared only for that banquet and what is left is destroyed afterwards, epitomised by the throwing of a glass against the wall by Constantine. During the meal, each of the guests makes a speech. It is here that the irony of the title comes in, as unwittingly, each speaker reveals the limits of the aesthetic life. While in the very throws of the aesthetic, their words reveal its emptiness.

The most famous slogans of the aesthetic life are ‘seize the day’ and ‘live for the moment’. This is often taken to be an appeal to hedonism, although this is not necessarily the case. Although there is a seducer at the banquet, there is also a young man who cannot bring himself to make advances to the women he desires. This is because such a move would take him beyond the instant into some kind of commitment, however small. His long speech can only produce a list of the pros and cons of a relationship, based on particular observations. But none of these particulars can provide a universally valid reason, nothing can persuade him to go beyond the moment. What unites all the characters is a lack of inner fulfilment and a despair that is a result of permanently having to reinvent oneself. “Live for today because you could be dead tomorrow.” But when tomorrow comes and

[3] See Collins [1954], pp42-50, especially p45.

you're still alive the process has to begin again.[4] This relentlessness demonstrates the futility of ever trying to "seize the day".

Hannay's description of the aesthetic life as being "a life ensnared in or dedicated to immediacy."[5] therefore seems accurate. The only time that really matters, or has any reality is now. Therefore life is directed towards the now. This does not mean that an aesthetic individual does not think of the future or have any plans. In attempting to live for the moment, the wise aesthete will realise that all moments of time, past, present and future, are for a moment the now, and that therefore the future is not irrelevant. Thus in the seducer's diary of *Either/Or*, we see a scheme to attain a young woman unfolding over quite a long period of time. But the goal of this scheme is a moment of pleasure and not anything else. Thus it is quite possible to desire only that which has immediacy and yet work towards attaining that thing some time in the future.

The aesthetic life is not too difficult to comprehend. But it is less clear how one leaves the aesthetic sphere and enters the second stage, the ethical. Judge Wilhelm, in Part two of *Either/Or* reinforces the point that it is not enough to think long-term to truly live ethically, as the seducer's planning has already showed. Here, he comments on the validity of "contracting marriages for a definite period of time“

> It is already clear...that associations of this kind are not marriages, since although contracted in the sphere of reflection they have nevertheless not attained the consciousness of the eternal, as has the ethical way of life, which is what makes the alliance a marriage.[6]

[4] Dorothy Parker expressed this succinctly in her poem *The Flaw in Paganism*:
Drink and dance and laugh and lie,
Love, the reeling midnight through,
For tomorrow we shall die!
(But, alas, we never do.)

[5] Hannay [1982], p56

[6] Kierkegaard [1843a], p396

So consciousness of the eternal is clearly an essential element of living ethically. But in what does this consciousness consist? Primarily it is concerned with the consciousness of ethics as universal law, that there are such things as right and wrong and that these things are right and wrong for all people at all times. By realising that we are proper subjects for these laws, we acknowledge the eternal within ourselves.

But the ethical has problems of its own. The ethical, by laying claim to universality, immediately uproots itself from the instant and attaches itself to the eternal. And yet it is fundamentally unable to do so. Man is attached to the finite. No ethical system can be created by appealing to universal values, because the infinite and the universal are beyond reach. We therefore find our universal ethics to be founded not in the infinite but the temporal, as for example, embodied in tablets of stone delivered from Mount Sinai. This is the paradox of the eternal made finite and temporal. This is also the paradox of Christianity. What is eternal - God's truth - depends upon the temporal and the historical - the resurrection of Jesus Christ. But it is not just religious conceptions of morality which suffer these paradoxes. Any ethical claim, for example, "murder is wrong" cannot be a claim that murder is wrong here and now but that it is wrong in an absolute sense. As Kierkegaard saw it, universality is what makes ethics ethics. And yet we have no absolute basis for this rule-making. As finite, historical beings we simply cannot reach over to this infinite sphere. So for Kierkegaard, certain forms of 'moral relativism' would be a contradiction in terms, if moral relativism means that any moral rules we may have are provisional and thus subject to change at any instant, without attachment to a permanent framework. For the ethical presupposes a rule, and a rule by definition cannot be something created just for the instant.

I would now like to make explicit the difference between the aesthetic/ethical distinction as concerns the way we live and how it concerns the way we are. When I talked about the aesthete living for the instant and the ethical person living under a universal law, I was clearly talking about the way we live. But certainly for Kierkegaard, whether or not such ways of life are appropriate depends very much upon whether these ways of life are appropriate for the kind of beings we are. Kierkegaard believes we cannot detach the question of what we are from the question of how we are to live. In order to understand this, we first need to consider how the categories of the aesthetic and the ethical apply to the ontology of persons as well as to the ways in which we live our lives.

The aesthetic way of life is concerned with the moment. Persons conceived aesthetically are likewise beings trapped in the moment. On this view, a person has no existence beyond the here and now. Kierkegaard places the body in the aesthetic category. The body, being a physical thing, is seen as limited, finite and bound to the moment. Persons conceived ethically, however, are beings whose existence spreads out over time. Kierkegaard goes so far as to say eternally. Kierkegaard places the soul in the ethical category. The soul, as a non-physical thing, is not bounded by time, nor in some sense by space either. It is pure, indivisible and eternal. The ethical person receives the data of experience in the here and now but his existence is not confined to the moment. The aesthetic person is so confined, and has no existence beyond current experience. The aesthetic is thus characterised by the finite, the temporal and the body, while the ethical is characterised by the infinite, the eternal and the soul. These crude caricatures of the aesthetic and the ethical as applied to ontology will be more fully explained as we consider their significance in the rest of the chapter.

Kierkegaard believes there is a necessary connection between the ontology of persons and the way persons live. There are two reasons for this, neither of which entirely convince. The first reason is that the way we are determines the way we should live. If we make the aesthetic assumption, that we are just finite consciousnesses, bound to the instant, then what reason could there be to live ethically? I am not eternal, ergo, I am not bound by eternal laws. Who cares if I do something that is, from the point of view of ethics, wrong? I have no immortal soul to be punished for it, nor any existence beyond the moment that can suffer the consequences. And what is more, how can a finite intelligence even understand what these universal moral laws are? Therefore, to live ethically would be absurd. Similarly, if we are in some way infinite, then it would be equally foolish to live life on the principle of grasping the fleeting moment. To misquote the Bible, what good does it do a man to have all the world (a vast, but finite space) if he loses eternity? So the ethical question of how we live is supposed to follow from the ontological question of how we are.

The second reason is, conversely, that the way live determines the way we are. This follows quite closely the form of a Kantian transcendental argument, although this is not explicit in Kierkegaard’s texts. That is to say, starting from the way something actually is, Kierkegaard deduces something which must be true in order for it to be possible. The argument is that in order to live a life in a certain way, one must be constituted in a way that enables one to live in that way. This is as obvious as the claim that, in order for a piano to be played in a certain way, it must be constituted so as to make playing in that way possible. Through the pseudonymous texts, Kierkegaard shows the possibility of living ethically and aesthetically. Therefore, persons must be constituted in such a way as to make both the aesthetic and ethical ways of life possible.

Neither of these arguments are strong. In the first case, there seems nothing incoherent in a finite being committing themselves for a finite period to a moral law, nor for an eternal being living in the here and now. As for the second argument, the conclusion that we are constituted in such a way as to make living ethical and aesthetic ways of life possible is too general for Kierkegaard's purposes. There are many ontological theories of personhood which would allow such lives, as is clear from the observation that an ethical being can live aesthetically and vice-versa. So it is in vain that we look in Kierkegaard for a logically necessary link between ontology and the way we live. We should rather concentrate on asking ourselves if Kierkegaard is right when he says that man can be conceived ethically and aesthetically, and that there is a tension between these views, just as there is a tension between the aesthetic and ethical ways of life. I believe he is, but in order to make this convincing, we need to make a further distinction.

This distinction is best understood by considering whether Kierkegaard's dichotomy between the aesthetic and the ethical is really the either/or choice Kierkegaard presents it as. The immediate and the eternal are not so much two opposites as two extremes on a spectrum. Between them would be moments of time of varying lengths, better called periods of time. At one end is the unbounded period of time, the infinite, and the other, that length of time which is capable of being grasped immediately in consciousness, the moment. There is therefore something between the instant and the eternal, which I call the finite continuous. The question now is, is the finite continuous subsumed into one half of Kierkegaard's dichotomy, or is it truly separate? This is where ethics and ontology come apart. From the ethical point of view, the either/or remains the eternal or nothing. For the ethical person, unless they can live by universals, somehow connected with the eternal, then there is no ethical basis for life, as is pointed out by the Judge's comments on fixed-term marriages quoted above. The finite continuous is as problematic a base for morality as the immediate. But from the ontological point of view it is slightly different. For the person who conceives of life and himself aesthetically, it does not matter if they think of eternity or a period of time; what is still apparent to them is the sense in which they are bound to the moment. The temporal continuous is therefore sufficient to break away from the immediate. The idea of persons as finite-continuous beings would create many problems for the Kierkegaardian. In place of the simple finite/infinite dichotomy that existed both ontologically and ethically, there would be two dichotomies: the infinite and finite for ethics, and the immediate and what could be called the transcendent or finite continuous for ontology. I am not going to examine how or whether the Kierkegaardian conception could accommodate this upheaval. I would suggest, however, that the possibility of the finite-continuous reveals an over-simplification in Kierkegaard's either/or.

Kierkegaard's original dichotomy was interesting because it pointed to a fundamental tension between two characteristics of the self which seems impossible to resolve. The tension is still present in my reformulation. If the aesthetic is bound to immediacy, then how can an aesthetic individual also be finitely continuous? This is the dilemma which seems to me to be at the heart of Kierkegaard's dichotomy between the aesthetic and the ethical. It is not so much that fact there are two different alternatives, take them or leave them, but rather a seemingly insoluble problem: How does one go beyond the aesthetic? Let us look closer at the nature of this problem with particular regard to persons. Conceived aesthetically, a person seems little more than a succession of moments. More precisely, a person is a series of particular, immediate experiences. All we can ever get from such a series is simply a series. The complete series which makes up the life of a person is still simply an aggregate of particular experiences. An aggregate remains an aggregate and does not become a whole simply through there being enough elements in the aggregate. This is the problem that forced Hume to say that there is no single, unified self, only an illusion of one due to the contiguity and connectedness of particular instances. In this case Hume resolved the tension between the immediate (aesthetic) and the finite continuous (ethical). But his resolution was unsatisfactory, as he recognised:

> But all my hopes vanish when I come to explain the principles that unite our successive perceptions in our thought or consciousness. I cannot discover any theory which gives me satisfaction on this head.[7]

This reinforces the point that we do not need the strictly Kierkegaardian distinction between finite and infinite to keep hold of the tension between two ways of viewing the self. Nor do we need to maintain the strong line that both the aesthetic and the eternal have a place in reality. Hume's theory failed because it failed to explain how we are able to perceive of our perceptions as being unified in one consciousness. Had he been able to do this, there would have been nothing wrong with his claim that the single consciousness is an illusion. What is given is not that reality is both aesthetic and ethical, but that we have ethical and aesthetic perceptions of ourselves. Any account of how ourselves are needs to explain how this is possible.

It must also be remembered that it is as wrong to leave the aesthetic unaccounted for as it is for the ethical. Kierkegaard's puzzle is: how can you go beyond the aesthetic and at the same time preserve it? For it is doubtless true that we are in a very real sense bound to immediacy. By focussing on immediacy, Hume let the continuous slip. It is just as bad a mistake to explain the continuous and let immediacy slip. So we are caught between two seemingly irreconcilable impulses: the impulse to accept ourselves as bound to the moment and our impulse to view this as too limiting and to want to hold that there is a part to being a person which is more than just being a series of immediate consciousnesses. The challenge remains for us today to explain how these two can co-exist. This challenge is no less real if we replace the infinitude of the ethical with the notion of the finite-continuous.

This is a departure from Kierkegaard, who talked of the infinite, but I hope it can now be appreciated that we do not need to go so far with Kierkegaard to keep a grip on the fundamental tension. As the tension arises from an apparent incommensurability, it would appear that one of the two aspects of the self must be illusory. This would be the Humean line, mentioned above. It is also possible that, rather than being both ethically and aesthetically constituted, we simply are constituted such as to make it possible for us to live in both ways, although this seems to be equally contradictory. However, Kierkegaard decides that both must be real and attempts to account for how the two can be brought together, and this is where he goes beyond rationality and into "the absurd". It is tempting to choose this as the point at which to part ways with Kierkegaard. But I feel that how Kierkegaard tries to resolve this contradiction is instructive, and I will go on to discuss his attempt shortly. But first, a little stocktaking is in order.

Kierkegaard talks about the finite and infinite within us. Taken literally, to suggest this is useful in the current debate in personal identity would be eccentric and hard to defend. But I have suggested that the real problem is not so much attaining finitude's opposite as finding a way of breaking away from the immediate. The historical aspiration towards an ethical view of man may show an excess of ambition to reach infinity, but it does reveal the feeling of being much more than just the being here and now. But what the weak transcendental argument shows is that there must be something in us which allows us to have this thought. Memory and contiguity are but two of the candidates put forward to suggest how we are able to think of ourselves as, as these accounts would have it, more than we actually are. I will return to the problems in these arguments in

[7] Hume [1739], Bk1, Pt4, p331

part two. What I feel Kierkegaard does is make the necessity of accounting for the continuity of the self appear as important as it should. Persons are such that they are able to think of themselves as both locked to and in some way free of the now . Any account of the self that fails to explain this, fails, period.

3. Two Elements To Be Resolved: Some Familiar Analogues.

Kierkegaard presents two given aspects of the self and asks how they can be synthesised. Let us look further into the nature of the two givens. The first is the aesthetic, which includes finitude, the temporal and the physical body. One thing interesting about this is that it ties in very closely with the idea of the self viewed empirically. When Hume applied empirical observation to himself, all he could find were particulars at particular times:

> For my part, when I enter most intimately into what I call myself, I always stumble upon some particular perception or other ... I never can catch myself at any time without a perception, and never can observe anything but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of myself, and may truly be said to not exist.[8]

There was no self beyond the immediate content of consciousness and no way of deducing the existence of one. The extreme finitude of being a person became apparent. For this reason, persons, when viewed empirically, seem to belong to the aesthetic.

This is confirmed when we take a brief look at the successors of Hume. There has been a long tradition, starting from Locke, of accounts of personal identity which rest on memory criteria. In these, the locus of the self is taken to be conscious experience, and what makes two mental events at two times those of one and the same person is the link of memory. Of course, this view has come in many shapes and forms, some quite sophisticated. But what all the forms have in common is that the building blocks of a person over time are taken to be specific instances of conscious mental activity. The memory event of a past event links those two events to the same person. Here again though, despite attempts made to provide continuity through memory, the conception of the self can be said to be ensnared in immediacy, in that whole selves are mere aggregates of specific instances. This whole approach to the self can thus easily be seen as being essentially aesthetic, in Kierkegaard's sense. The notorious problem for this approach is how it accounts for the unity of consciousness and of persons, a problem which Hume for one admitted he could not solve.[9] Such an incompleteness in the account is only so much as Kierkegaard would expect. His account of the aesthetic focussed more upon psychological and existential incompleteness. Hume finds the same incompleteness, although this time it is discovered through a rigorous critique of his own arguments, which were based on empirical observation.

The ethical category is comprised of the infinite, necessity and the non-physical soul. Looking back over philosophy's past, it appears that persons have been conceived of in this ethical way when viewed as subjects rather than objects of consciousness. As thinking subjects, we by necessity think of ourselves as wholes over time and cannot simply conceive of ourselves as a series of particulars. Possessing such a unity is part of what it means to consider oneself as a subject of experience. Descartes's method of investigating the self conforms to this approach in that, rather than try and undertake empirical self-observation, he considered what it meant to be the thinking thing he knew himself to be. This

[8] Ibid, Bk1, Pt4, p301

[9] See page 11

method, in direct contrast to Humean empiricism, yields the result of a necessarily existent, indivisible, unperishable soul:

> We cannot understand a body except as being divisible, while by contrast we cannot understand a mind except as being indivisible. ...even if all the accidents of the mind change, so that it has different objects of the understanding and different desires and sensations, it does not on that account become a different mind; whereas a human body loses its identity merely as a result of a change in the shape of some of its parts. And it follows that while the body can easily perish, the mind is immortal by its very nature.[10]

This account too has its problems. How can we move from pure reason to facts about the world, especially facts about infinity and eternity? This question echoes Kierkegaard's point concerning the unattainability of the ethical. And also, how can we find room in Descartes's account for the empirical truths concerning persons that suggest finitude and divisibility? Again, this echoes the inability of the ethical to leave behind the aesthetic.

What this shows is that Kierkegaard's dichotomy has very striking parallells in other philosophies. The big difference is that Kierkegaard tries to unite both types of approach, whereas the more usual thing has been for philosophers to latch onto one or the other. One exception to this is Kant, who distinguished between persons as empirical objects and transcendental subjects. But the big difference between Kierkegaard and Kant lies in the fact that Kant concluded that the two ways of viewing persons had to be kept separate, whereas Kierkegaard attempted to draw them together. What is also interesting is that Kant explains the so-called contradictory nature of persons by distinguishing the different ways of looking at persons. As has been remarked, it was necessary for Kierkegaard to make us see the aesthetic and the ethical from the inside. This also accords with the idea of there being different ways of looking. The paradox arises when we try to see both at the same time from a neutral viewpoint.

The parallells I have suggested here between Kierkegaard's thought and those of perhaps more familiar philosophers should neither be over stressed nor ignored. That there are striking similarities is more than just interesting, but this should not be taken to be a claim that the positions are the same. But what makes Kierkegaard unique is that he endeavours to find some way of linking the two approaches. He also expresses with some urgency the importance the issue has for selfhood, ethics and rationality. So finally we must turn to see how this "impossible" reconciliation is made.

4. Kierkegaard's Solution.

Kierkegaard failed to find satisfaction, intellectually and personally, in the either/or of the aesthetic and the ethical. So what could the next stage be? The answer is the religious, and although to explain this will initially divert us from the self, a proper explanation requires this short digression.

The religious enters the fray when we consider the limits of the ethical. The attachment of the ethical to the eternal comes through the telos of the ethical system, an absolute end to which all points in the system have their reference. Kierkegaard's most thorough examination of the ethical comes in *Fear and Trembling*, a work which Kierkegaard himself saw as his most important. It is, like all his pseudonymous works, described as an "aesthetic work". This is because the ethical, although concerned with the eternal, actually fails to reconcile itself to the eternal, and it is therefore examined through the eyes of the aesthetic. Kierkegaard's paradigm case for this is Abraham's attempted sacrifice of his son,

[10] Descartes [1641], para 14

Isaac. The central problem for the book is to explain how this sacrifice can be justified when the ethical maxim is “thou shalt not kill”. Kierkegaard writes “The ethical as such is the universal, and as the universal it applies to everyone, which from one angle means it is applied at all times. It rests immanent in itself, has nothing outside itself that is its telos but is itself the telos for everything outside itself”.[11] On this definition, Abraham’s act was one of attempted murder, pure and simple. There can be no question of it having an ethical validity, as ethics is universal and therefore exceptionless. The answer to this apparent problem is to say that we must accept Abraham not as a paradigm of morality, but as the paradigm of the man who goes beyond ethics altogether and thus overcomes the insoluble paradoxes of the aesthetic and the ethical by entering the religious sphere. Hence the religious is not a form of the ethical, but a distinct sphere of existence.

To show that religion goes truly beyond ethics, Kierkegaard asks whether there can be a “teleological suspension of ethics” that can allow for Abraham’s actions. On this account there cannot. The reason is that if Abraham had to suspend the ethical system in order to fulfil its highest telos - obedience to God’s will - then the ethical is not an ethical system at all per se. It would lose its universality, which is essential for any ethical system, and would merely be a tool to some higher end. Hence ethics looses its universality, leaving it as an incomplete system that requires something more, the religious, for its fulfilment.

The author of *Fear and Trembling*, Johannes De Silentio, cannot express what Abraham did in ethical terms. Hence the pseudonym - De Silentio; mute in the face of a deed that goes beyond the ethical. This is because the deed can only be understood from the religious sphere. Note once more the inability and futility of trying to analyse the deeds executed in one sphere from another. The reason why

[11] Kierkegaard [1843b], p54

it is inexplicable is that it can only make sense by appeal to an absolute that lies beyond the temporal and hence beyond the reach of humanity. And yet humanity has to make that connection for the ethical system to have any basis. This is what induces the "fear and trembling" into Abraham and us. The religious basis for ethics can only come through faith and faith is a leap into the uncertain. To return to the ethical without the religious is to return to a castle built on sand. To return to the aesthetic is to dive into a void.

Kierkegaard concludes that having exhausted all possibilities in the ethical and the aesthetic, only the religious is left. But if the other two are exhaustive, what is left? With two elements, the finite and the infinite, there are four possibilities; one or the other, neither or both. For Kierkegaard, having accepted the reality of both, the only solution can come through a synthesis of the finite and the infinite. Logically, this is an absurdity. But Kierkegaard believes that it is an absurdity we are forced to confront. The aesthetic and the ethical logically exhaust all possibilities, and yet neither one can provide the complete account required. Therefore we must go beyond what logic can offer if we are to complete our understanding of ourselves and our lives. The next question then is, what is this thing that we can choose that is both finite and infinite? This is the religious sphere. Religion, for Kierkegaard, meant Christianity, and it is certainly a criticism of his work that no other religious doctrine was even considered. By the standards of rationality, religion is absurd. Jesus Christ was supposed to be both man and God. But how can anything be both finite and infinite, historical and ahistorical? It is precisely this absurdity, however, which for Kierkegaard marks the religious life out as the true way forward for man. Man considered aesthetically is incomplete, man considered ethically lacks the necessary finitude. But in Christianity we have a meshing of the two, a reconciliation of both diverse elements into one world-view. However, the price of accepting it is that it lacks rational foundation. To become a Christian takes a true leap of faith. Reason says it to be absurd. Nothing can prove it to be true. This is why, Kierkegaard believes, Abraham’s deed is the supreme act of faith which commands such respect from Christians: Abraham had the faith to do God’s will, even though it mean transgressing the objective moral law. Although anyone would do something if God actually asked them to do it, how could one know it was God? Could it be some mischievous demon? Think of all the crimes, the Yorkshire ripper cases being most notorious, that have been committed by people who claimed God had told them to do it. How does Abraham know he is different to these deluded psychopaths? He doesn’t know, he simply makes an act of faith. Any Christian who has said, “I take it all on faith“ and smiles should read Kierkegaard, and know that the correct reaction to faith is not a smug smile, but Fear and Trembling.

How does this apply to the self? In *The Sickness Unto Death* Kierkegaard discusses this by considering what he calls “despair”. But this is not what we ordinarily mean by despair. It is rather an existential despair that comes from contemplation of what we are. One of Kierkegaard’s formulations of this is:

> Despair is a sickness of the spirit, of the self, and so can have three forms: being unconscious in despair of having a self (inauthentic despair), not wanting in despair to be oneself, and wanting in despair to be oneself.[12]

Why should anyone despair of being oneself? More puzzling yet, how can anyone be unconscious of having a self? The answer to both these questions comes from the idea that we are often in ignorance of what we ourselves really are. But then to realise what we are is not to end the despair, but rather the beginning of a more authentic despair, as the reality is hard to accept. Continuing the passage above, Kierkegaard explains this reality:

> The human being is spirit. But what is spirit? Spirit is the self. But what is the self? The self is a relation which relates to itself. ... the human being is a synthesis of the infinite and the finite, of the temporal and the eternal, of freedom and necessity. In short a synthesis. A synthesis is a relation between two terms. Looked at in this way, a human being is not yet a self.

Expressed more briefly in *The Concept of Dread*, the thought is:

> [Man] is a synthesis of soul and body which is sustained by spirit.[13]

This is an interesting variation on traditional Christian dualism, in which a third element, spirit, is added as the relation between the body (the aesthetic) and the soul, (the ethical). However, this spirit is not a third 'thing', it is rather the relation itself.

There is thus a sense in which selfhood is something to be attained, as is suggested by the line "a human being is not yet a self". Ergo, spirit is something not automatically resultant from the combination of soul and body, as spirit is the self. Indeed, Kierkegaard's main criticism of the Denmark of his time was that it lacks spirit[14]. Throughout *The Sickness Unto Death*, Kierkegaard describes true attainment of the self as being reached through increased self awareness, culminating in the realisation that "the heightened consciousness of the self is knowledge of Christ, a self directly before Christ".[15] It is only when we are aware of ourselves as being before God and Christ that we can truly understand what we are. It is important for Kierkegaard that entering into the religious sphere is necessary for true self understanding. The discussion of *Fear and Trembling* is his explanation why: only Christianity is able to attain the synthesis between the ethical and the aesthetic. Without Christianity, it is simply inconceivable that a

[12] Kierkegaard [1849], p43

[13] Kierkegaard [1844], p44

[14] See prologue to Hall [1993]

[15] Kierkegaard, [1849], p146

synthesis is even possible. But once one accepts Christianity, the synthesis becomes possible.

It is for this reason, according to Kierkegaard, that man needs to enter into the religious to truly understand his own nature. By accepting the contradiction inherent in religion, man accepts the contradiction inherent in himself and then can truly be said to have found spirit, the true way of synthesising soul and body. The state of man prior to achieving this synthesis is summed up in the expressions “Finitude’s despair is to lack infinitude”[16] and “infinitude’s despair is to lack finitude”.[17] The former case is easier to understand. The despair of the person who considers themselves only aesthetically is caused by failing to get a grip on any permanence within themselves. The second case is perhaps subtler. If we consider ourselves as eternal, then we are, in a sense, denying a part of what seems to make us really ourselves. Our lives have chiselled out the personalities we now have, our desires and projects are based around our limits. This is why so many religions make sense of eternal life as being some kind of dissolution of the ego. Kierkegaard puts it succinctly in *The Sickness unto Death*:

> The self is simply more and more volatized and eventually becomes a kind of abstract sensitivity which inhumanly belongs to no human, but which inhumanly participates sensitively, so to speak, in the fate of some abstraction, for example, humanity in abstracto.[18]

The aesthetic/ethical distinction thus leads to an impossible two pronged dilemma. On one prong, you’re trapped in the moment, on the other you loose individuality. Either way, you’re not a complete self. That can only come through the religious.

[16] ibid, p63
[17] ibid, p60

5.Creating a Self.

To show why this is of relevance, I would like to offer an interpretive summary of what we have just discussed. In contemporary philosophy, one would be expected to clearly differentiate what was meant by the different terms; person, human being, man and self. Indeed, the critical reader may have questioned the apparently slack use of these terms in this section. However, it is not at all clear that Kierkegaard himself applied his terms with the degree of consistency we would hope for. However, one distinction which does seem to be clear is that between "human being" and "self", as brought out by the phrase in *The Sickness Unto Death*, "a human being is not yet a self". Kierkegaard talks of the self as a synthesis of two elements: the aesthetic (the finite and the body) and the ethical (the infinite and the soul). Two elements combine to form a third, the self. But the self is not an automatic result of the combining of these two. A person contains both within them. But becoming a self requires action on behalf of the person, based first on despair and then entering religiosity. In this way, we can see that the ethical and the aesthetic are the givens, but the self is created.

The given is simply a person. In order to truly become a self, a person has to synthesise the aesthetic and ethical, elements within themselves. But what does it mean to say that prior to this synthesis there is no self? Kierkegaard is clearly talking about some form of high-redefinition of what the self is, it cannot be what is ordinarily meant by that term, that is, something like a being, conscious of itself and its environment. He says, for example:

> The biggest danger, that of losing oneself[19], can pass off in the world as quietly as if it were nothing; every other loss, an arm, a leg, five dollars, a wife, etc., is bound to be noticed.[20]

[18] ibid, p61

[19] More accurately, failing to attain the self, as one cannot lose what one does not yet have.

Self, for Kierkegaard, is rather the full recognition of what a person is, a recognition that for him can only come through recognition of being a self before God. Becoming a self is therefore a kind of life project, something that should be aimed for.

However, this seems to be getting ever further from what we are doing in the present debate over persons. When we ask the question, “What is a person?” we are asking either a factual or conceptual question. Either way, the analysis is of the given. It would certainly cause puzzlement if we were to suggest that the philosophers job was to construct a person. But if Kierkegaard is right, then the project of simply analysing the given is going to hit a dead end. We cannot simply answer the factual question because there are two aspects of the given person which are in contradiction. The enterprise will end either missing one half or in contradiction. So in order for our understanding of the notions of person or self to be completed, we will have to find some way of reconciling the two elements. What was a life project for Kierkegaard becomes a philosophical project for us. Kierkegaard saw it as uniting two elements to become a self; our project is to unite the two elements to give an account of the self. What Kierkegaard saw as a matter of existential urgency becomes a matter of intellectual urgency. I do not mean to suggest that the two can be made totally separate. Harrison Hall, in this interpretive summary of Kierkegaard’s thought, makes it easier for us to see the connection between the existential and intellectual concerns:

> The human self consists, in part, of two opposed sets of needs: needs for the finite, the temporal and the necessary on the one hand, and needs for the infinite, the eternal and the free or possible on the other...one

[20] Kierkegaard [1849], p62-63

> solves the problem and becomes a self in the full sense only if he recognises and satisfies all of his contradictory needs.[21]

Existential satisfaction requires recognition of what our needs consist in. But unless we can make sense of what we are, of why we have these different sets of needs, then it is difficult to truly comprehend what these needs are. Intellectual satisfaction is one of those human needs.

Would not satisfying this need to make sense of two apparently mutually incommensurable ways of looking at the self be best met by junking them and finding a third, more consistent one? Not exactly. We can't just junk the aesthetic and ethical viewpoints. They are not constructs but givens. I have revised somewhat the sense in which they are givens. Firstly, I have argued that, ontologically speaking, the ethical need not be seen as necessarily eternal, but is best viewed as being that which goes beyond the immediate. Secondly, although I take it that it is given that we view ourselves and the world in these two ways, that does not mean the world necessarily is both aesthetic and ethical. What we need is a way of understanding that allows us to keep hold of the necessity of seeing the world in both these ways. The self conceives of itself and the world both ethically and aesthetically. The question is not whether this is true but how can this be true? We start from what is true, the Kierkegaardian givens, and have to answer how they can be true. If we are to avoid contradiction, the answer must take one of the following forms: Only one is really true of the world, but the other is a necessary illusion in our conceiving of it, which means we cannot but see it as true. This works rather like the Müller-Lyer illusion, where two lines of equal length appear as different lengths. They are the same length when measured, different when looked upon. When we understand that the latter is simply an inescapable illusion, we can understand how, while both must seem true to us,

[21] Hall [1984], p18

one is really an illusion. Another solution states that neither are true of the world and that both are necessary illusions. This solution is clearly less economical than the previous one, and in general it is best not to postulate more phenomena to be illusions than is absolutely necessary. A third solution is that both are true, but whereas they necessarily appear to apply to the same thing, in fact they apply to different things which we confuse. A fourth solution is that both are true and apply to the same thing, but to different aspects of that one thing.

6. Conclusion.

Kierkegaard does not offer any conclusive arguments as to why we can consider persons as both aesthetic and ethical beings, but he rightly spotted that there are these two conflicting ways of looking at the self and that the aesthetic and ethical characterisations of the self follow from these different approaches. I have given the Humean and Cartesian accounts of the self as examples of the two ways of seeing the self in action. Kierkegaard's great contribution to the debate on selfhood is not only to point out these two different views but also to make their resolution a fundamental requirement of any explanation of what a person is.

The conclusions of this section are three: (1) A person can be seen as both aesthetic and ethical, and this is the given in that a person will be seen in this way simply by virtue of being considered from a certain viewpoint and not through being understood in a theoretical way. The most well known paradigms of this are the Humean and Cartesian accounts given above, in section 3. There is a kind of necessity in these viewpoints in that, no matter how we may go on to understand the self, seeing the self in these ways will still be unavoidable, so long as we adopt the relevant viewpoint. (2) As the self can appear to us in both the aesthetic and ethical ways, and that these views exclude each other, to explain the truth in both these positions, understanding must go beyond the given. This means an account of the self cannot simply describe the self as it appears to us, as it appears to us in a contradictory way. (3) Such an account may reject one or both of the aesthetic and ethical, but is obliged to explain why it is we are forced to see ourselves in these ways. This obligation is what I call the Kierkegaardian requirement. Any account of persons which does not explain the aesthetic and ethical in persons does not meet this requirement and consequently, it leaves too much unexplained for us to be able to embrace that account.

Explaining the Kierkegaardian requirement and why it is important is a more lengthy and difficult task than seeing whether a theory meets it or not. Furthermore, it is best employed to judge a completed theory. For these reasons, the requirement will not be used until the end of this thesis when we look at my critical development of psychological reductionism and pass judgment on it.

Our first requirement has thus been formulated. I now turn to the second.

Chapter Two

The Relevance Requirement

Most recent work on persons in philosophy has focussed on the issue of personal identity over time. This is usually understood as being the search for the set of necessary and sufficient conditions for a person X at one time and a person Y at another time being the same person, or a demonstration of the impossibility of formulating these conditions. Conditions of identity are sought, and only if an impasse is reached with identity are other ideas of, as Parfit puts it, “what matters in survival” examined. This approach seems natural. When we consider the future, we generally want to know if I will survive and this seems straightforwardly to be a question as to whether or not myself and any future person will be the same person.

However, we have to be careful when we undergo this investigation not to lose sight of our primary concern. If our primary concern is not identity, but persons, then we have to be careful as to how relevant the concerns of identity are to our endeavour. To put it simply, a study of personal identity which is focused on the personal should not emphasise identity at the expense of the personal. And if it could be shown that any investigation was indeed telling us more about the nature of identity than the nature of persons, we should reconsider the importance placed on personal identity. In this chapter, I claim that there are indeed reasons why sometimes, when we appear to be interested in personal identity, in fact, it is not identity which is our primary concern. In my discussion I distinguish between survival and identity. I take personal identity over time to be token identity of the individual person. Any view which abandons token personal identity in favour of type personal identity, abandons the idea of personal identity as it has been understood historically. Survival, however, is a broader concept. Something can be said to survive not only if there is a token identical continuer of it, but also sometimes if there is a type identical whole or part continuer of it. I explain in more detail the differences between these different sorts of identity in section one.

There are two types of cases which support my claim. Firstly, I would like to consider hypothetical cases of fission. I contend that in such cases, considerations of personal continuity and survival are more relevant to our concern in persons than considerations of personal identity are. By considering fission in terms of identity we are lead to solutions which are either indeterminate or unsatisfactory. Much more satisfactory solutions can be offered if we abandon talk of identity altogether. Secondly, there are cases when the nature of identity over time in general requires us to deny personal identity where a good claim could be made for personal survival. Here, there is a solution to the problem which is clear for the identity theorist, but which leaves other problems unresolved, problems of great importance when we consider our survival. What these illustrate is that there are at least two sorts of cases in which judgments concerning personal survival are best based on considerations other than those of personal identity. The fact that this is possible shows that personal survival and personal identity can be considered separately.

These arguments exemplify the importance of what I call the Relevance Requirement. This is because what my argument fundamentally depends upon is the idea that if any account of persons is to be sufficient for our purposes it has to be directly relevant to our interests in persons. When we talk about personal survival, identity or continuity we are not engaged in a merely abstract debate, but are dealing with concepts of direct relevance to the way we see ourselves and our place in the world. It is not enough to provide a ‘solution’ to the problem of personal identity if that solution fails to address our concerns as persons ourselves in our future existence. To talk of identity if identity does not pertain to what is most relevant to our interests is thus to fail to meet the relevance requirement. These ideas and the reasons for holding them should become clearer as the arguments of this chapter progress.

Two things are necessary for the line of argument outlined above. Firstly, we need to go back right to the start and ask just what my present concern is in the debate over personal identity. Secondly, it is necessary to look at how the concept of identity over time works in general. By doing this at some length in the first two sections of this chapter, the arguments of the subsequent two sections can be presented more easily.

1. The Issue.

We must make clear what the concern of this thesis is. If the concern were for personal identity, then clearly I could not deny that identity is what matters. But I am discussing personal identity for a purpose and not just for its own sake. It is this purpose, this underlying interest that the subject has for us which I want to get clear on. Parfit appears to try to address this fundamental interest when he uses the expression “what matters in survival”. But he fails to make clear just what this means, which frustrates many readers who wish to know in what way and for whom “what matters” matters.[22] There is not a great deal in the literature, therefore, to help us define this basic interest.

There are, of course, many things about survival which may concern us. Reputations, legacies, families, our own health and soundness of mind, to name but a few. Obviously, the debate over personal identity cannot address all these concerns. But nor is it very illuminating simply to state that the debate is about identity or survival. So perhaps it is best to start with the question, just what is it about personal identity which seems to be so important and of such interest to us? We, as persons, are beings who are aware of our existence over time. Our existence over time is something which we usually take for granted. We also take for granted the maxim that ‘people change’. There are many differences between a person at one time and a person at another, distant time. If there is an afterlife, these differences may well be even greater. Senile dementia also preys on our minds, threatening our sense of self-identity. And with the advent of artificial limbs and brain surgery, there are future prospects for quite extensive change in our material make-up. Even now, with drugs such as Prozac, we are confronted with issues of how much we can be altered by circumstances and technology and still be us. All of these are real issues which are truly puzzling. Add the thought experiments of brain-swaps, teletransportation and fission[23] and the possibilities are bewildering. When we consider ourselves undergoing these various changes, the natural question to ask is, “will that person still be me?” However, this question is slightly ambiguous. There are at least three different interpretations of the question. Firstly, and most obviously, this is a factual question of identity. It is either true or false that a person at time X and a person at time Y are numerically the same person. However, it is far from obvious that this can always be a matter of fact.[24] Given that people can change to many different degrees, there are reasons for believing that in some cases, there is no factual answer. Even if there is a factual answer, there is still a second interpretation of the question: Should I, now, and that person, then, consider the later person to be my survivor, who I have reason to care about just as much if he were me, regardless of whether we

[22] I try to dispel this confusion in chapter six, section 5.

[23] See chapter three for a description of these various thought experiments.

[24] The most convincing argument for this is Parfit’s discussion of the combined spectrum. Parfit [1984] pp236-243.

are numerically identical? I call this the first-person question of survival. It may be the case that the facts dictate that there are two, not one persons in this situation, but that the connection between the two is so strong and of such a kind that the persons involved cannot but think of themselves as the same in all important respects, no matter what the facts concerning their numerical identity are. In section four I argue that this is the case in teletransportation. Related to this question is the third person question of survival: Should other people consider myself and that future person as if we were numerically the same person?

On some views, these three ways of putting the question all hang together. That is to say, the reason why I and others should (or should not) view a person at time X and a person at time Y as if they were the same person is that as a matter of fact they are (or are not) the same person. However, my claim is that, although there is a factual question of identity, when many of us consider the issue of personal identity, it is not this question, but the first person question of survival which really is of concern to us. When we consider the various transformations that can occur, we want to know, if that were to happen to me, should I have the same concern for that resulting person as I would for a numerically identical survivor? This is why the problem of personal identity is of interest to those other than philosophers interested in the logic of diachronic identity: because we are persons ourselves who are interested in how we should view our own future and past existence. My intention in this thesis is to weaken the link between the factual question of identity and the first person question of survival and to show that the second question is the more pertinent to the general concerns of persons. Put another way, I intend to show that our view of personal survival does not entail personal identity. In this section I show that personal survival does not *necessarily* entail personal identity. In the remainder of this chapter I show how addressing the question of personal identity may not address the first person question of survival, and that if this is so it fails to address the concerns relevant to this study. And in the thesis as a whole I develop a view of personal survival from the Parfitian conception which attempts to address the first person question of survival without needing to answer the factual question of identity.

Firstly then, how can it even be possible for the first person question of survival to be separated from the factual question of identity? To begin to answer this, we must distinguish between token identity and type or qualitative identity. If X and Y are token-identical it means that they are materially and qualitatively the same, and that everything that is true of X is true of Y. It is the form of identity that Leibniz's Law deals with. If two things are type-identical it simply means that they are qualitatively identical. Some believe that this is a real, absolute identity of universals whereas others claim it is more of an identity by analogy to the real identity of tokens. It doesn't matter for us which view is correct. Type identity is usually thought of in terms of whole objects. For some reason, billiard balls seem to be the paradigm. Two billiard balls of the same size, colour and construction are type identical. But type identity can equally hold with respect to parts of objects. The vitamin C in an orange is type identical to the vitamin C in a potato, despite differences in distribution, concentration and so on. Type identity is also a very broad notion in that it doesn't only apply to measurable, material objects. It is possible for two people to have type-identical political beliefs, or senses of humour.

These features of type-identity are very important when we consider how the importance of identity can depend on our particular interest in each case. If we set before us a plate of potatoes and an orange, obviously the orange is not token identical with the potatoes. Nor are they type identical food items. But now let us consider the situation from the point of view of a person who wants a certain amount of vitamin C. There is a type identity of the vitamin C in both foods. So, all other things being equal, for the purposes of this person, there is nothing to choose between the two different foodstuffs. For her purposes, they are identical. All other features of the foods can be disregarded and what we are then left with is equal amounts of vitamin C. For this reason, we often utilise something like the following principle:

> The Principle of type-identical parts: Two non-type identical wholes that contain type identical parts can, ceteris paribus, be treated as type-identical from the point of view of someone whose interest in those wholes lies solely in those type-identical parts.

When I say “can be treated as type-identical” I mean that the two wholes are as substitutable for each other as two type identical wholes are. The ceteris paribus clause is also important. It will, of course, make a lot of difference if, in order to get the same quantity of vitamin C, a person has a choice between eating half an orange and a kilo of raw fish, or if a person doesn’t like oranges.

This principle is a kind of lazy shorthand we use, when strictly speaking there is no identity of wholes, but because we are only interested in certain parts, we treat the wholes as identical. We should now look at how this principle operates in practice. Imagine someone whose love of Britain is based on a love of the British national character, which he thinks consists in our sense of humour, losing gallantly at sport and the enjoyment of fish and chips. Here, the whole is Britain and the parts, those items on his list. This person could also be a staunch Euro-federalist and support the dissolution of the British state into a European super state. That region that remained Britain would be neither type nor token identical with the Britain of today. But as long as the remaining British region continued to have its sense of humour, fish and chips and gallantry in defeat, ceteris paribus, that person would have no reason to prefer the old British state to the new

European super-state. Indeed, he may be prone to say, “It’s still the same old Britain.” Similarly, a former member of the Social Democratic Party could feel the party’s identity remained after the merger with the Liberals. As long as the distinctive features which that person felt vital to the SDP continued with the Liberal Democrats, it can make as little difference to her which party it is as it does which tea-bag she picks out of the box each morning. Again, she may be prone to say, “It’s still the SDP, really.” In both these cases, we have two non-type identical wholes (British state/British region; SDP/Liberal Democrats) with type identical parts (national character; distinctive policies). Therefore to the person whose interest lies in the type-identical parts, the different wholes can be treated as type-identical.[25]

Such a principle may seem not too important philosophically, as it is no more than an elliptical way of referring to type-identity of parts. But it does reveal something quite important about the way we think. The fact is that we are prone to talk about wholes as identical even when it is only type-identical parts we are really concerned about. Someone going to a certain restaurant for the first time in years who says, “same old restaurant,” does not need to be told that, in fact, much of the decor and menu has changed quite considerably. Only certain features need to be type-identical for her to think of it as the same old place. Could it be possible that when we talk about our survival, we are also talking somewhat elliptically? This seems to be the implication of Parfit’s conception of personal identity, and if so, understanding this could neutralise certain criticisms directed at Parfit’s claim that personal survival consists in Relation R –

[25] There is also the analogous principle of type-identical effects: Two non-type identical wholes that have type identical effects can, ceteris paribus, be treated as type-identical from the point of view of someone whose interest in those wholes lies solely in those type-identical effects. Although I think this is a natural extension of the principle of type-identical parts, I do not require it for my thesis and thus keep it as a separate claim.

psychological continuity and connectedness[26]. Some have tried to criticise Parfit along the following lines:

(1) Persons are of type X. (e.g. human beings)

(2) Relation R can survive transformation of type X to type Y

□ (3) Relation R can survive transformations which human beings cannot.

□ (4) Personal survival cannot consist in Relation R.

The idea is that, if persons just are human beings, but relation R can hold between a person and say, a machine, or between two different persons, then Relation R cannot be all that is required for survival. The problem is that the phrase “personal survival” may be being used here somewhat elliptically. It may not refer to survival of the person, but only to survival of that part of the person which is of importance to us, which Parfit claims is Relation R. To say a person survives a certain transformation may simply mean that there is a type identity of certain vital parts of that person before and after the transformation, rather than that there is strictly survival of the whole person.

This possibility emphasises my point that when we are concerned with personal survival, it is not obvious what precisely this concern is for. It could be for token identity, type-identity of wholes or type identity of certain parts. Without good reasons, we cannot presume that the desire that I continue to exist presupposes one rather than the other. We often have a desire for something to continue to exist when in reality a type identical whole or part will do. X may want his political party to continue to exist, but finds that an apt merger will do just as well. Y wants her £100 to continue to exist, but of course doesn’t mind if the bank gives her new coins. Z wants himself to continue to exist. Why must we assume that this is a desire for token identity of the human being? If it is not, then we may

[26] See Part Two for a more detailed discussion of Relation R

find that our natural propensity to employ the principle of type-identical parts leads us to make misleading statements about sameness of persons, even though these statements are natural and in some sense perfectly acceptable, if taken to be what they actually are: elliptical references for type-identity of parts.

The possibility that a type-identity of parts is what counts raises many curious issues. For example, nothing in such an identity entails uninterrupted continuation of existence. Some of these issues will be discussed in this thesis. For now I only wish to establish that, if our concern in the personal identity debate is for our continued existence, this in itself does not, contrary to appearances, entail that our concern is for token identity.

2. Identity Over Time.

Heraclitus was the first to throw doubt on the possibility of identity over time. The apparent paradox of identity over time is how anything can change and yet remain the same. There are two main conceptions of how an object continues to exist over time. The first view is that objects endure, which is to say that an object exists in its entirety at any one moment in time and so long as it does not cease to exist from moment to moment, it continues to exist. On this view, objects are three-dimensional wholes that move through the fourth dimension of time. The alternative view is that objects perdure. That is to say, objects are four dimensional and thus have temporal as well as spatial extension. Just as a slice of salami is a ‘space slice’ of a longer, whole salami, a stick of salami at one time is just a ‘time-slice’ of a whole salami that exists over a longer period of time. I would agree with Lewis that in fact objects perdure rather than endure[27,] although nothing in my arguments depends upon this. However, it will be easier when discussing identity over time to work on one model or another, so I will discuss identity over time in terms of perdurance. Where it is helpful or important to do so, I offer an endurance view of what I am discussing

Although what follows is not an argument in favour of the four-dimensional perdurance conception, I hope to show its attractive neatness. The following statement is necessarily false:

Fa & ¬Fa

However, the following is not:

Fa at t^1 & ¬Fa at t^2

where t predicates temporal or spatial location. So the deduction:

John is Ugly

Smith is handsome

John _ Smith

is valid, but we cannot argue:

John at t^1 is ugly

Smith at t^2 is handsome

John _ Smith

Adding a temporal or spatial qualifier allows us to attribute opposing properties to one single object. In the spatial case, the reason for this is obvious. Take for example the Greenwich meridian. It is true both that the Greenwich Meridian at 55^0 latitude is in the North Sea and that at 45^0 latitude it is in France. If we had two different singular terms, (The Greenwich Meridian at 55^0) and (the Greenwich Meridian at 45^0) then the fact that one is in France and one is in the North Sea would mean that they could not be the same thing. But, of course, this is not the case. There is rather one meridian which has a part in France, which is not identical with another part, which is in the North Sea. The two parts are non-identical, but this does not imply the existence of two different wholes. By

[27] *On the Plurality of Worlds*, 4.2

accepting that a spatial object has parts, it is easy to see how one object can have two conflicting properties, just so long as those properties are assigned to different parts.

The perdurance view treats time the same way as space and attributes to any four-dimensional object temporal as well as spatial parts. So, John can be ugly and Smith can be handsome, yet John can be Smith. This is because John Smith can have one temporal part, which is an ugly pre-pubescent and another temporal part which is a handsome young man. This is neat, but it does entail a strange ontology where what exists at any moment of time is not a whole person but merely a time-slice of a four-dimensional person. The question of identity over time is therefore not “What makes X and Y at two different times the same?” but “What is it that makes two time-slices at two different times part of the same perdurer?”

Before turning back to the particular case of persons, we need to consider the general problem of individuation of perdurers. Perdurers are ordered series of time slices. This means any number of time-slices can be arranged into an ordered series. So there exists an object of which my pen today is one part and the crown jewels tomorrow is another. The universe is filled with such strange objects, but all of these exist as surely as do my pen and the crown jewels. Two time slices can be part of a whole perdurer no matter how tenuous the link between them is.

Fortunately, we have good reasons for preferring some perdurers to others. We can see why by making use of the distinction between natural kinds, which have real essences, and non-natural kinds, which have only nominal essences.[28]

[28] Locke [1694] introduced this distinction, but I am basing my use of this distinction on Wiggins [1988]. In making use of this distinction I am not committing myself to Wiggins’ wider doctrines. It seems to me, however, that any ontology must account for the distinction that Wiggins makes between natural and non-natural kinds.

For a natural kind to have a natural essence means that there are natural laws which explain its individuation and essence. As Wiggins puts it:

> When there is a dispute concerning an object identified under a natural kind, then one can readily conceive of getting more facts. [...] Identity questions about members of natural kinds can be expected to find the notion of identity at its best.[29]

In other words, there is something which it means to be a natural kind regardless of how developed our understanding of that thing is. On the other hand, something has a nominal essence if it is individuated or classified under functional descriptions. There are no natural laws governing what things are clocks or cars and so likewise there are no natural laws governing their identity as clocks or cars. In this case, if we are in disagreement about a non-natural kind, the dispute is not likely to be solved by getting hold of more facts. To see how this affects identity conditions, some examples will help.

Things which belong to a natural kind, such as Tibbles the cat, form natural four-dimensional 'worms' and we can give good non-arbitrary specifications for what it would mean for a time-slice to be part of Tibbles. These conditions would doubtless involve contiguity in time and space and certain causal relations. In this way, we can see how there is a difference between gerrymandered perdurers such as the 'pen-jewels' and natural objects, such as cats. Only those time-slices which are linked in the necessary nomic ways can be apart of Tibbles. This also rules out the possibility of Tibbles having any gaps in her existence. If there is a period of time where there is no Tibbles then that means that perdurer has come to an end, and any perdurer that turns up at a later date will have to be a different one. Thinking along spatial lines will help us again here. A petal in Edinburgh and

[29] Wiggins [1988], p159

a stem in Glasgow cannot be part of the same whole, living flower, unless this is one very large flower![30]

But for anything that is not a natural kind, identity is more problematic. Take Hobbes' famous example of Theseus' ship.[31] This ship is put into dry docks for repairs. Parts are removed and replaced. It is a long process, but actually, by the end of the process all of the parts have been replaced. As it happens, someone has been collecting all the old parts and has reconstructed the ship from them. So which is Theseus' ship? Here, we are in possession of all the facts concerning the time slices. But several conflicting four dimensional perdurers can be constructed from them, depending upon what we take the nominal essence to be. Firstly, we could say Theseus' ship is whatever ship fulfils a certain role for Theseus, which would mean it is the one that was repaired. Secondly, we could say it is the ship which was originally built at such and such a time and it is the one that was, in effect, dismantled and reconstructed. Thirdly, we could say it is the ship that was originally built up until the time of its dismantling, in which case we would say the ship was destroyed, and that a new replica and a reconstruction were made.[32] There are possibly other explanations. The four dimensional ontology alone does not help us decide which is the correct answer. The way to decide seems rather to depend upon what nominal essence we attribute to the ship which in turn depends on what our interest in the ship is. As an object that fulfils a particular function, it is best to see the ship as having undergone repairs. If we are more concerned with constitution, because, for example, we need to get forensic evidence, the reconstructed original would be the one to go for. If it is ownership

[30] On the endurance view, the identity of natural kinds appears simpler. Tibbles falls under the natural kind 'cat'. For Tibbles to continue to exist over time entails that the three-dimensional cat 'Tibbles' not only continues to exist but continues to be a token of the natural kind 'cat'.

[31] Hobbes [1839], pp135-138

[32] On the endurance view, the same dilemmas arise. But it is not now a question of which perdurer is Theseus' ship but rather under which circumstances the ship ceases or continues to exist.

we are concerned with, it will be the repaired one. If our interest is sentimental, again the original may be more truly ‘Theseus’ ship’ than the other, and so on. No one of these claims is more or less supported by the four dimensional ontology. Their justification is rather dependent upon where our interests lie. Identity over time is not given as it is with natural kinds.

Nor is there the same restriction on contiguity. Consider first a spatial example of a road that has been interrupted, usually by development, but which re-appears on the other side of the development. A temporal example is the Azores High, a high-pressure weather system which is part of Northern Europe’s meteorological map every summer, but which simply does not exist for the rest of the year. These both have nominal essences. The Azores high is that high pressure system which fulfils a particular role at a particular time of the year, while a particular road is whatever stretch of tarmac, concrete, brick or whatever that links certain places. Both the road and the system are individuated by their fulfilling a functional role. This gives us a freedom to legislate concerning their identity which we don’t have with Tibbles. Meteorologists could decide that naming this weather system is misleading, and they could instead give each annual Azores-functioning system its own name. Similarly, it is possible for the council to decide that a road which has been divided should one half renamed. Our ability to legislate is restricted by custom, practice, convention and so forth. But in no such cases is identity determined by natural laws.

We do not need to decide whether this means there is no real identity here at all, although there is an identity of type or universal, which certainly doesn’t contravene Leibniz’s law. What is important here is that only tokens of continuous, uninterrupted natural kinds can have any claim to a natural identity. The identity of anything with only a nominal essence is either a useful fiction, a matter of linguistic, social or conceptual legislation, convention or practice.

What I have so far said has concerned the question of how identity over time in general can be possible. I have concentrated on the perdurance conception of identity, on which diachronic identity is a matter of time slices forming a single perdurer. A natural-kind perdurer must be continuous in space and time and its time slices must be contiguous. A natural kind endurer, on the other hand, is simply an object that continues to exist over time and continues to be a token of the same natural kind. I will start from the assumption that persons have real essences and therefore real identity conditions. I will then go on to show that if this is true, then facts concerning identity will not be relevant to certain concerns we have in problem cases of personal survival. Instead, we have to fall back on more fundamental questions about persons and in doing so, we challenge the view that identity is the main concern of persons in their survival. The conclusion then must be that either persons have only nominal essences or that identity cannot be what is important to persons about survival.

What if persons are non-natural kinds? Would it then make sense to claim that personal identity is important? For non-natural kinds in general, whether or not the time slices form a single object is a matter for one or other of stipulation, legislation, convention, or culture. As we saw with Theseus' ship, the criteria we set down for identity of a non-natural kind depends upon our interest in the object. In the case of persons, the fact that we would have to decide what our interest in persons is before we can establish the criteria of identity already suggests that there are concerns prior to that of identity. But it could still be argued that despite all this, identity is still very important indeed. Consider an analogous case of the embryo and the child. Most would agree that there is no fact waiting to be discovered which would reveal to us when the embryo becomes a full-blooded person. But it can still be argued that it is very important indeed how we legislate, in the conceptual and legal sense, to distinguish between foetus and child. This is certainly a possibility, but in such cases we need to be given additional reasons as to why such legislation is important. It doesn't automatically follow from the fact that identity conditions can be given that these identity conditions are important. Similarly, if persons are non-natural kinds, we would need to know just why personal identity is important. Until such reasons are given, I am happy to settle for the conclusion that there is no reason for supposing that non-natural kind identity is important in itself. There certainly may be reasons we could offer to show why it is important, but even if they are good ones, they would follow indirectly from what our interest in persons is rather than from the fact of identity itself. The importance of identity is weakened, if not refuted. So for the remainder of this chapter, I will discuss personal identity on the assumption that persons are natural kinds.

3. Fission and the Impotence of Identity.

Many recent discussions on personal identity have focussed on hypothetical cases of fission. We are to imagine the possibility of a person splitting amoeba-like into two separate organisms. This has been taken in many different ways to pose problems for personal identity. On both psychological and physical criteria, it seems that neither of the two resultant persons can be identical with the original person. Parfit argues for this view persuasively using the traditional endurance model.[33] The arguments on the perdurance view are very similar. There is a series of time-slices before the split and another series after, and it is of course impossible for any series of time-slices to be identical with any other. The meaningful question that remains is, which of these series are part of a single perdurer? There are at least four possibilities all perfectly compatible with the four-dimensional ontology. Firstly, we can say that what we have here is a single, branching perdurer. All the time-slices, before and after the split, belong to one, single object. Of course, this does not mean that any time-slice on one branch is identical with any time-slice on the other. A second possibility is that the perdurer consists of all the time slices up until the split, and thereafter all the time-slices on one branch only. The second branch can then be seen as a breakaway, separate perdurer. The fact that both branches are equally contiguous with the original is irrelevant, as the same could be said for a mother and child. However, the obvious difference in this case is that there is no significant difference between the two branches. A third possibility is the same, except that we choose the other branch as the continuer. The last possibility is that at the point of branching, the original perdurer ceases to exist and two new ones come into existence.

The problem is that each of these four possibilities is equally supported by the four-dimensional ontology. Indeed, there are more. Lewis' own solution to the problem is to claim that there are two persons present all along. It is simply that the two people share a common stage.[34] This is rather like there being two roads which share common stages. I believe this is actually the case with the A2 and the M2 roads. I do not wish to discuss the merits of Lewis' solution. All I wish to say is that the perdurance view of existence over time no more favours this view than any of the others. Therefore, if we wish to talk about personal identity in this case – understood as concerning the question: which series of time slices form a single person? – there will be several views to consider. We are now in the same position as we were with non-natural kinds. As with Theseus' ship, there are simply different perdurers from which we must choose which one is the single person. Nor does the endurance view make anything easier. In such a case,

[33] Parfit [1984] pp253-266

[34] Lewis [1976] p27

there just doesn't seem to be a natural law which determines which, if any, of the fission products is identical with the pre-fission person.

From this point, I see two ways of proceeding. We can make what I shall call an identity-affirming response or an identity-neutral response. The first of these entails simply taking up the challenge of answering the identity question as framed above. There are a number of different bases upon which we can make our choice between the various competing perdurers. One of these requires us to adopt a stance based upon our interest in persons, in much the same way as we need to decide what our interest is in Theseus' ship before we can make an identity judgment there. What exactly this interest is would be a subject for further debate. But as I argued in section 1, it seems rash to assume that our interest as persons ourselves is in survival of the token. I suggest that our interest is in fact simply survival, which as I said earlier[35] is a broader concept than identity, and so we would have to see which of the competing perdurers ensures the greater degree of survival, without prejudging the need for token, or part or whole type identity. However, my argument does not hinge upon what we decide our interest is in this case and I certainly don't think that what I have said establishes that survival is our real interest. All I wish to establish here is that the strategy of deciding our interest and then determining which ordered series of time-slices constitutes an individual person on that basis is a coherent one.

The problem here though is that we are now in a situation indistinguishable from that of the identity of non-natural kinds. We are put in the position of legislators who are not out to discover which series of time slices is a single individual but deciding which series is to be considered a single individual. The same problems thus occur here as I pointed out at the end of the previous section. Firstly, our main topic of enquiry becomes persons' own interests in their survival and identity is relegated to a secondary concern. And secondly, the importance of identity becomes questionable and requires establishing independently. With the identity of a person resting upon a decision or custom rather than a fact, we need to know why our decision is an important one.

Alternatively, we could make an identity-affirming response based, not on the stance we adopt in relation to persons, but on our prior idea of what a person essentially is. On this view, the function of puzzle cases such as fission cases is to provoke deeper thinking into the nature of persons, so that we revise our view of what persons are until we reach a solution to the puzzle case. So we ask, which of the competing perdurers that can be consistently constructed from this case matches the concept of a person?[36] If we could answer this question, then we would be able to say what is required for a series of time slices to form a single person and would therefore have solved the problem of personal continuity over time. For example, Carol Rovane has argued that what it means for us to be persons entails a unique continuer and therefore concludes that fission, for us as we are now, would destroy a lot of what it means for us to be persons. I do not wish to discuss the merits of this view here.[37] I mention Rovane as an illustration of how resolving the dilemma of identity posed by fission requires us to start from what being a person really means. Having done this, we can then choose which of the perdurance options is best and then come up with an account of identity. But in this case, identity again loses its place as the primary topic of investigation. We do not discuss identity to learn about persons, but we study persons and thus discover their identity. And, of course, it still leaves open the possibility that when we consider what a person is, diachronic identity is not that important for persons.

[35] page 25 (First page of this chapter)

[36] It seems to me that such an approach would entail abandoning the view that a person is a natural kind.

[37] I consider Rovane's views in more detail in §6.4.

So although this sounds like a strategy that will lead to a sensible answer to the identity question, what it really does is leave a lot open whilst shifting the nub of debate away from identity and back to persons themselves.

Now to the second, ”identity-neutral” response. We have already acknowledged that solving the puzzle case requires us to go back to thinking about what persons are and/or what our interest is in them. If this is our starting point, why not ask a question like this one: If I were faced with the prospect of fission, which, if any of the perdurers would I have good reason to view as a continuer of myself? There are possible answers to this which do not require token identity. It is possible for me to consider a whole or part type-identical continuer as a continuer of myself. Of course, in these cases we would not have a single, natural-kind perdurer or endurer, and we are currently working on the assumption that person is a natural kind. But there is absolutely no reason to believe that if we are natural-kinds, the fact that we are such creatures is what matters. In section one, we saw how the sort of continuation of existence we want for things can vary from case to case. Token identity continuation is just one of those. If what is important about being me is preserved in a different form of continuation, then the fact that the continuer is a numerically different person to the one I am part of now need not necessarily matter. While personal identity was the primary topic of investigation, this approach would have appeared radical. But when we consider the fission case, it seems that we cannot but be pushed back to considering other issues surrounding persons and their survival before we can go on to consider the identity question. Having made this step-back from identity, the possibility is opened up that our next step forward may not take us back to identity. Of course, nothing I have said so far shows that identity may not turn out to be important. All I wish to show is that this is far from self-evident. And certainly the fission case offers no reason why it must be important.

To summarise, the dialectical response to the fission case leads to the rejection of the primacy of personal identity. To decide whether there is an identity between pre- and post-fission persons, we need to examine more closely what it means to be a person to solve the puzzle. This is the normal procedure. But, once invited to examine what it means to be a person, we also need to consider what is our concern here, as I outlined in section 1. When we do this, we will question whether there is any need for our continuer to be part of the same perdurer or endurer as the question of identity demands. We will then be able to validly question whether the issue of personal identity is the crucial one. The more fundamental question is, I believe: what sort of continuer satisfies a person's interest in his or her future and past? An identical perdurer or endurer then becomes just one among many possibilities, and we have managed to move away from the idea of the primary importance of identity. Then, even if we do conclude that identity is important, we will have been in a position of demonstrating that it is so and not just assuming so.

4. Teletransportation and the Limits of Identity.

Some may not be convinced that the identity requirement can be eliminated. They would reject the idea that a natural kind can have an adequate continuer which is not numerically the same. They would argue that although the fission case leaves open several possible candidates for the single, perduring person, at least restrictions of identity narrow down the possible candidates for persons. I would now like to suggest that in fact it narrows down the possibilities too much. To illustrate this we must look at Parfit's science-fiction thought experiment, where a person is teletransported from earth to Mars. The facts about what happen are clearly laid out:

> When I press the button, I shall lose consciousness...The scanner here on earth will destroy my brain and body, while recording the exact states of all my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the replicator on Mars. This will create, out of new matter, a brain and body exactly like mine.[38]

The question is, is the person who walks out of the replicator the same person who entered the teletransporter on earth? The answer to this should be apparent from what we have said already: no. In section two of this chapter, I concluded that if persons are natural objects, then requirements of contiguity and connectedness must apply. These conditions are not met in teletransportation. So if identity is what matters in survival it would seem the teletransportation case would hold little interest for us. The simple fact is, the person who walks out of the teletransporter is not the same as the person who walked in. But I would contend that, on the contrary, there is plenty to interest us in the case. Let us imagine that Star Trek is not a work of fiction but an accurate representation of what is going on in outer space. The Starship Enterprise returns to earth and the crew members go to be reunited with their families. But they are stopped by law enforcers who tell them that, since they have been away, philosophers have shown that the teletransporter they have been using destroys personal identity. Therefore, the crew of the ship is not the same as the crew that left earth previously. They are not biologically related to the families they thought they had and the high court has ruled that they have no claims on the property of the original crew. Doubtless the crew would be stunned. From their point of view such a declaration would be absurd. And it is also hard to believe that the families, no matter how intellectually convinced by philosophers, would find it hard not to think

[38] Parfit [1984] p199

of these people as their relatives returning from space. The crew would doubtless conclude that, no matter what lack of formal identity there is between them and the original crew, everything that matters in ordinary survival was present within them and that they have the right to be treated in the same way as those original people.

Who is right? By the end of this thesis perhaps we will be in a position to tell. But there is certainly nothing in the fact of non-identity which automatically dispels the puzzling issues surrounding teletransportation. The reason for this is that we have not been offered any reasons why identity should be all that matters. The problem is that it has been assumed too easily that by solving the problem of personal identity we will have found out what matters about our futures. What this science fiction case shows is that we can think of the identity issue as being solved without giving us all the answers. What if I discovered that I was the result of such a process? Does this mean that I must think of 'my' family and 'my' past differently to others? The fact of non-identity does nothing to eliminate a posteriori the fundamental questions that would face those who underwent, or knew someone who underwent, the teletransportation process. These questions about what matters about being a person and what matters in survival, which I put in section one, still apply. Identity may play a role in the answers, but it is not our starting point. Insistence that identity must be what matters fails to explain why teletransportation seems so problematic.

My argument could be criticised for being emotive. But in this case, we cannot detach the emotive elements in the issue, because the issue concerns what matters for us in survival. If it was identity that mattered, then the fact of non-identity should make us view teletransportation as death. But when we know all the facts, the issue is not so clear cut. So it is hard to see why identity could be all that matters. Again, it may well be part of what matters, but all that means is that it is one factor amongst others and is not in itself the key focus of our study.

5. Conclusion.

In this chapter I spent some time outlining some of our concerns in the personal identity debate and what it means for something to be identical over time. Once I did this, I discussed two cases. In the fission case, in deciding which perdurer should count as a person, or whether there is a single endurer before and after fission, we were forced to question more deeply whether we were right to assume our continued existence depends on a single person continuing to exist. In teletransportation, identity was a clear cut matter, where what matters for us, as persons, was not. What do these cases together show? It shows that identity cannot be assumed to be the crucial factor when we consider our futures. My approach does share some common features with Parfit. In particular, Parfit claimed that the fact that there are cases where questions of identity are empty – which is to say that there are cases where there is no determinate answer to the question, "does X at t^1 = Y at t^2?" – supports the view that identity is not always what is important about our continued existence. My approach differs in two ways. Firstly, because I have argued that the factual question of identity can be distinguished from the first personal question of survival, the former question does not arise only when questions of identity cannot be solved, but rather as a question worth answering in its own right. Secondly, I have argued that even when there is a determinate answer to the question of identity, which I believe there is in the teletransportation case, important questions concerning our first personal view of survival are not necessarily also answered. So even if there could be some way, in the fission case, of resolving the identity issue, there is no reason to believe that this would resolve what is important when we consider ourselves in the fission situation. To sum up, Parfit’s arguments rely heavily on the supposed indeterminacy of identity. My arguments, however, rely not so much on the indeterminacy of identity, but on the logical distinctness of the factual question of identity and the first personal question of survival.

This conclusion has important consequences for how we approach the philosophical issue of persons. Personal identity may well turn out to be something that is very important and is certainly a distinct area of study. But we cannot simply assume that all of what we refer to as “the question of personal identity” is primarily concerned with identity at all. Parfit concluded that personal identity was not what mattered. In my view, what we should be trying to clarify is to what questions identity is relevant and to which questions other factors, such as survival, are relevant. The debate which started with Locke must not just be slavishly continued. We must distinguish the different questions and decide which are relevant to our particular concerns.

As my concern is with the first person question of survival, the relevance requirement can be stated thus: Any account of personal identity which solves the factual question of identity but not the first personal question of survival fails to meet the relevance requirement. This is not to say that the identity question is not important. But as I claimed in section 3, from our point of view as persons ourselves, the identity question does not appear to be in itself the most important one. Unlike the Kierkegaardian requirement, the relevance requirement will be used throughout this thesis as well as at the end to assess the final position reached.

Chapter Three

Persons and Thought Experiments

In the previous chapter I introduced some examples from Derek Parfit's *Reasons and Persons*. As I indicated in the introduction, this thesis is to a large extent a critical development of Parfit's work. Before we look at any of the details of his arguments, we have to consider his general approach to the subject, which several philosophers have found objectionable. They have done so primarily for two reasons. The first is his reliance on the concept of 'person'. What kind of sortal term is this supposed to be? Does it refer to a natural kind, for example? The second is the importance of thought experiments to his arguments. We have already been introduced to the ideas of fission and teletransportation in chapter two. But how much can we infer from these fantastic science fiction stories? Several writers have wanted to deny the validity of both features of Parfit's approach and many, notably Johnston, Wilkes, Williams and Robinson[39], have seen the two 'errors' as being somehow related. Their unease can by summed up in the idea that fanciful thought experiments about ill-defined 'persons' lead us too far away from what we actually are and how we actually do continue to exist. Thus the critique can be seen as an attempt to make the philosophy of persons less speculative and more empirical. In particular, they recommend ditching 'person' in favour of a more naturalistic sortal such as 'human being'.

In this chapter I defend the Parfitian approach against such objections. With such a growing body of criticism directed at these foundations, it is necessary to check if they are sound. First, I describe the thought experiments that feature most prominently in the remainder of this thesis. At this stage I do not draw any conclusions from them, as our present concern is the validity of using them in argument at all, not what conclusions they are taken to support. Then I consider two objections from Wilkes and Robinson against thought experiments before turning to Mark Johnston's critique, which looks more broadly at how thought experiments are used by Parfit in what he calls 'The Method of Cases'. Finally, I turn to the concept of person and consider criticisms made by Wiggins in particular.

1. The Thought Experiments.

(i) Fission.

In Chapter Two we considered a person dividing amoeba-like into two. Some would ask if this possibility is really imaginable. It requires us to allow that a hitherto single, unified conscious life may branch out in two directions. Parfit claims that "psychological continuity has, in several actual cases, taken a dividing form"[40], and that what has actually happened must be possible. This claim is rooted in some interesting actual experiments on sufferers of severe epilepsy[41]. It was found that by cutting the neural connections between the two hemispheres of the brain (performing a commisurotomy), the fits suffered by epileptics could be significantly reduced. But an unintended result was that patients ended up with what appeared to be "two separate spheres of consciousness"[42]. Parfit describes an experiment which, while it was not one of the tests actually done, gives an accurate representation of what can occur:

> One of these people is shown a wide screen, whose left half is red and right half is blue. On each half, in a darker shade, are the words 'How

[39] Johnston [1987], Robinson [1988], Wilkes [1993], Williams [1973].

[40] Parfit, p259

[41] As described by Nagel [1971], Sperry [1966] and Eccles [1965].

[42] Parfit p245. Parfit's interpretation is backed up by Sperry's own reports. "Everything we have seen so far indicates that the surgery left these people with two separate minds, that is, two separate spheres of consciousness." Sperry [1966], p299.

> many colours can you see?' With both hands the person writes, 'only one'. The words are now changed to read, 'which is the only colour you can see?' With one of his hands the person writes 'red', with the other he writes 'blue'.[43]

The startling thing about this case is that each sphere of consciousness seems unaware of what the other side is aware of. Each hand, 'speaking' for one sphere of consciousness only, reveals via their reports that the colour one side of the brain sees, the other does not.

A second empirical discovery Parfit uses is the fact that people can survive the destruction of one of their hemispheres. Parfit takes these two discoveries to show that, despite the fact that we normally do use both sides of our brains, only one is required for us to be fully-conscious human beings. The two hemispheres are usually responsible for different functions. Speech, for example, is controlled by the right hemisphere. But it is possible for the other hemisphere to take over these functions, although this does usually require complete relearning. Parfit also notes, although he doesn't quote his source for the claim:

> It is also believed that, in a minority of people, there may be no difference between the abilities of the two hemispheres.[44]

So, the scientific evidence shows that, each hemisphere can be fully conscious without the other and that the two hemispheres can be separated so as to create two separate centres of consciousness each unaware of what the other is perceiving, at least for some set of experiences. There is at the moment one limit on this. There is only one brain stem, and at the moment it has not been shown that a brain hemisphere can survive separation from the brain stem, or

[43] Parfit, p245

[44] *ibid* p246

that the stem itself can be divided. Parfit does not worry about this possibility, noting:

> Given the aims of my discussion, this does not matter. The impossibility is merely technical. The one feature of the case that might be held to be deeply impossible - the division of a person's consciousness into two separate streams - is the feature that has actually happened.[45]

In the light of these facts, Parfit borrows a thought experiment from Wiggins[46] It is to be assumed in this case that the person involved is one of those people with no difference in the abilities of each hemisphere and is one of three identical triplets:

> My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine[47]

Henceforth, when I refer to the fission thought experiment, this will be the one I am referring to.

(ii) Teletransportation.

This thought experiment has already been described in sufficient detail in §2.4. However, in addition to this 'simple' teletransportation, Parfit also describes the 'branch-line' case. Here, everything works as in normal teletransportation, except that, instead of destroying the original body, the scanner leaves it intact. What we now have is the original person still on earth and an exact replica walking around Mars.

[45] *ibid* p255

[46] Wiggins [1971], p50

(iii) The Combined Spectrum

Parfit develops the combined spectrum thought experiments from an objection to his personal identity thesis put forward by Bernard Williams.[48] Here we must imagine that we have a machine which is capable of replacing certain features and/or constituents of a person. It can do this in one of three ways. The first is to alter the neural arrangements in the subject's brain so as to bring about a change in that subject's psychological characteristics. After such a change, the person is still made up of the same substance. It is just that the substance, specifically the neural connections, have been rearranged. To employ a computing metaphor, the hardware has been reprogrammed with different software. This machine can operate to any degree. That is to say, it can alter 1% of the person's neural connections right through to 100% of them. Assuming that our psychologies are dependent on these neural wirings, it can then alter our psychologies very slightly or change them completely. If we imagine a line which represents the degree of change undertaken, with 0% at one end and 100% at the other, this line represents what he calls The Psychological Spectrum. The machine which executes these changes could rearrange the neural connections in the brain so that they are exactly like those of a different person, or it could rearrange them according to the whim of the operator.

The second machine alters the physical substance of a person, without changing the psychology: a change of hardware that preserves the same software. Thus parts of the body are removed and replaced with exact replicas. This would result in an analogous physical spectrum. There are also two ways of going about these changes. One would be by replacing the original physical parts with synthetic replacements, the other would be by swapping parts between two

[47] Parfit,p254

[48] "The Self and the Future," in Williams [1973].

subjects. The third machine combines physical and psychological changes to produce a combined spectrum. In this case, both 'hardware' and 'software' are replaced to varying degrees, ranging from replacement of parts of the brain and body with different parts right up to a total change of brain, body and psychology. The possible outcomes of this process form The Combined Spectrum.

(iv) Brain Transplantation.

This is self-explanatory. The most discussed example is the one given by Shoemaker[49] :

> Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumours, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson"... When asked his name he automatically replies "Brown." He recognises Brown's wife and family...

These are the four sets of thought experiment upon which much of Parfit's thesis is built. We must now see why many have argued that we are wrong to base any conclusions on them.

2. Two Types of Objection to Thought Experiments.

Kathleen Wilkes and Carol Rovane share similar similar misgivings about what we can learn from thought experiments.[50] They are both worried that thought experiments often leave out the background conditions against which the thought experiments take place. Wilkes offers the most developed version of this

[49] Shoemaker [1963], pp23-24

[50] Wilkes [1993] and Rovane [1990].

objection. She starts by asking us to consider how a scientific thought experiment works, for example, Einstein imagining what someone would see if, as is actually impossible, one could travel at the speed of light:

> Experiments, typically, set out to show what difference some factor makes; in order to test this, other relevant conditions must be held constant, and the problematic factor juggled against that constant background. If several factors were all fluctuating, then we would not know which of them (or which combinations of them) to hold responsible for the outcome.[51]

So, if in a thought experiment we want to be able to deduce something from the altering of a constant, we must be sure that all the other relevant constants are unchanged. We should also be careful that it is genuinely one constant only which is being altered. We cannot justify any thought experiment by declaring that there is only one way in which the thought experiment alters reality, and that is that the world is not as it actually is. Wilkes' accusation is that in many of the thought experiments listed above, such attention to the background constants is severely lacking. Consider the case of amoeba-like fission. Wilkes lists all the unanswered questions about this case:

> How often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be induced, or prevented?...The entire background here is incomprehensible.[52]

Rovane makes the similar point that whether we know we are going to branch or not is crucial to how we would react to the prospect, in fact and in thought experiment[53].

[51] Wilkes [1993], p7

[52] *ibid*, p11

[53] In Chapter Six, especially section 4, I consider how important the differences between fission and ordinary survival are.

Interestingly, this is not an argument against thought experiments per se, but rather a demand that if we are going to use them, then we'd better be sure that we've got our background conditions sorted. Do Parfit's thought experiments meet this demand?

Wilkes' choice of amoeba-like fission for her example is certainly not a random one. Of all the thought experiments it is the most outrageous. But although Parfit does indeed use this case[54], it is not one of the key thought experiments which I described above. In the key cases described above, the control of the background conditions does strike me as very careful. Brain transplantation requires only that we allow one variable from our actual world now - that this operation has been perfected. The experiment requires no other change from our actual world. Fission, brought about by operation, again only requires that certain things are possible in the operating theatre that are not so now. The combined spectrum and teletransportation cases require much more technological development. But even here, to use the experiment, we need only imagine one thing being different from our current world - that the technology exists to perform these operations. These do indeed seem to entail the alteration of one constant and one constant only which Wilkes demands of thought experiments. (But see the second objection below).

Wilkes' and Rovane's objections sound persuasive because they ask us to consider the wider social and personal differences that a world in which we divided, teletransported and who knows what else, would entail. But in all four thought experiments, there need be no consequences for society or people in general. All four can be conceived as one-off possibilities for single individuals. We could imagine that two scientists have been working on these technologies in a private laboratory and only performed the experiments on themselves. The question of their identity over time would still arise even though they lived and grew up in a society where such changes are alien.

So, if Parfit asks "what if I was teletransported?" there is no requirement for him to spell out the wider implications of this weird technology on society for the question is equally valid when applied to the first and maybe only person to use the device.

There is, however, still a sense in which there is room for doubt over the background conditions. We might wonder whether the laws of physics have to be changed in order for us to imagine these cases. If they would, then we might not have a case of just one fact about the world being altered but something more like "several factors all fluctuating". This thought leads onto the second type of objection, based on considerations of imagination and possibility.

Again, Wilkes is the clearest on this issue. She first considers different types of possibility. First, there is logical possibility. Anything which doesn't entail any contradiction is logically possible, such as pigs that fly, or sheep living on Mars. But this puts only a small constraint on what is possible. Most philosophers do not accept that everything that is logically possible is actually, or in Wilkes' terminology, 'theoretically' or 'in principle' possible. This could also be called physical or nomological possibility, but I shall stick to Wilkes' terminology. Take, for example, the case of iron floating in water. There is certainly nothing contradictory about this, but with all we know about iron and water, many would like to argue that this is not in fact a real possibility. Although we can say we can imagine iron floating on water, this is not enough to make this possible. Either what we are imagining is something like iron floating on (something like) water or imagination is simply no guide to possibility at all.

[54] Parfit, §100/101

These issues of imaginability and possibility are complex and involve considerations of meaning and reference, externalism and internalism that neither Wilkes nor I have given much time to. Certainly, for Wilkes' point to be truly conclusive we have to take a lot of other philosophy on faith. But Wilkes doesn't depend upon us reaching any firm conclusions about this. Indeed, the very contentiousness of these issues works in her favour. Her point is really that any thought experiment which does not work within the parameters of what is known by the physical sciences is open to the charge of depending on an impossibility and thus being at best inconclusive and at worst, irrelevant. For example, a thought experiment that requires us to hold that pigs can fly may not to convince us of very much. Whether or not we decide this a real possibility or not, the charge can still be made that this is irrelevant because in our world pigs can't fly.

In essence, this point is an extension of the first criticism about specifying background conditions. Here, the laws of physics are considered relevant background conditions. If a thought experiment flouts these laws, then we may not learn much from it[55], as there is no single factor being altered against the constant background which any form of experimentation requires. An example of this sort of objection at work can be seen in John Robinson's critique of Parfit's argument from fission, who claims:

> Recent experimental work by Colwyn Trevarthen and his associates suggests that such a conclusion [there are two streams of consciousness in commisurotomy patients] is incorrect.[56]

[55] It is possible that a thought experiment might require the laws of physics to be broken, if this is precisely what the thought experiment is concerned with. For example, when Hume discusses causation and the logical possibility of something not causing its usual effect, he imagines the laws of physics not holding. But in the cases discussed above, it is clear that alternative laws of nature are not part of the thought experiments.

[56] Robinson [1988], p325. Eccles also agrees with Robinson against Parfit and Sperry, claiming "The goings on in the minor hemisphere, which we may refer to as the computer, never come into the conscious experience of the subject." Eccles [1965], pp30-31

Here, it is taken as read that establishing the actual possibility of the phenomena of divided consciousness is important for the thought experiment. Parfit is ambivalent about this. Although he uses the commisurotomy experiments to refute the objection that divided consciousness just isn't possible[57], he also says "it can be useful to consider impossible thought-experiments."[58]

In order to clear up the issue of the relevance of possibility to thought experiments, we first need to get clear on the different sorts of possibility. We have already seen Wilkes' distinction between logical and theoretical possibility. Let us now turn to Parfit's conceptions. Parfit distinguishes between that which is "deeply impossible" and "mere technical impossibility".[59] In the fission case, it may be technically impossible to divide the brain, as perhaps the brain stem cannot be cut or divided. But according to Parfit, what is still deeply possible is that consciousness can take a dividing form. In both cases, empirical data can be important. The importance of the commisurotomy cases is that it shows something which might have been deeply impossible to be possible. As long as that's established, if it turns out to be technically impossible to divide the brain, Parfit thinks that's unimportant.

It is unfortunate that there are so many terms used here to describe two types of modality:

(1) Logical possibility. X is possible because it entails no contradiction.

(2) Technical possibility (Parfit) or theoretical possibility (Wilkes). X is possible because it doesn't break any natural laws, even if X cannot be achieved now.

By 'natural laws' we mean more than the laws of physics. Wilkes uses the example that a fish cannot be a whale as being theoretically impossible, on the

[57] Parfit, p259
[58] ibid, p255

assumption that this is not analytically true. Where does "deep possibility" fit in? The most likely explanation is that it is Parfit's term for logical possibility, but because there is some uncertainty over this, I have omitted this term.[60]

Technical possibility can be important for a thought experiment, but not always to all parts of it. Sometimes a thought experiment as a whole can be technically impossible, but the part of the experiment which is truly vital may be technically possible, and that the rest of the thought experiment is merely a device to highlight this one, crucial feature. Perhaps the best way to understand this notion is to consider how science fiction works. Consider H.G.Wells' story, *First men on the Moon*. In this delightfully whimsical tale, man first travels to the moon by using a substance called "anti-grav", which enables their spaceship to escape gravity and rise up to the moon. One could dismiss this story as impossible. It is technically impossible, because there is no technically possible substance "anti-grav". This is true, but then, it could be replied, "anti-grav" is merely a device employed by the author to get his men on the moon. What is still technically possible is that man could go to the moon. What is important here is distinguishing the relevant from the irrelevant aspects of the story. In this case, it could be argued that the mechanism is simply a device and what is of real relevance is the possibility of moon travel.[61]

In the fission thought experiment, the fact that brain division may be technically impossible doesn't matter because branching consciousness is

[59] ibid

[60] It is possible that Parfit uses "deeply possible" simply to mean "technically possible in all *crucial* respects" He says, "Does it matter if ... this imagined case of complete division will always remain impossible? Given the aims of my discussion, this does not matter. This impossibility is merely technical. The one feature of the case that might be held to be *deeply* impossible – the division of a person's consciousness into two separate streams – is the feature that has actually happened." (Parfit, p255). Here, we could substitute "technically" for "deeply" and the explanation would still make sense, the point being that division of consciousness is technically possible, but not by completely dividing the brain. Hence the crucial part of the thought experiment is technically possible, even though the thought experiment as a whole isn't. See present discussion for importance of this distinction.

[61] As a matter of fact, it matters not a jot to Wells' story whether moon travel is possible or not.

technically possible in human persons. The relevant part of the thought experiment is divided consciousness. Parfit believes that the commisurotomy cases have shown that divided consciousness is possible. The point of considering fission is to enable us to consider the consequences of divided consciousness for personal identity as clearly as possible. Thus, even if the mechanism Parfit employs in the thought experiment – brain division and transplant – is technically impossible, the relevant detail of the thought experiment, i.e. divided consciousness, is not. Hence, the thought experiment remains useful. The same applies to teletransportation. Though perhaps technically impossible, the relevant feature of the thought experiments are technically possible, namely, that the physical matter out of which a person is constituted can be changed without effecting their psychologies.

However, such a defence is not enough to save all of Parfit's thought experiments. Consider the combined spectrum. The transfer of type-identical mental states from one person to another is far from obviously technically possible. And even in the teletransportation case, the technology to perform it is far beyond our reach. Wilkes' criticism is that accepting these far-fetched thought experiments as technically possible, even in their essential elements, takes too much for granted. Can we really accept these possibilities so readily when they concern matters where our knowledge is so limited? The criticism is pertinent, but what it really does is to urge caution rather than invalidate the thought experiments. Given what we know about the human body, mind and brain, none of Parfit's thought experiments strike me as being technically impossible even though to actually perform them would require technology at best way beyond our current capabilities and at worst beyond all *human* attainment. But it must be accepted that they may strike others less favourably, and in those circumstances, resolving our differences would be very difficult. Certainly we would have to consult scientific experts.

If it is uncertain whether a thought experiment is technically impossible, is it enough that it is logically possible? Technical impossibility is not automatically an obstacle to thought experiment. Wilkes discusses Einstein’s thought experiment about travelling at the speed of light. Here, the experiment is still admissible because the experiment is not telling us about what it is like to travel at the speed of light but rather something else about the nature of light, speed and time. Wilkes tries to explain how this is acceptable by saying that the impossibility in question is not relevant to the experiment’s purpose. However, she doesn’t explain this concept of relevance in any detail. So what is relevant in the hypothetical thought experiments involving persons? When we imagine what would happen if we could divide, be teletransported and so on we are not so much interested in what that tells us about such hypothetical phenomena, but what it says about ourselves. What is relevant is what the thought experiments reveal about what we think ourselves to be, not what they tell us about what may happen in far-fetched scenarios. After all, Parfit is not of interest because we think we may divide, just as Einstein’s thought experiment is not interesting because we think we might travel at the speed of light. Considering fission is interesting because it reveals interesting things to us about the way we actually are. Without wishing to preempt the remainder of this thesis, we could see how such considerations may tell us something about the importance of one-one relations with future selves, the relative unimportance of bodily over mental continuity and the unimportance of identity. By stretching our concept of a person and asking us to consider ourselves in perhaps impossible circumstances, we can learn something about the way we actually are. In thought experiments, we can isolate elements which we consider important to our survival and see whether they are as important as we believe them to be.

Consider a parallel in the philosophy of language. Wittgenstein famously claimed, “If a lion could talk, we could not understand him”.[62] A talking lion must surely be as impossible as a dividing person. The brain power and vocal apparatus of lions would make it as impossible for a lion to talk as it would for us to divide the brain stem. Indeed, it may be more impossible, as for a lion to ever be able to speak would require changes in lions themselves to the extent that it would doubtful whether we still had the same animal, whereas it is possible that dividing the brain may only require advances in technology. But Wittgenstein’s interest was not in talking lions, but in the importance of shared social practices for meaning. Imagining the impossible case of a talking lion is merely a way of crystallising the issue. In the fission case, we could equally imagine the perhaps impossible case of a divisible brain. But our interest isn’t really in dividing persons but how much importance we give uniqueness, our bodies and our mental lives in personal survival, and therefore how important such factors are in ordinary life.

To summarise: Wilkes’ makes some telling criticisms of the use of thought experiments. Against this attack I have put up three lines of defence. Firstly, on the requirement that the background conditions must be fully specified, I answer that in Parfit’s key thought experiments they are. In all of them, we need only imagine one type of operation existing which does not exist now. There is no reason to suppose that they are not technically possible. But it can also be argued that the thought experiments rely on our accepting certain scenarios as technically possible when there are not convincing reasons to suppose that they are so possible. Varying one constant, for example, may entail a greater departure from our world than is at first apparent. Therefore they may not be technically possible. The second line of defence is that sometimes the thought experiments as a whole are not technically possible, but the crucial elements of the thought experiments, such as the division of consciousness and the replacement of physical parts of the body, are technically possible. This, too, can be called into question, particularly in thought experiments where mental states are transferred. Given our limited knowledge, it is too early to say whether these processes are technically possible. Given this uncertainty, it is unwise to trust their conclusions too much. The last line of defence is to argue that even if they are technically impossible, the experiments are not concerned with dividing people but with our actual selves. By imagining ourselves in even technically impossible situations, we can learn as much about ourselves as Einstein did about the nature of the world by imagining himself travelling at the speed of light, which is also technically impossible.

Wilkes has another sort of objection. She claims that there are plenty of real cases that can furnish our debate, including foetuses, commisurotomy and multiple personality. Given this rich source and the problematic nature of hypotheticals, why bother with thought experiments at all? I think we should judge by results. Her own work eschewing thought experiments, though fascinating, is too inconclusive too often. One feels that the value of thought experiments is that real phenomena can be pushed further to a logical conclusion and in so doing, the issues can be crystallised. All of Parfit's thought experiments have their start in real life. Fractured consciousness, bodily transplants and tissue renewal all exist. To see how important these are, we imagine them going that little bit further. If mental continuity is all that counts, what if there is only mental continuity and none of body (teletransportation)? If identity is what matters, what if there is continuity without identity (Fission)? Even if these are technically impossible,

[62] Wittgenstein [1953], p223

considering them is a way of isolating the factors we consider to be important and seeing if they really are the crucial factors in identity or survival.

3. The Method of Cases.

Mark Johnston's criticisms of Parfit's methodology isn't so reliant on a criticism of the use of thought experiments in general. Johnston is more concerned with how 'puzzle cases', be they actual or imaginary, are used to further argument. Johnston refers to this way of argumentation as "the Method of Cases". Firstly, we must see what this method of cases actually is. Johnston describes it thus:

> Cases real and imaginary are produced. Competing accounts of the necessary and sufficient conditions for personal identity are then evaluated simply in accord with how well they jibe with intuitions wrung from these cases.[63]

One such case is that of fission, described in the previous section. The point of considering such a case is to help decide between "competing accounts of the necessary and sufficient conditions for personal identity". Psychological reductionism gives an account of identity conditions in terms of psychological continuity, while physical reductionism claims the necessary and sufficient conditions for identity are to be found in bodily continuity. In the Brownson case, we are asked to consider who Brownson could best be said to be. The intuitive answer seems to be Brown. Given that he does not have Brown's body, this is used to support the psychological criteria and damage the physical criteria account. Of course, this is not the end of the debate. The supporter of the physical criteria view will try to come up with a counter example, or modify his position to, for example, the claim that it is not all the body which determines identity, but only key parts of it, namely, the brain.

This argument seems paradigmatic of Johnston's method of cases. There is the real or imaginary case, and the evaluation of competing criteria through how they "jibe with intuitions". But Johnston has at least three criticisms of this procedure. Firstly, he claims that it actually presupposes that reductionism is true. Secondly, he criticises the method's reliance on intuitions. And thirdly, he claims it makes ordinary re-identification of persons problematic. These points form part of an ongoing argument in Johnston's paper, but each criticism stands independently of the others. Let us consider each of these criticisms in turn.

For the first criticism, Johnston states that the method of cases can only be viable if we presuppose the reductionist requirement. This is:

> Our grasp of the concept of being the same person should be able to be correctly represented as a grasp of the necessary and sufficient conditions for the application of the predicate "is the same person," conditions that could be cast in terms of statements about continuity and dependence, statements not themselves to be explained in terms of statements about personal identity.[64]

Johnston says that if we choose to give up the reductionist requirement, that means we would have adopted what Parfit calls the "further-fact view".[65] On this view, personal identity is not constituted purely by facts concerning mental and physical continuity. Certainly, it is the case that such facts constitute evidence for personal identity, but they are not that in which personal identity consists. If we adopt this view, then the method of cases loses all validity. In the Brownson case, for example, we are presented with a bizarre situation where two factors which normally constitute good evidence for personal identity – psychological and bodily continuity – are separated. But then there is nothing in the cases that could make

[63] Johnston [1987], p59

[64] ibid, p60

us decide whether personal identity holds or not, as the cases do not describe what happens to that in which personal identity does consist. We may feel the evidence points more towards one way or the other, but it would seem bizarre to reach any conclusions about personal identity from such inconclusive evidence. So for any holder of the further-fact view, puzzle cases offer no hope of deciding the issue.

Johnston is right in that such cases as the Brownson example would hold no interest for the further-fact theorist. Anyone who believes that personal identity is a matter of the continued existence of an immaterial soul, for example, would simply find the descriptions incomplete. However, this is not in itself an argument against the method of cases. Firstly, if we could have already eliminated the further-fact view by some other form of argument, there would be nothing wrong in deciding between competing reductionist views by the method of cases. It is only if the method of cases is the only method available for the reductionist that any inadequacy could be attributed to it. But secondly, there seems to me to be no good reason why the method cannot have a place in the assessment of further fact views. Indeed, one of the great criticisms of further-fact views was given by John Locke, using a version of the method of cases:

> Let anyone reflect upon himself and conclude that he has in himself an immaterial spirit, which is that which thinks in him and in the constant change of his body keeps him the same, and is that which he calls himself, let him also suppose it to be the same soul that was in Nester or Theorists at the siege of Troy. [...] But he now having no consciousness of any of the actions either of Nester or Thersites, does he or can he conceive himself the same person with either of them? [66]

[65] ibid, p62

[66] Locke [1694], Chapter XXVII, Para. 13

Locke had already defined ‘person’ as a “thinking intelligent being that has reflection and can consider itself as itself, the same thinking thing in different times and places”[67] . But the thought experiment with Nester and Theorists is not there simply to show how the person with the same soul cannot be the same person given Locke’s definition. It is rather that this example provides evidence in favour of Locke’s view over that of the soul view. What it suggests is that, whether or not there is some further fact concerning ourselves, it is wrong to place our identity in this further fact. In this way, it is a perfect example of the method of cases at work; an imaginary case is produced and the two competing accounts of personal identity are judged as to how well they fit our intuitions about the story. And there seems no way in which the further-fact theorist can simply claim that this argument does not apply to him.

So, not only is it possible for the method of cases to be of relevance to the further-fact theorist, even when this is not so, this cannot be a criticism of it unless the reductionist uses it as his only tool. So much for the claim that the method of cases presupposes reductionism. What about Johnston’s second claim, that the method relies too much on intuitions? Before discussing this further, it is worth making the general point that we must not overstate the role of intuitions in this methodology. Consider Parfit, whose work is the subject of much of Johnston’s attack. Two facts should set off a few warning bells for Johnston. Firstly, Parfit claims his view is a revisionary one[68], which surely must entail overturning some intuitions we at present have. Secondly, many of Parfit’s conclusions even Parfit claims to be counter-intuitive.[69] These two facts sit uncomfortably with a view that the ultimate arbiter for Parfit is intuition. The point

[67] ibid, Para. 9

[68] Parfit, Introduction, *x*

[69] “We are naturally inclined to believe that that our identity is always determinate ... this natural belief cannot be true.” ibid, p217

is that Parfit is also looking for overall coherence and consistency. Where intuitions are in conflict with consistency, it is the intuitions that go, not the consistency. Although these comments should make us wary about Johnston's caricature of Parfit and general critique of intuition, Johnston is also making a specific criticism about the results the use of intuitions yield in one particular case.

With these warnings, let us turn to the problem case at the heart of Johnston's critique. He takes up Bernard Williams' conundrum. We are to imagine a machine which is able to record brain states in two different people and then re-programme the brains so that the original brain states of A are now in B and vice-versa. On the assumption that brain states are nomically connected with mental states, the result would be that A and B had swapped psychologies. The effect of this swapping would be the same as that of a brain swap, without the messy surgery. A supporter of the psychological view of personal identity would use this to support the claim that A was now B and vice-versa. This is because, if we were to imagine ourselves as either A or B before the operation, and we were asked to choose which person would be given a lot of money as opposed to being tortured after the operation, we would choose the person who would wake up with our psychologies. To show that this is right, we can imagine that the person who wakes up in B's body would think to themselves "I am A", and would be relieved if he remembered requesting riches for B's body and sorry if he had chosen torture.

But what now if we imagine something slightly different. We are to imagine that we require very painful surgery that cannot be performed under anaesthetic. But what the doctors suggest is that, before our operation, our brain states are read into a computer and the brain states of another person are read into the brain. Then, after the operation, our original brain states will be read back into our brains. Johnston says that:

> A might reasonably retort that he is being asked to undergo a double assault. First, his brain is to be fiddled with in a fairly drastic way so as to produce radical psychological discontinuity, and then he is to be caused to feel severe pain. And this reaction is in accord with the intuition most of us would have about the case.[70]

But, of course, this situation should provoke the same intuition as the original case. For in both of them, according to the psychological criteria theorist, there is no psychological continuity between the original person and the person who, in the same body, is to suffer severe pain. But in fact, the second story wrings out the intuition that our bodily continuity does matter.

What Johnston claims this shows is that, if it is so easy to show how ostensibly the same case can wring out totally different intuitions, then how can intuitions be of any use here? As he puts is:

> How can intuitions be reliable if we can be got to react so differently to the very same case?[71]

The cases are not literally “the very same”, but in both cases brain states are copied and wiped from the brain of the human being before that human being is tortured. At this point we should disentangle two different points in Johnston’s critique. Firstly, there is the local objection that in these cases of exchanges of psychology, intuitions are of little use because, depending on how we tell the story, we can wring out conflicting intuitions. Secondly, this specific case should cast doubt on the general validity of trusting our intuitions. To corrupt Descartes: Johnston has found that intuitions deceive, and it is prudent never to trust completely those who have deceived us even once. It may well appear that Johnston only needs to establish his first, specific point to make his objection

[70] Johnston [1987], p66
[71] ibid, p67

stick. But this is not the case. The fact that intuitions can conflict in a puzzle case leaves us with three possibilities. Firstly, it could be that the puzzle case itself is unsuitable and should be set aside. But unless we reject the use of thought experiments in general, it cannot be right to declare any particular thought experiment unsuitable just because our initial intuitions about it conflict. Given that Johnston does not critique thought experiments in general and given our considerations in section 1 of this chapter, we can set this possibility aside. That leaves two possible courses of action. Either we find some way of deciding which of our intuitive reactions is correct by probing them further, or we attempt to solve the problem without the use of our intuitions at all.

It is clear from Parfit's comments about the very counter intuitiveness of his position that he would go for the first of these courses of action. That intuitions conflict is not the end of the story as there is more to the method of cases than brute intuitions. Nobody who uses the method of cases would want to claim that intuitions were infallible. In a case where intuitions were in conflict, the task would then be to see which of the intuitions has to be eliminated. There are various ways of doing this. If it is the case that the view supported by either of the intuitions is incoherent or contradictory in some way, then that would provide an excellent reason for claiming that one story succeeds in eliciting a misconceived intuition. We could also place the particular thought experiment within the context of other thought experiments and arguments. If it is the case that other thought experiments support one intuition rather than the other, then this would suggest which of the intuitions is more suspect. We could then try to see if either of the intuitions rests on a mistaken belief or a trick of the mind in much the same way as Hume tried to explain how we get the idea of causation.

How could we explain away the intuition that we would have something to fear if our brains were wiped and then that body were to be tortured? Imagine that instead of brain-wiping, the operation will be performed with a general anaesthetic, but at a time before anaesthetic was generally known about. If we explained what was going to happen to A, he may well have a lot of apprehension about what will happen to him. The state of being anaesthetised is simply so alien to him that he cannot imagine being operated on and not being conscious of it. It is hard for us to imagine, being so used to the concept of anaesthesia, just how possible it would be for someone unfamiliar with it to have the terrible intuition that they will not be able to operated on without feeling pain. They too may feel as if they were gong to be subjected to a "double assault"[72] of anaesthesia and being cut open. It is open for the supporter of the psychological criteria of personal identity to explain our intuitions about the second case in a similar way, and to argue that, if we were familiar with brain-wiping technology, we would not react to an impending operation undertaken in this way as a "double-assault". Familiarity would make it as innocuous as an operation under anaesthetic.

Another way of explaining away the intuition would be to appeal to Nozick's idea of the closest continuer[73] . Nozick has put forward a theory of personal identity that claims that I am identical with whatever future person is the closest continuer of me who is not a closer continuer of someone else. The inadequacies of this as an account of personal identity have been pointed out many times.[74] But the view does suggest a psychological point that we will tend to view such a closest continuer as myself unless we have good reasons not to. Hence in the absence of any fully-formed beliefs about personal identity that conflict with the intuition, we will view the mind-wiped future person as ourselves. A Parfitian view may well provide the fully-formed beliefs that will explain why we actually have nothing to fear from such a process.

[72] ibid, p66

[73] Nozick [1981], Chapter One.

I need to stress that these explanations of how we come to have a wrong intuition play no part in deciding which of the intuitions is correct. We do this by examining them for consistency, both internal and in relation with other cases we consider and other intuitions we have. Only then do we do the necessary work of explaining how it is that our intuitions came up with conflicting views. This is how the reductionist should deal with cases where intuitions conflict. What Johnston needs to show is that this strategy is defective. If he doesn't, then the fact that intuitions conflict in itself cannot be an objection to the reductionist programme.

But Johnston does not deal with this head on. Instead, he offers an alternative strategy which he claims is better because it depends on intuitions less. His method has two stages, the first of which concentrates on ordinary reidentification, by which he means the way in which we identify people in actual everyday life. He then adds:

> Of course it may be that a number of competing theories survive this first stage. The surviving competitors may then be evaluated in terms of their compatibility with our reactions to those puzzle cases which provide situations in which the competing theories diverge in their pronouncements. But these intuitive reactions are defeasible judgments, which we can defeat by showing they are over generalisations from the ordinary run of cases or are produced by some distorting influence or are outweighed by other judgments that we have reason to respect.[75]

I have argued already that no supporter of the method of cases would dispute that judgments wrought from intuitions are defeasible. In which case, the difference between Johnston's position and the one he is attacking seems to be dissolving. For the only difference that now remains is that Johnston considers

[74] Notably, by Noonan [1985] and Parfit, Appendix E, pp477-479.

[75] Johnston [1987], pp63-64

that other avenues should be explored before adopting the method of cases. But even here, I’m not sure that Johnston is accurately representing his opponents. After all, considered historically, it could be argued that the method of cases only has the priority it has now precisely because previous attempts to trace personal identity through more commonplace routes proved inconclusive. Seen in this way, Parfit et al are working in Johnston’s second stage, because the first stage is taken to have been thoroughly worked by their predecessors.

Johnston’s second criticism does not therefore convince. He admits that intuitions have a role in the debate, and so the only real criticism that remains is that they are overused. I would argue that there is little evidence that they have been overused. What Johnston’s argument does highlight is that we must be very careful to make sure that this really is not the case in the specific arguments we formulate. But Johnston has not shown that the use of intuitions should be at all eliminated from our methodology. In the particular case he highlights, he has given us no reasons why it should be abandoned. Certainly, if psychological reductionism depended entirely on this one case, then there would be a problem. But as it is only one argument amongst many, the fact that different intuitions can be elicited from it is not reason enough for us to abandon the use of intuitions altogether. And Johnston has not shown that psychological reductionism does depend on intuitions wrought from this one case. But what is more important for this thesis, Johnston has failed to establish that the method of cases is fundamentally flawed, even if doubts still remain as to the best way to interpret the brain wiping case.

Now we must look at Johnston’s final objection, that the method of cases yields criteria of personal identity that makes ordinary reidentification problematic. This objection seems to me the most puzzling of the three. The argument is that, taken to its logical conclusion, the method of cases yields the result that we are

’bare loci’ of mental life. Given that he reaches this conclusion based upon misreadings of the method of cases which I outlined above, I will not concern myself with that argument here. But once this conclusion is reached, he states:

> The crucial point is that our ordinary claims to know that our friends and familiars were continuously where their bodies were when they were unconscious or in deep sleep rests on nothing like the employment of ...a theory to rule out the possibility that any number of bare loci came to be associated with their bodies during such periods.[76]

In other words, the bare loci view allows for the possibility of a succession of bare loci inhabiting the same body, and yet as normal re-identification makes no allowances for this and yet still works, the bare loci view must be suspect. Firstly, this seems no more than an appeal to the principle of ordinary re-identification anyway. We have not yet seen why this principle is so important. Secondly, as Snowdon points out, sleep poses no particular problem at all:

> His example of sleep can be given the obvious answer that there is very good evidence that people do not quit their normal body links during sleep, namely, that, however random the timing of reawakening they are always there when reawakened. [77]

It can thus be seen that it is not at all clear that the method of cases does yield a result that conflicts with principles of ordinary re-identification, and that Johnston seems to have no argument for holding to this principle at all. Indeed, he introduces the principle of ordinary re-identification with the line, “Here’s a suggestion”[78] . But is there anything to be said for the principle? Certainly, any philosophical view that results in a radical scepticism over something we believe to be unproblematic is sure to raise suspicion, although that in itself is not an

[76] ibid, p74

[77] Snowdon [1991], p124

argument against it. But it seems odd that Johnston requires the theory of personal identity to allow reidentification “in just the way in which we reliably and unproblematically reidentify ourselves and each other over time”.[79] Take a mundane substance like gold, for example. We have our own reliable and unproblematic way of identifying gold, even though the true test of goldness requires chemical analysis most of us don’t even understand. But this scientific principle does not cast doubt on our ordinary ability to do the job. Certainly, we make mistakes with our everyday method, but we also do so when re-identifying people. Similarly, there seems no reason why the principles of personal identity over time cannot be based on factors quite different from those of our ordinary methods without casting doubt upon them. So if, for example we choose a psychological criterion of personal identity, because, as a matter of fact, we neither perform brain transplants nor teletransportation, our ordinary methods of reidentification do not come into doubt. And in actual puzzle cases, such as senility or global amnesia, our ordinary methods of reidentification do not yield any clear results anyway, so there is no unproblematic standard to measure our account against.

In conclusion, then, I would argue that Johnston’s critique of the method of cases fails, for the following reasons: Firstly, the method does have a place in arguments against further-fact views. But even where it does not, it can be used to help decide between reductionist views where we have already eliminated further-fact views by other methods. Secondly, it is not a method that relies solely on intuitions, and therefore has a better foundation than Johnston believes. Consistency and explanatory value are two further requirements that our intuitive judgments must meet if they are to be acceptable. Thirdly, there is no reason to

[78] Johnston [1987], p63

[79] ibid

believe that the method yields results that would oblige us to give up our ordinary, unproblematic methods of reidentification over time.

It is as a result of all of these considerations that Johnston concludes that:

> [..] it is plausible to take our pure concept of a person, understood as whatever it is that our reactions to these puzzle cases manifest, as being too unspecific to be of much interest, and as unspecific in a way that will typically lead us wrong in many puzzle cases.[80]

In which ways does 'person' mislead for Johnston? It leads us (arguably) to accept a bare locus view of the self that makes ordinary reidentification problematic. And it leads to imaginings or thought experiments which are "idle" because they rest upon the employment of untrustworthy intuitions. I have answered both these accusations. But, nonetheless, Johnston's suggestion that the term person is "too unspecific" does raise the important issue of just how wise we are to employ this term.

4. Persons, Animals and Humans.

Locke made a very compelling distinction in his Essay which greatly influenced subsequent discussions:

> It being one thing to be the same substance, another the same man, and a third the same person, if person, man and substance are three names standing for different ideas; for such as is the idea belonging to that name, such must be the identity.[81]

To dissect precisely what Locke meant here would require us to look at what 'ideas' meant for Locke. Fortunately, we do not need to do so, because what I am chiefly concerned with here is how Locke's basic distinction between man and

[80] ibid, p71

[81] Locke [1694], Chapter 27, para.7

person has been used by philosophers rather than with Locke's particular theory. As such, I am not at this stage interested in Locke's particular conception, but with how we can make sense of the supposed person/man distinction.[82] The basis of this distinction is a principle that applies to far more than men and persons. It is the idea that there are many things which fall under more than one sortal concept, and as the identity conditions for each sortal it falls under differs, so an object can cease to exist qua one sortal but continue to exist qua another. For example, something can be both a statue and a lump of bronze. But if this is melted down, although it ceases to exist qua statue, it continues to exist qua lump of bronze.

Applied to ourselves, the idea is that we are 'persons' and 'men'. A person is a "thinking, intelligent being that has reason and reflection and can consider itself as itself"[83] whereas a man is simply a member of the species homo sapiens. What has provided the whole possibility of a debate over personal identity is the idea that the lives of a 'person' and a 'man' can come apart. This can be imagined in cases of a person changing bodies, or a human being losing all faculties deemed necessary for it to be a person, but for that human still to be alive. Furthermore, there is the possibility of non-human persons. If we deny the possibility of person and human being having different identity conditions then there seems to be little to distinguish the philosophical problem of personal identity from that of the identity of any other animal.

There are two types of objection to the idea that our existence qua person may differ in extension to our existence qua human being. One is to deny the general possibility that one thing can fall under different sortals to which different identity conditions apply. The other is to object to the more specific distinction

[82] I shall return to Locke's particular view in §7.1

[83] Locke [1694], Chapter 27, para.9.

between ‘man’ and ‘person’. To do the first type of objection any justice would require a study beyond the scope of this thesis.[84] Therefore in what follows, I shall be considering only the second kind of objection. These will of course imply more general points, but I must leave the development of these to others. What I shall be doing here is looking at some arguments which at first glance could seem extremely important for personal survival, but in fact, still leave questions begging. In what follows, I shall stick to the two examples already given of human/person and statue/bronze.

One counter argument is to claim that rather than there being one object which falls under two sortals in cases such as those of the statue and the lump of bronze, there are in fact various different objects present all along. For example, it has been argued that the only way to make sense of the bronze and statue case is to propose that, contrary to common sense, there is not one thing there which is both a statue and a lump of bronze, but rather two separate things which happen to be co-extensive.[85] The evidence for this is that they do in fact cease to be co-extensive when the statue is melted, which is precisely the same evidence cited as a reason for saying there is one thing which falls under two sortals! Similarly, we are wrong to think that we are one thing which is both a man and a person, but rather there are two distinct but coexistent things, a human being and a person. The problem now becomes: which of these two objects is the true referent of ‘I’? This way of explaining how one object can apparently fall under different sortals can be attacked on many fronts, not least on the grounds of its ontological extravagance. There is, however, no need for us to take issue with this viewpoint, because it reinforces rather than challenges the view that ‘person’

[84] My final position does not require an answer to this general problem.

[85] If we consider the four-dimensional ontology favoured by Lewis, as outlined in §2.2 , we can see quite easily how two different perdurers or ’four dimensional-worms’ can share a common set of stages, just as two roads can have a stretch in common.

and ‘human being’ can come apart. As long as this remains a possibility, we are free to continue our investigation.

A second possible line of objection that has a prima facie relevance is based on the work of Saul Kripke on necessity. Kripke has argued that if X and Y are rigid designators, that is to say, terms that designate “the same object in all possible worlds”[86] then ”for any objects x and y, if x is y, then it is necessary that x is y.”[87] Kripke mainly discusses necessary identity relations, but Kripke uses the same principle where ‘is’ refers to constitution.[88] For example, he would argue that “water is H_2O” is necessarily true as both “water” and “H_2O” are rigid designators.[89] It appears that this issue is very important for personal identity. For if, as many reductionists claim, persons are not “separately existing entities, apart from our brains and bodies”[90], and it can be shown that persons are necessarily their brains and bodies i.e. human beings, then it is hard to see how a person can survive the destruction of their body.

Does Kripke’s view entail that persons are necessarily human beings? It would seem not. For anyone who accepts Locke’s definition of a person, or something similar, ‘person’ is not a rigid designator. This is because anything which matches the definition of person, or any similar definition, would be a person. In other possible worlds, ‘person’ may designate something other than the actual persons in our world. Compare this with water. One could hold up a glass of water and pointing at it say, “water is whatever natural kind *this stuff* turns out to be”. It turns out to be H_20, and so water cannot be anything else. But one does not turn to a person and say, “a person is whatever natural kind this

[86] Kripke [1971], p172

[87] ibid, p163

[88] In the sentence ‘a person is a human being’, is the ‘is’ one of identity or constitution? I take it to be the ‘is’ of constitution, but as the same Kripkean necessities apply for constitution and identity relations, one could read it as an ‘is’ of identity without affecting my arguments.

[89] The actual example Kripke give is of ‘heat’ and ‘The motion of molecules’.

being turns out to be". It is part of the concept 'person' that a being can be a person regardless of the natural kind to which they belong. We cannot conclude that because the persons around here are human beings, a person must be a human being. Wiggins denies this, and I shall return to this denial shortly.

Although Kripke's view does not entail that all persons are human beings, it could be argued that for any particular human person, the fact that that person is that human being is a necessary truth. Consider Kripke's discussion of a wooden lectern. Of course, there is no general necessity for lecterns to be made of wood. But the more interesting question is, "could this very lectern have been made from the very beginning of its existence from ice?"[91] He concludes, rightly I believe, no. But as he acknowledges, "the question of whether it could afterward, say in a minute from now, turn into ice is something else". These considerations help to reduce the relevance of essentialism in our debate on persons. The essentialist holds that, if this lectern is made of wood then had it not been made of wood then it could not be this lectern. Similarly, if I (this person) am constituted by this body and brain, i.e., this animal, then had I not been this animal I would not have been this person. This says nothing about whether I could, in the future, cease to be this animal.

On their own then, there is nothing in Kripke's arguments concerning the necessity of identity and constitution that makes a distinction and possible separation between persons and human beings impossible. I am not saying that Kripke's ideas have no place in the personal identity debate. All I am saying is that there is large gap between the state of the debate now and it having anything conclusive to say about personal survival.

[90] Parfit p216

[91] Kripke [1971], p179

Wiggins has followed the second line I outlined above, which I believe causes more difficulties for the reductionist. This strategy focuses not on the general issue of sortals and objects but on the particular case of human beings and persons. There seems to me to be two distinct lines of argument that characterise Wiggins' work in this area. Firstly, there is the argument that although we are persons, to understand what we essentially are we have to consider ourselves as human beings. Secondly, there is the argument that 'human being' and 'person' are co-extensive terms. I shall start with the first of these arguments.

As I discussed in §2.2, Wiggins makes use of Locke's distinction between natural and nominal essences. While real essences are out there in the world whether we discover them or not, nominal essences are constructs which we fit the world into. Hence a cardboard box has a real essence which is its molecular structure and so on, while it's nominal essence depends on what it is used for, storing objects or as a table. for example. A nominal essence is definitional, that is to say, a table is whatever is used as a surface for writing, eating off or placing drinks and so on. 'Person' seems to be a nominal term, because whatever fits the definition of a person, be that definition Lockean or a variant, is a person. Locke, Dennett et al, in their attempts to formulate conditions of personhood are trying to get at the best definition of the term, not conducting an empirical investigation as to what persons are in the way that metallurgists discovered what gold really is. 'Human being' is quite different. We can discover what a human being is only by scientific investigation. It is our assumption that there just is a natural kind, homo sapiens, of which it is our job to discover the essence.

When we consider what we are, we discover that we are persons and we are human beings. Which of these terms captures best what we essentially are? Wiggins' view is that if we really want to know what we are then it makes more sense to consider our natural essence rather than our nominal essence. As natural essences are concerned with the way the world is carved up, rather than with the way we carve up the world, our natural essence should give a more accurate picture of what we really are. So to discover what we really are, we should consider ourselves human beings. Our conditions of identity are then determined in the same way as any other animal's conditions of identity, in a way which fits the natural continuation of a living organism.

This is certainly a strong objection against any form of psychological reductionism that defines personal identity in non-physical terms. But it runs up against the same problems I discussed in chapter two. Whether or not 'person' is a nominal or natural term, the first person question of survival remains open. It may be that our identity conditions should be determined in the same way as any animals, but if this only answers the third person question of identity and not the first person question of survival, then the relevance requirement has not been met. We would only have answered one of the puzzles of personal identity and this is not, I have argued, the question most pertinent to our interests in persons, as persons ourselves. Wiggins' argument is, however, still important, because if we were to conclude that the answer to the first-person question of survival involves identity, then it would rule out psychological reductionism.

Wiggins' point may actually strengthen the case that identity is not part of the answer to the first person question of survival. Consider myself and a bodily continuous future person who has fallen into madness, recalls nothing of my life up until this descent into madness and has a totally different character to me. How should I think of this person? Should I make provision for him? These are important questions. Now, I am told that this person will be me, because identity is determined by the identity of the human animal. This doesn't seem to answer my worries. It may well influence my decisions in some way, but it leaves other questions like, "Should I consider my relation to that future person as being as important to me as the normal relations I have with myself at other times?" And these questions seem to be the important ones.

Wiggins' second line of argument has been to argue that the concepts of a particular person and a particular animal cannot be disentangled.[92] If we are careful enough in our analysis of the term 'person', then prima facie cases where there are persons without human beings, or human beings without persons will be discredited. I have already suggested why, if we accept a Lockean definition of a person, (and I see no reason why we should not), then the sortal "person" is not constricted to the natural kind actual persons happen to be. Snowdon also rejects Wiggins' view, claiming that it ultimately depends upon the claim that animals are "of necessity the unique possessors of mental states," and that "neither Wiggins nor anyone else has given persuasive grounds for this claim."[93] But he concludes his discussion of Wiggins with an interesting suggestion:

> 'Person' in its psychological use should be thought of as a term which can apply to entities of different types, which does not pick out what we fundamentally are, and does not pick out a sort sharing criteria of identity.

This is only a suggestion and Snowdon as yet has no argument to back it up, but it does seem to me to express quite succinctly the essence of the doubt people have about this term 'person'. What follows is my attempt to unpack this suggestion and may or may not represent Snowdon's own thoughts.

Let us return to Locke's original definition of a person. What we have here can be read as a list of attributes. A person is a thing which can reason and think. But as to what sort of thing it is which has these attributes, the definition is neutral. Locke's definition allows for carbon, silicone or spiritual persons. In that sense, Snowdon is right to say that 'person' "can apply to entities of different types". One

[92] Wiggins [1987]

[93] Snowdon [1994]. MS

can think of many other terms which apply across types. Scavengers can be many different sorts of animals, ‘fast things’ can be mechanical, organic or chemical. But in these cases, the thing which qualifies as a person, scavenger or fast thing is not usually individuated by that attribute. The formula one car that breaks down or the scavenger that turns vegetarian ceases to be a fast thing or a scavenger, but does not cease to exist. Similarly, if a human being ceases to think and reason, it may cease to be a person but does not cease to exist. Furthermore, transference of an attribute need not take identity with it. Let us say that scientists remove the scavenger gene (which almost certainly does not exist, but let us imagine it does) and transfers it into a sheep and that engineers remove the formula one car’s engine and put it in a Lada. We do not say that the Lada now is the formula one car or that the sheep is the very same animal as the now ex-scavenger. Similarly, so an argument may run, if we transfer the attributes of reason, memory and so on from one organism to another, surely this does not take identity with it. This kind of reasoning must be what motivates the sort of objection Snowdon offers. Person is a term that covers anything with a certain set of attributes, but it is unable to pinpoint what anything which falls under this term fundamentally is. So it is in vain that we look for identity conditions in the continuation of these attributes. Instead, identity conditions can only be obtained by looking at the natural kind which supports these attributes, in our case, human beings.

Wiggins says something very similar in *Sameness and Substance*. He entertains the possibility that person is a “concept whose defining marks are to be given in terms of a natural kind determinable, say animal, plus what may be called a functional or systematic component”.[94] This would make ‘person’ rather like ‘vegetable’. Vegetables are any number of natural kinds whose root, fruit or leaf is savoury and edible. So although they are defined functionally, their identity conditions are given by their natural kind. He dubs this view 'animal attribute view'. He goes on to say that 'person' differs from 'vegetable' because "the definition of 'person' is not something we conceive for ourselves in the way in which we have conceived for ourselves the nominal essences for 'hoe' or 'house'."[95] The attributes of persons are therefore discovered by finding out the attributes of human beings, for Wiggins, the only real persons there are.

These considerations of natural kind and attribute put the issues before us sharply in focus. I agree that for token identity, we would be foolish to trace a continuation of attributes. But I have already rejected the idea of identity as being unproblematically a crucial part of the answer to the first-person question of survival. This allows for the possibility that in cases where there is continuation of attributes there is also what is necessary for personal survival. I have argued that what matters may not be identity but continuation of some kind. If then all the attributes that make me, now, a person, continue, and what is important is that I am a person and not this animal, then can I not view this as survival? It is certainly not survival of me as I am now, but it may well be survival of all that matters about me now. I would be in a similar position to the British person facing Euro-integration that I described in §2.1. The thing which I care about will not continue to exist but what I care about in that thing will continue to exist. Nostalgia may induce a touch of sadness but I have no real reason to feel that the future is any worse from my point of view than it is now.

I should stress once more that this possible view of personal survival is not one which I have presented strong arguments for. What it rather does is open the door for a way of understanding personal survival in a different way. The crucial

[94] Wiggins [1988], p171
[95] ibid, p173

point is simply that unless token identity is shown to be crucial to answering the first person question of survival, then there is no reason to suppose that we cannot talk about survival of the person as distinct from the survival of a numerically identical human being.

If this is correct, then the animalist could be right without threatening my position. All the following can be true:

(1) Persons are humans.

(2) Personal identity is identity of the human animal.

(3) We can find out more about persons by studying actual persons i.e. Human beings.

(4) Personal survival[96] does not imply personal identity but the continuation of those aspects of the person which are most valuable, namely those aspects to do with consciousness, character and memory.

This justifies looking at the issue of *persons*. If instead we looked solely at human beings, then we would be rejecting the possibility of (4) being true out of hand.

In this chapter I have considered the twin bases of our enquiry – its key concepts and methodology. The criticisms that have been directed at these two bases all have something important to say and our review of them has revealed important things about our enterprise. We have found that person is unlikely to be a suitable concept to gather identity conditions under, but that nonetheless, persons, not human beings, can still be our primary concern. We have got clearer on why and how thought experiments work and we have also seen some of the traps the ‘Method of Cases’ must avoid. Now, we can begin.

[96] As defined in chapter tw

Part Two

Introduction

Having set out certain requirements for a philosophy of persons and defended the use of certain methodologies and concepts, I will now turn to consider one particular position on the issue, namely Parfit's version of psychological reductionism.[97] I shall not be arguing for this view from first principles. The pedigree of psychological reductionism in general has been established historically, and developed ingeniously by Parfit, and so there is nothing I could add in support of the basic principles of Parfit's conception. However, I do believe that there are significant flaws in Parfit's position and my aim is to remove these flaws whilst retaining the essential elements of the Parfitian conception.

I argue that there are features of Parfit's psychological reductionism which appear essential to it, even though they can and should be removed. In part two, I distinguish the essential elements of Parfitian psychological reductionism from certain unfortunate features which have come to characterise the account and then come up with a revised view of what psychological reductionism is which I hope leaves out those features of it which are unacceptable.

Before beginning, it would be worthwhile giving an overview of what is to come, even though there may be a certain opacity in this summary that I hope will be removed over the remainder of this thesis. Parfit's position will Parfit's position has three characteristic features, all of which I believe can be retained after my revisions. The first of these features is the explanation of what is required for a unified mental life in terms of psychological connectedness and continuity. I will argue that Parfit's account of this, Relation R, is flawed, but that we can replace it with a better explanation which fulfils the same basic function as Relation R. A unified mental life can be explained in terms of psychological connectedness and continuity, but not in the way Parfit attempts to do so. The second feature is that there can be a unified mental life without personal identity. The third feature is that it is the unified mental life, not personal identity, which matters in survival. The question of what is meant by "what matters" will be dealt with in chapter six.

It is clear that the main one of these claims is the first, namely that what is required for a unified mental life can be explained in terms of psychological connectedness and continuity. Such an explanation has two parts: there are the relata of such an account, namely thoughts and experiences, and the connections between these relata. Chapters four and five deal with important problems in the Parfitian conception of persons concerning these two elements. In chapter four, I consider whether psychological reductionism requires its relata – thoughts and other mental events – to be independent of the subjects who have them. I argue that there is more to be said on this account than Parfit considers, but that a proper appreciation of the issue need not threaten the Parfitian conception. Equipped with this better understanding, I turn to the relations between thoughts and experiences in chapter five, where I argue that Relation R fails to perform the function required of it, but that there is room for an alternative which still retains the distinctive Parfitian features. In chapter six I take the lessons learned form the largely negative arguments of chapters four and five and offer a positive account of how psychological reductionism should be described. It will then be the job of part three to assess this revised conception and test it against the requirements of part one.

Chapter Four

[97] All referencees hitherto to Parfit will be to *Reasons and Persons* unless otherwise indicated

Parfit and "I"

At the risk of over generalising, it could be said that historically, psychological reductionism's strength has been in its general principles, but it has often proved weaker when it comes down to fleshing out its accounts with detail. For example, Locke famously located personal identity in the continuity of the same consciousness and offered several convincing arguments for this.[98] But on closer examination, his criteria for personal identity were shown to be inadequate. Hume's claim that we cannot locate the self through introspection was persuasive, but that left him totally unable to account for the unity of a person's life.[99] Parfit makes the opposite type of error. He does try to fill in the detail of his position, but in doing so, he loads his thesis with unwelcome and unnecessary baggage. In this chapter, I try to unload some of that baggage. Parfit claims that reductionism can account for the unity of consciousness over time solely in terms of the mental events and the relations between them which constitute that life. What is wrong with this and how his errors can be corrected is the subject of this chapter.

The key question that is being asked in this chapter is, "Does the nature of thoughts in general, and first-person thoughts in particular, entail any facts about the subjects of those thoughts?" In other words, can we learn anything about persons by considering the nature of the thoughts which constitute a person's mental life? The relationship between thoughts and the subjects of those thoughts is a matter of great disagreement. Parfit claims that we can describe thoughts and mental events independently of the persons who have them.[100] What

[98] Locke [1694], Book 2, Chapter 27.
[99] Hume [1739], Appendix.
[100] Parfit, §81

precisely this means, whether or not it us true and the importance of this issue for Parfit's conception of persons are three issues which I consider in this chapter.

I begin by explaining what a theory's being reductionist requires and how Parfit's claim that thoughts can be described independently of thinkers fits into this reductionism. I then consider two arguments that apparently conflict with the claim that thoughts and thinkers can be separated, that recently put forward by Quassim Cassam and its archetype, that of Immanuel Kant. I then examine the consequences of these claims for the Parfitian approach to personal survival. I argue that a Parfitian position can be made compatible with both Kant and Cassam's accounts, though in reconciling these differing positions, revisions will have to made to Parfit's account.

1. What is Reductionism?

Reductionism means different things in different disciplines and even different areas of philosophy. I will not attempt to give a general definition of reductionism but will stick to reductionism about persons. Parfit says that all reductionists claim:

> A person's existence just consists in the existence of a body, and the occurrence of a series of thoughts, experiences, and other mental and physical events.[101]

Prima facie, this is a pretty anodyne definition, and would seem to make reductionists of everyone bar those who hold that a person is essentially a soul or indivisible Cartesian ego. But this definition does contain a few assumptions which when made explicit make it more interesting. Firstly, what characterises the reductionist position is not just that reductionism makes it possible to make a list of what a person consists in, as above, but that those items on the list are taken to be the constituent parts out of which a person is constructed. Parfit's definition suggests that we have a set of ingredients which, when blended together, produce a person. The way Parfit explains his reductionism implies the possibility that these ingredients could all have an independent existence, without there being any persons. Not everyone who at first glance would accept what Parfit's definition says about persons would accept this. Many would claim that although a person is just a body and a series of mental and physical events, there is no way any of these events could exist without the existence of a person. I shall consider precisely what these claims concerning the independence of thoughts amount to shortly.

Secondly, the reductionist account should go at least some of the way towards explaining the unity of a person's life. It is not just that all the elements listed together constitute a person, but that the way they come together explains how and why persons are what they are and how they are. If the reductionist position is true, then it will be possible to explain what makes thoughts and events part of the life of a particular person in terms of the constitutive elements listed in the definition. Again, I think Parfit fails to explain how this is so, and I attempt to show why in the next chapter.

There is more than one reductionist position however, although all would agree with the general definition. Parfit describes three positions: Identifying Reductionists (IR), Constitutive Reductionists (CR) and Eliminative Reductionists (ER). Their differing claims are:

IR: Persons just are bodies.

CR: A person is an entity that has a body, and has thoughts and other experiences. Though persons are distinct from their bodies, and

[101] Parfit, p216

from any series of mental events, they are not independent or separately existing entities.

ER: There really aren't such things as persons; there are only brains and bodies, and thoughts and other experiences.[102]

Parfit claims that he is a constitutive reductionist. Although he claims that we could describe the experiences and relations between them which go up to make a person "without claiming that these experiences are had by a person"[103], this does not make him an eliminative reductionist. This no more follows than it does to say that, because we can describe all the parts of a car and the relations between them without claiming they are parts of a car, then cars do not really exist. The claim that we can describe the constitutive elements of persons without mention of persons can be expressed as the claim that there can be a personless description of what constitutes persons. The possibility of a personless description is a distinctive feature of Parfitian reductionism. However, it is not clear what precisely is meant by a personless description and how important it is to Parfit's argument. I conclude in this chapter that in order to strengthen the Parfitian view, we will have to rule out the possibility of a personless description.

A constitutive reductionist could be a physical or psychological reductionist. Both would agree that a person simply is their body, thoughts, events and so on. A physical reductionist would claim that it is virtue of continuity of body and or brain that these elements are part of the life of a single person, whereas a psychological reductionist would claim that it is in virtue of the mental features that these elements are part of the life of a single person. For the psychological reductionist, a person is defined and individuated by the series of psychological events which go up to make the life of that person and not by any facts

[102] Parfit's notes on the 1993 Jacobson Lecture, University of London.
[103] Parfit, p217

concerning their bodies or brains. In this way, although actual persons are living organisms, it is in virtue of their psychological attributes that they are persons, and the conditions of identity over time for persons are psychological ones.

Let us consider in more detail what this position entails. It appears to me that to be a psychological reductionist entails undertaking two different reductions. The first is the “whole to parts” reduction, which is described by Parfit’s general definition of reductionism. A person is simply those physical and mental elements that together constitute the life of a person. Here, the complex whole – the person – is reduced to the sum of its simpler parts. But in order to be a psychological reductionist, a further, different form of reduction is required. Of all the elements which make up a person, one type of element – the mental – is bracketed off. It is in virtue only of these elements that a person is the particular person that he is. For example, it is held that if all the physical matter from which a person is made up is changed, as long as there is a continuity of mental life, the person will remain the same. This kind of reduction I call a “total to essentials” reduction. In this type of reduction, we take any object or phenomenon which is constituted by several aspects and reduce it to those aspects of the object which are crucial in individuating it. For example, the British Labour Party is constituted by its members, organisation, policies and so on. But some would argue that what makes the Labour Party what it is is, for example, clause four of its constitution. Everything else about the party could change, but as long as clause four persisted, it would still be the Labour Party. As is clear in this example, the total to essentials reduction is an attempt to capture two things. Firstly, it captures the defining features of a thing. A person, for the psychological reductionist, is a person precisely because it has a mental life. It may require a body to have this mental life, but it is not in virtue of the body that the organism is a person. Therefore, despite the fact that a human person is a complex organism, its psychological features are what define it as a person. As a consequence of this, secondly, it attempts to capture what is required for something to remain that thing over time, i.e. it pinpoints those elements which must form part of its conditions of identity, or survival, over time.

Although we have seen that there is a two-part reduction here, in Parfit's view the reduction has not yet gone far enough. The problem is this: if persons are to be individuated by their mental lives, then how in turn are these to be individuated? It is not enough to explain the unity of a person's life in terms of the unity of a mental life, we must also explain the unity of the mental life. It is in answering this question that Parfit's reductionism gets more specific and more reductive still. Parfit, having already reduced the unity of a person's life to the unity of mental life, now attempts to reduce the unity of mental life to the thoughts and experiences that make up this mental life. The claim is that mental events belong to the life of a single person because they are interrelated in certain ways and not because they are intrinsically owned by one person or another. As Cassam describes this position:

> It is because the thoughts and experiences themselves and the relations between them can be described in impersonal terms that facts about the ownership of mental states or events by a subject can be said to have received a reductive explanation.[104]

Parfitian reductionism, according to Cassam, is thus the attempt to reduce the personal to the impersonal. In other words, thoughts which are thoughts of a person can be described without reference to that person. Furthermore, it is this feature which Cassam claims makes the account reductive. On this point, I disagree. As I have explained, the reduction of the personal to the impersonal is a further reduction that comes after the initial two-stage reduction. So an account that didn't have this last reduction would still be a reductive account. There would still be, of course, a need to explain the unity of a mental life, but this explanation would not need to be reductive for it to be part of a wider reductive account.[105] Nonetheless, Parfit does offer a reduction of the personal to the impersonal, although there is an ambiguity in what this could mean, which Cassam again notes:

> Shoemaker has pressed the question whether Parfit's impersonal description is simply one in which there is no actual reference to persons, or whether the entities referred to in impersonal descriptions are thought of as capable of existing without there being persons.[106]

Cassam thinks Parfit holds the latter, stronger view, but Cassam's interpretation of Parfit is not unequivocal. The question for Parfit is then, do the thoughts and events which make up a person's life require the existence of a person for their own existence or could they exist if there were no persons at all? The weak version of Parfitian reductionism would claim that there is no need for thoughts to be capable of independent existence. All that is required is that we do not need to refer to persons in our account of what persons are. The strong version of Parfitian reductionism says we can imagine being able to describe the mental events which could go to make up a person's life without there actually being a person at all. What explains the existence of thinking, conscious subjects is the existence of thoughts, desires, memories and so on which just so happen to be interrelated in the relevant ways. But persons only exist because of these interactions and there is no contradiction in holding that any one of the mental events which go towards constituting the life of a person could exist without there being any persons at all. Cassam calls this Strong Reductionism. This strong

[104] Cassam [1992], p363

[105] I offer such an account of mental unity over time in chapter six.

view is *prima facie* far more implausible than the weak version. And Cassam argues that strong reductionism is a position that is not ultimately tenable because “it is distinctive of a life of a person or subject to be self-conscious, and hence to include the thinking of first-person or I-thoughts. [...] and the content of a particular I-thought cannot adequately be characterised without ascribing it to a particular person or subject.”[107]

Deciding which view Parfit holds is no easy task, as Parfit is far from explicit on this point. However, I am not so much concerned with the interpretation of Parfit as with the consequences of denying the independence of thoughts. There are at least two ways in which this independence can be denied. The first is Cassam’s claim that I-thoughts depend *ontologically* on persons or subjects. The second is Kant’s claim that thoughts *formally* entail a subject. In the remainder of this chapter I shall consider and explain both these views and examine their consequences for the Parfitian conception. I shall argue that neither claim is fatal to the Parfitian conception, but that both require changes to it. I shall begin with Kant’s arguments.

2. Kant: No Thoughts Without a Subject.

Kant’s views on the relations between a thought and its subject are famously difficult and I obviously cannot undertake a full account of them here. What I will do is simply to try to state key aspects of his position and sketch some of the reasons he has for holding them. The account I am focussing on comes in his *Critique of Pure Reason*.

Kant argues that for it to be possible for there to be experience of the world, there has to be a fundamental unity within the subject of experience which allows

[106] Cassam [1992], p361

[107] Cassam [1992], p362

the world to be perceived. Kant calls this the transcendental unity of apperception: the unity of inner experience necessary for any particular experience of the world to be apperceived as part of a wider, unified whole. Closely related to this idea is Kant's claim that there is an "I think" which can necessarily accompany all thought. It is this "I think" which allows the manifold of experiences to be grasped as a coherent whole in conscious experience. Without it, consciousness could only be of a psychedelic series of dissociated experiences.

The fact that any thought can be accompanied by an "I think" entails the formal requirement that every thought must have a subject. But Kant argued that we cannot make any inferences from the fact that thoughts must have a subject to the nature of that subject. The transcendental unity of apperception is not an object of experience, but that which makes experience possible. Therefore knowledge of either what the subject of consciousness is or of the phenomenon of consciousness in general cannot be obtained from the unity of apperception. This can only come through "intuition", i.e. experience of objects in the world. So there is a distinction between the a priori knowledge of the unity of apperception and the empirical knowledge of subjects of consciousness. Problems arise when the two are confused.

So Kant actually agrees with Descartes' starting point, that the cogito, the "I think", is necessary for all thought. It is simply that one cannot move from that to any conclusions as to the nature of this cogito. Kant writes of the cogito:

> As is easily seen, this is the vehicle of all concepts, and therefore also of transcendental concepts, and so is always included in the conceiving of these later, and is itself transcendental. But it can have no

> special designation, because it only serves to introduce all our thought, as belonging to consciousness.[108]

The fact that I think makes all judgments possible. This form of knowledge, knowledge of myself as a subject of thought, is what Kant calls transcendental, meaning it is not concerned with objects of intuitions, but the form of understanding. All our judgments have the possibility of self-ascription; they can all be described in the form “I think X”. So “I think” is simply the wrong sort of thing to be an object of the understanding. It is rather the form in which understanding is held.

But Kant continues:

> However,... it [the cogito?] yet enables us to distinguish, through the nature of our faculty of representation, two kinds of objects. ‘I’, as thinking, am an object of inner sense, and am called ‘soul’. That which is an object of outer sense is called ’body’.[109]

So although nothing can be inferred from the ‘I think’, there is still the soul, the ‘I’ of inner sense to deal with. This is where Kant goes on to explain the ‘rationalist error’ of Descartes, in what Kant calls ‘transcendental paralogisms’. A logical paralogism is simply a fallacious syllogism. But “a transcendental paralogism is one in which there is a transcendental ground, constraining us to draw a formally invalid conclusion”.[110] In other words, because of the way we necessarily see the world, we are seduced into making an invalid deduction. This is demonstrated in the first paralogism:

[108] Kant [1781], A342

[109] ibid

[110] ibid,

> That, the representation of which is the absolute subject of our judgments and cannot therefore be employed as determination of another thing, is substance.
>
> I, as a thinking being, am the absolute subject of all my possible judgments, and this representation of myself cannot be employed as predicate of any other thing.[111]
>
> Therefore I, as thinking being (soul), am substance.

In the second edition, the same argument is put differently:

> That which cannot be thought otherwise than as subject does not exist otherwise than as subject, and is therefore substance.
>
> A thinking being, considered merely as such, cannot be thought otherwise than as subject.
>
> Therefore it exists only as subject, that is, as substance. [112]

The ‘I’ of the first edition becomes the ‘thinking being’ of the second. But in both cases, what seems to be the source of the confusion is a mix-up between substantial and formal subjects of thought. The first premise seems to concern thinking subjects as objects of thought. Both the ‘thats’ of “That, the representation of which... ” and “that which cannot be thought... ” seem clearly to be objects which are being thought of and being represented. It follows from Kant’s empirical realism that objects that are represented exist and as such are substance. The subject here is thus substantial: its existence is empirically given. The second premise uses a different employment of ‘I’ or ‘thinking being’. In this

[111] ibid, A348

[112] ibid, B411

case, there is no object of intuition, rather, we are talking about the way in which a subject of thought must be considered. In Strawson's words, in this case "I" is used as "criterionless self-ascription".[113] That 'I' am subject is analytic; it tells us nothing about that to which 'I' refers. The subject here is thus formal, i.e. its existence is given, but not empirically, thus there is nothing that can be said concerning its substantiality. Now we can see why the conjunction of the two premises is a necessary paralogism. The second premise is that we cannot be thought of other than as subjects. Though this cannot actually tell us of anything beyond the formal requirement for a subject, because in the first premise 'subject' is substantial, when we conjoin the two premises, which use 'subject' in different ways, we get a synthetic conclusion apparently proving the substantiality of the soul. We must remind ourselves that 'subject' is ambiguous, in that it can be formal or substantial, and that no conclusion from premises which conflate these two meanings can be sound.

This is very important for the argument in the third paralogism concerning personal identity:

> That which is conscious of the numerical identity of itself at different times is in so far a person.
>
> Now the soul is conscious, etc.
>
> Therefore it is a person.[114]

In this case we again need to consider the distinction between "the whole time in which I am conscious of myself" and the "outer observer who represents me in time".[115] In the former case we again have an a priori statement which reveals nothing of the real existence of myself.

[113] Strawson, [1966], p164-5

[114] Kant [1781], A361

[115] ibid, A362.

> The identity of the consciousness of myself at different times is therefore only a formal condition of my thoughts and their coherence, and in no way proves the numerical identity of my subject.[116]

There is neither an inference nor an intuition that goes along with this self-identity. To be conscious of oneself and conscious of oneself as the same self is one and the same thing. The latter is entailed in the former. As Strawson puts it:

> When a man ascribes a current or directly remembered state of consciousness to himself, no use whatever of any criteria of personal identity is required to justify his use of the pronoun 'I' to refer to the subject of that experience....It would make no sense to think or say: I distinctly remember that inner experience occurring, but did it occur to me? There is nothing that one can encounter or recall in the field of inner experience such that there can be any question of one's applying criteria of subject identity to determine whether the encountered or recalled experience belongs to oneself - or to someone else.[117]

Consider how this differs from ascribing identity to an external object. Kant's incomplete description is:

> If I want to know through experience, the numerical identity of an external object, I shall pay heed to that permanent element in the appearance to which as subject everything else is related as determination, and note its identity throughout the time in which the determinations change.[118]

In this case, there is something substantial in identity statements. As it does concern objects of intuition, there is a substantive referent for the identity statement. "Same person" is meaningful because "different person" is

[116] ibid, A363.

[117] Strawson [1966], p165

conceivable. But “different I” makes no sense, and so “same I” doesn’t inform us of anything.

But there is something even worse about using ‘I’ as an identity fixer. In this case, Kant is unusually clear:

> Despite the logical identity of the ‘I’, such a change may have occurred in it as does not allow of the retention of its identity, and yet we may ascribe to it the same-sounding ‘I’, which in every different state, even in one involving change of the [thinking] subject, might still retain the thought of the preceding subject and so hand it over to the subsequent subject. [119]

Here, “subject” is used in the substantial sense. So what this quote means is that the use of ‘I’ is compatible with there in fact being a series of materially different subjects. Consciousness could pass baton-like from one substantial subject to the next, and the assertion of identity would be empty. An important phrase here is the “same sounding ‘I’”. The ‘I’ in thought is same sounding because when I think of what I did or thought yesterday, it seems to me that the ‘I’ which did those things is the same ‘I’ that is now thinking of them. From my point of view, it appears that there is an identity between ‘I’ today and ‘I’ yesterday. Kant’s point is that I am not entitled to conclude from this that it is the same substantial ‘I’ yesterday as today, as its appearance as the same ‘I’ is compatible with there being various different substantial subjects. In this way, the ‘I’ only appears, or “sounds”, the same – it may not *be* the same.

We can sum up the core of Kant’s arguments about the “I think” by considering a short quote from the Critique:

[118] Kant [1781], A362

[119] *ibid*, A363

> I do not know an object merely in that I think, but only in so far as I determine a given intuition with respect to the unity of consciousness in which all thought consists. [120]

We have knowledge about the way thinking things are solely in virtue of the fact that we are able to perceive them as objects of our consideration. But the way we do this cannot itself be the object of knowledge. We cannot, as it were, turn the camera of mind back onto itself.

In conclusion, there are two key points we need to take from Kant's work. Firstly, all thoughts have the possibility of having the "I think" attached. In other words, even when there is no occurrent belief "I think", the thought is in such a form that the "I think" could be appended to it. But secondly, we cannot use this fact to reach any conclusions about what subjects of thought really are. The "same sounding I" can be found across thoughts which are not part of one materially single subject. I have not offered any arguments for these conclusions and have only broadly outlined Kant's. The fact that these are Kant's conclusions does give them some weight. The least I have done is shown how a position which holds both of the above conclusions is not only possible, but has a notable historical precedent.

I shall consider the implications for Parfit of accepting Kant's argument shortly, but first I need to consider another view of how thoughts require subjects.

3. Cassam: No I-thoughts Without a Subject.

Cassam offers a short, simple argument in favour of his view that I-thoughts always require subjects and then takes much more time discrediting various arguments against this view. He starts with the general claim that the content of a thought cannot be fully specified unless the truth conditions of that thought have

[120] *ibid*, B407

been determined. Hence, to specify the content of the thought “that tree is burning,” we need to know which tree is burning. Similarly, the content of the thought “I am in pain” cannot be fully specified unless we know who the thinker of the thought is. This entails that the content of I-thoughts cannot be specified unless there is a thinker of that thought. As Cassam puts it:

> In order to know what the thought is one must know who the thinker is.[121]

A corollary of this is that to know there is a thought you must know there is a thinker. We must be clear precisely what this argument requires for first-personal thoughts. On certain definitions, it is not a person. Cassam gives as an example Dennett’s conditions of personhood. On Dennett’s formulation, persons are rational, subjects of intentional ascriptions, language users, self-conscious, require a certain attitude to be taken towards them and are capable of reciprocity.[122] These last two conditions are required for persons to be moral agents. Cassam notes that even if the last two conditions are lacking, a being would still be capable of first-person thoughts. Cassam then makes it clear that his thesis is that “I-thoughts are only properly ascribable to subjects.”[123] Cassam simply assumes subjects are “like persons except that they do not necessarily meet the ethical conditions of personhood.”[124] Whether or not Cassam is right, for the moment I shall accept the possibility of his conception of what it could mean to be a subject, but not a person.

This point is important, as it opens up a gap between what can be called a subject of thought and a fully-fledged person. This invites investigating whether although the thought with a subject cannot be reduced to the subjectless thought,

[121] Cassam [1992], p365

[122] Dennett [1992], p177-78

[123] Cassam [1992], p370

[124] ibid

the personal can still be reduced to the impersonal, by which I do not mean third-personal but more specifically, whether reference to persons is eliminable.

Whereas Kant argued that a subject was a formal requirement for thoughts to be had at all, Cassam's conclusion is rather that a substantial subject is required. There needs to be a subject in order for an I-thought to be a genuine "self-ascriptive, subject–predicate thought."[125] The "I" needs to refer to something, namely, a person or subject. Cassam has notable predecessors, such as Evans, who were also prepared to argue that 'I' entails a substantive subject.[126] So although Kant and Cassam agree that thoughts entail subjects, they differ considerably in their reasons for why this is so and in what sense a subject is required. Given these differences, we should consider the consequences of Kant's and Cassam's arguments for Parfit separately.

4. Parfit and Cassam.

If Cassam is right, how can Parfit retain his claim that it is possible to specify the thinker of an 'I' thought without reference to persons? The key is that Cassam's requirement to specify the thinker of a thought can be met solely by subjects-at-a-time, without the need to specify persons. This can be shown by reconsidering some of Parfit's thought experiments. Imagine someone is in the duplicating teletransporter, which has developed further faults. It not only does not destroy the original person, it also sends two copies on to outer space. Just before losing consciousness in the booth, X thinks "I am cold". On waking up, she and her two copies, Y and Z all think, "so I'll put on a jumper". Cassam's key claim was that "in order to know what the thought is one must know who the

[125] ibid, p365
[126] Evans [1982], chapter 7.

thinker is".[127] In this case, it is true that we know the thinker of "I am cold" is X, and X, Y and Z are all thinkers of the type-identical thoughts, "So I'll put on a jumper". Here, we have specified the substantial subjects-at-a-time of the thoughts in question. The problem here is that we still don't necessarily know which *persons* are the subjects of these thoughts, as it is precisely the point of such thought experiments to show that there is a problem of personal identity in such cases. This would be especially true if X had been destroyed and only the replicas remained. There is no such problem of the identity of the substantial subjects-at-a-time of each thought. The fact that there is a mystery concerning which persons are present doesn't prevent us from being able to locate and pick out the substantial subjects-at-a-time involved.

The force of Cassam's point that "in order to know what the thought is one must know who the thinker is" can in this way be somewhat weakened. Certainly, in order to know what the thought is we need to know that it has a subject. But the subject of thought can be located, and thus the content of the thought known, without knowledge of which person is having the thought. For example, in the teletransportation case, we need to decide which, if any, of the pre- and post teletransportation subjects-at-a-time are stages in the life of a single person in order to attribute the thought to a particular person. But before we do this we are able to know the content of the thought, as we are able to pinpoint the substantial subject at the time of the thinking of the thought. We can thus fully specify the content of each occurrence of the thought "So I'll put on a jumper" without reference to persons. As long as we know the subject of each thought, X,Y or Z, we know enough.

In this way, the sense of "subject" which Cassam's argument requires is something less than that of a fully-fledged person. A subject is here simply the

[127] Cassam [1992], p365.

substantial being which has the thought at a particular time. But the concept of a person involves the way in which a being's existence extends over time. This opens up room for the impersonal description of persons Parfit requires. A person can be described as a certain series of substantial subjects-at-a-time. So I-thoughts do not entail anything about the substantive persons we are.

The case of fission makes the point even more clearly. Strawson discusses this case and labels the person who splits P and the two resulting persons P^1 and P^2. From the inside, both P^1 and P^2 will be forced to see themselves as the same person as P. But we know that they cannot both be the same person as P.[128] If we are to attempt, as the rationalist does, to make substantive judgments about our identity over time merely because of the form of our thoughts, we will conclude that because we think we are the same as the original person, we are the same as that person. Strawson neatly highlights this error by imagining P^1 and P^2 both saying in unison:

> The rational psychologist is right! On the inside one does detect the asymmetry, for I know that I am P and my opposite number is not. [129]

In this case too, even if we know the substantive subjects-at-a-time of the various thoughts are P, P^1 and P^2, we don't yet know whether these three subjects are one, two or three separate persons. In this way, the concept of a subject-at-a-time is more basic than that of a person, and the knowledge of oneself as subject does not provide any further knowledge of oneself as a person. This is shown by the fact that P, P^1 and P^2 all know which thoughts they are the subjects of, but they can still be mistaken over which person they are. So to say "P is subject of thought T" is an impersonal description in the sense that P need only be described as a subject of thought and not as a person. This makes it

[128] See §2.3 for the reasons why both fission products cannot be identical with the pre-fission person.

impersonal in the strict sense of "without reference to persons", rather than third-personal.

This response requires giving up the idea of thoughts as ontologically independent, but it still allows for an impersonal description of persons. The key features of Parfitianism can survive this revision. It would still be possible to account for a unified mental life over time in terms of psychological connectedness and continuity, but it would now be in terms of connections between subjects-at-a-time rather than individual thoughts and events. Furthermore, this psychological relation between subjects-at-a-time is different from the identity relation, as is shown by the fact that it would hold in the fission example above. Nothing of importance is lost by talking about psychological connectedness and continuity between subjects-at-a-time rather than individual thoughts, but rather more credibility is gained.

Although I believe this is a tenable response to Cassam, why not go further with Cassam? To respond in terms of subjects-at-a-time entails a view of persons as being some kind of aggregate or product of strings of subjects-at-a-time. But as these subjects do not lack Dennett's ethical conditions of personhood, what is there to distinguish these subjects-at-a-time from persons-at-a-time? Why insist on talking about subjects-at-a-time, when the subjects in questions are clearly persons? Certainly, we would lose the "personless description" but this seems neither an important nor well-defined feature of the Parfitian conception.

I need not decide between these two options. What I believe to be more important is to assess the relevance of Cassam's point to the Parfitian enterprise. As I argued in part one, I believe the first-person question of survival to be different from the third-person question of identity. Further, the psychological reductionist tradition has seen first-person survival as a matter of a unified mental

[129] Strawson [1966], p87

life continuing over time. This position is not threatened by Cassam's argument. What is threatened is Parfit's view that a unified mental life is a product of thoughts and other mental events, interrelated in certain ways. Now, if we follow Cassam and do not allow these thoughts to have an independent existence, then we are lead to the conclusion that thoughts are always of a person-at-a-time, or a subject-at-a-time. This means mental unity over time can no longer be seen as a matter of connections between independent mental events, but it can be seen as a matter of the relation between persons-at-a-time.

This is a change to the Parfitian conception I favour adopting. In chapter six, where I offer my positive account, I will try to show how this change can be incorporated. But first, we must see how Kant's arguments effect the Parfitian conception.

5. Parfit and Kant.

Cassam argues that for I-thoughts to have any content, it must be possible to state the truth conditions for that I-thought, and to do this, one must know who the thinker of that thought is. Cassam made it clear that what is actually required for I-thoughts is a subject. For Kant, thoughts formally entail a subject, but nothing concerning the substantiality of the subject can be deduced from this. Putting Cassam's arguments to one side, what would accepting the Kantian conception entail for the Parfitian view?

Parfit argues that mental unity over time is a product of connectedness and continuity between individual thoughts. Kant's view is not necessarily opposed to the atomism of this position. But what makes each thought formally the thought of a single subject is that the "same sounding 'I'" can accompany each thought. It is in virtue of this formally identical 'I' that there is a unity of thought, and indeed the same sounding 'I' is a precondition for the unity of thought over time.

The way in which this is opposed to the Parfitian view can be seen through the words of Patricia Kitcher:

> His [Kant's] reflections on the nature of mental states show that we can acknowledge something to be a mental state only if we can acknowledge the existence of, not a mere bundle, but a synthetically connected set of mental states.[130]

In order for something to be a mental state, it all ready has to be part of a set of mental states connected by the same sounding 'I'. If this is correct, Parfit cannot be right to say that thoughts can exist somehow unowned, and only become the thoughts of a subject once they are suitably connected. An unconnected thought is, on Kant's view, not a thought at all. A Parfitian account of the unity of consciousness over time, therefore, would seem to require to take account of this feature of thought. It is simply not enough to say that it is only because of the relations between thoughts that the thoughts are connected. There is also something about the form in which thoughts are had which contributes to continuity of consciousness.

But, again, I do not see why this should threaten the core tenets of Parfitianism. If psychological connectedness and continuity is a result of the form in which thoughts are had, then so be it. The important fact is that the unified mental life over time can connect persons-at-a-time who are not identical. The formal identity of the "I-think" is distinct from the substantial identity of the person. The gap is still there to answer the first-person question of survival in terms different to that of the third person question of identity. In fact, Kant's conception can facilitate this distinction, as Kant places a wedge between the inner unity given by the "I-think" and the notion of substantial identity, which is concerned with objects of intuition. What we would have to change is the explanation of what a unified mental life consists in, not the views about the importance of a unified mental life and its distinctness from substantial identity, and it is the latter views, not the former, which I believe are more crucial to the Parfitian conception.

In my positive account, given in chapter six, I shall try to show that it is possible for a revised Parfitian account to incorporate Kant's ideas concerning the necessary unity of thought.

5. Conclusion.

When Parfit claimed that thoughts are independent, he could have meant one of two things. He could have meant that they are ontologically independent, literally capable of independent existence. Cassam's arguments cast considerable doubt on this claim. Cassam argued that at least one sort of thought, I-thoughts, entail a substantial subject of thought. This argument is not quite as strong as may appear. All we need are substantial subjects-at-a-time. Alternatively, he could have meant that they are conceptually independent, that they can be thought of without thinkers. Kant argued differently. Thoughts formally require a thinker both at a time and over time. But again, this is not quite as strong a claim as may first appear. Nothing whatsoever about the substantiality of this subject is entailed by the existence of thoughts.

There is nothing to prevent both Cassam and Kant from being right. My concern has been to show how a Parfitian conception is still possible if both Cassam and Kant are largely correct, as I believe they are. Some radical restructuring of the Parfitian conception is therefore necessary. The idea that mental unity at a time and over time is simply to do with relations between individual thoughts must go. At a time, we need to accept that thoughts are had

[130] Kitcher [1984], p11

by a substantial person. Over time, there is something about the way in which thoughts are had that explains the unity of thought over time. If we accept both these claims, the key Parfitian claims can still be retained. A unified mental life occurs where there is psychological connectedness and continuity between persons-at-a-time. This continuity is explained at least partly by the form in which thoughts are had. This unity does not presuppose identity, and it is this unity, not identity, which is important for the first person question of survival. For now, I have merely made room for this account. In chapter six, I shall show how it is possible. But first, we must turn to Parfit's conception of psychological connectedness and continuity and see how that is flawed.

Chapter Five

Parfit's Relation R.

Can the reductionist account for the unity of a person's life solely in terms of the thoughts, events and actions which constitute that person's life, in spite of the problems for the psychological reductionist outlined in chapter four? To answer this question I must first look at Parfit's own account of what is required for the unity of a person's life: what he calls Relation R – psychological connectedness and continuity. Firstly, as I discussed in the previous chapter, the thoughts and mental events which form the relata of this account are supposed to be in some sense independent of the subject who has these thoughts. I have already argued that this conception is flawed. However, in this chapter I shall set aside this criticism so that my further criticisms do not themselves rest upon these earlier criticisms. I shall be considering the concepts of mental connectedness and continuity and arguing that they are ill-defined and inadequate to fulfil the task required of them. I thus argue that Relation R fails to perform the function it needs to perform, namely, to account for the unity of a mental life. It is because I believe Relation R fails on its own terms that I want to keep apart the criticism of the relata of relation R from my criticisms of relation R itself. I contend, however,that there is room for an alternative Parfitian account which can do the work Relation R cannot. The lessons learned from this critical account will then be used to positive effect in chapter six.

1. Psychological Reductionism and Diachronic Unity

Traditionally, psychological reductionists have held that the identity over time of a person is determined by the existence of strong psychological connections of memory, personality and intentions and so on holding over time. Various different accounts have been offered as to what these relations consist in, but the fundamental point was agreed. However, it came to be believed that the sorts of relations considered vital for personal identity could possibly exist in cases where identity did not hold. For example, in the fission case, it is claimed that two persons at a later time can have memories, personalities and intentions continuous with only one earlier person.[131] This seemed to threaten psychological reductionism, as it suggested that identity could not solely be accounted for in terms of psychological relations. But Parfit countered by arguing that where the psychological relations which usually determine identity branch, although we do not have identity, we still have “what matters in survival”.[132] By doing so, he makes more explicit the psychological reductionist’s intuition that it is our mental life which really counts.

On Parfit’s view, the unity of a person’s life is determined by diachronic unity of consciousness. Although a brain and body are required to sustain this unity of consciousness, it is in virtue of the unity of the mental life that a person’s life is a coherent whole and the unity of consciousness over time doesn’t require the same body over time. To give an account of what is required for the unity of a person’s life therefore it appears the psychological reductionist need only give an account of what is required for the diachronic unity of consciousness. To avoid any confusion between unity of consciousness at a time and diachronic unity of consciousness, I will from now on talk of the latter as a unified mental life. The concept of a unified mental life is intended to be a minimal one, and so fairly unproblematic, even though the way in which it is filled out will result in various

[131] Because the concepts of memory, and to a lesser degree intention, seem to presuppose the continued existence of one particular person, the terms quasi-memory and quasi-intention were introduced to cover those cases where there is an apparent memory or intention without personal

different positions. A unified mental life is a temporally continuous, though not necessarily uninterrupted, series of mental events which are related to each other in the same important ways as are our normal mental lives. I need to make three clarificatory remarks concerning this definition. Firstly, by using the term 'mental events' I am remaining neutral as to whether or not mental events can exist independently or not. The term neither assumes nor precludes the possibility of mental events being discreet and/or capable or independent existence, although following the previous chapter it will be clear that I do not think them so capable of independent existence. I also follow Parfit in defining mental events broadly so that the term covers "even such boring events as the continued existence of a belief, or a desire,"[133] as well as occurrent thoughts and memories. The second point to note is the lack of a requirement for uninterruptedness. This is, of course, to allow for periods of sleep and other times when we lose consciousness. I assume that the idea of continued consciousness with such gaps is not problematic.[134]

Thirdly, and perhaps most importantly, is the requirement for mental events to be related in the same important ways as our normal mental lives. The paradigm for a unified mental life is the mental life of actual persons. The relations that connect different temporal parts of these lives can be described functionally or materially. What relates a memory with an earlier event is both that there is a chain of neural causes internal to the same brain stretching back from the memory to the event, which is the material relation; and that the memory provides an unmediated source of information (accurate or otherwise) of the person's past experiences, which is the functional relation. For there to be a unified mental life,

identity. See p13ff and chapter six. The concept of quasi mental states is however far from being uncontroversial. See Wiggins [1992] and Hughes [1975] for critiques of q-memory.

[132] See §6.5 for a detailed explanation of what is meant by 'what matters'.

[133] Parfit, p211

do the mental events in that life need to be related in the same functional and material way as the mental events in our lives are in fact related, or is only the functional or the material relation important? The answer to that question is clearly problematic and lies close to the heart of the debate concerning reductionism. For that reason, my definition of a unified mental life does not presuppose any one answer. But it does presuppose that the question has an answer.[135]

Parfit claims that what matters, and what usually determines personal identity in a world without fission, teletransportation or other forms of duplication, is what he calls Relation R. Relation R is defined as "psychological connectedness and/or continuity, with the right kind of cause."[136] Where Relation R holds there will be a unified mental life and therefore the life of a person. Given its importance in *Reasons and Persons* it is perhaps surprising that Parfit neither gives a thorough explanation of just what this entails nor explains why he introduces it. Rather in the manner of a character in Plato's dialogues he simply offers it as a candidate for the criterion of personal identity over time, expands upon it *ad hoc* as and when the argument demands and allows it to emerge as the best candidate among rival theories. In order to examine Relation R in detail, we need to get much clearer about what it is.

As I have said, I contend that essentially, Relation R is a description of what is required for a unified mental life. The building blocks of Parfit's account are particular mental events, be they thoughts, feelings or memories. In Chapter four we saw that Parfit claimed such experiences can be described independently of their subjects. Therefore to understand what Parfit means by Relation R we must take all the connected mental events to be discreet elements which come

[134] Locke [1694] makes this point clearly in Bk 2, Chapter 27.

[135] I partly answer this question in Chapter six.

together only in virtue of their interrelations. This is one area where Parfit's explanation of a unified mental life diverges from my general definition. Whereas my definition was neutral as to the independence of mental events, Parfit claims that they are at least conceptually independent.

Relation R requires three distinct elements: Connectedness, continuity and the right kind of cause. I shall briefly discuss what Parfit means by each one of these terms. All of the remainder of this section is exegetical and thus accords with Parfit's views, except in one very important respect. Where I talk about a unified mental life, Parfit talks about persons. Parfit can do this because, as I have argued, for the psychological reductionist, all we need to do to account for the diachronic unity of a person's life is to account for the diachronic unity of mental life. For Parfit, connectedness and continuity with the right cause are the only things that matter in survival. In effect this means a unified mental life is alone what matters. The fact that nothing in the account is lost when I substitute 'unified mental life' for 'person' itself demonstrates the virtual synonymity Parfit accords the terms. I however would prefer not to use 'person' interchangeably with 'unified mental life' so as not to beg the question as to whether there is more to the unity of a person's life than mental unity. While Relation R describes nicely what on the Parfitian view is required for a unified mental life, I would like to keep it clear in my exegesis of Parfit that we cannot assume that it also shows what is required for unity of a person's life.[137]

The first element in Relation R is psychological connectedness. Parfit describes three kinds of psychological connections; those of memory, intention and 'psychological features':

[136] Parfit, p216

[137] I shall examine more closely the issue of what matters in survival in Chapter six and will consider some of the other factors which unify a person's life in chapter seven.

> Between X today and Y twenty years ago, there are direct memory connections if X can now remember having some of the experiences that Y had twenty years ago. [...] There are several other kinds of direct psychological connection. One such connection is that which holds between an intention and the later act in which this intention is carried out. Other such direct connections are those which hold when a belief, or a desire, or any other psychological feature, continues to be had.[138]

Note that these connections are described functionally, without reference to their material basis. However, there must be a material basis, as connectedness implies a cause, and so psychological connection is a type of causal connection. But the material basis is covered separately when Parfit considers what is the right kind of cause for Relation R. But that there must be a cause is not at issue.

It is clear that Parfit's explication of connectedness is somewhat minimalistic. A memory connection, for example, just is that causal connection, whatever it may be, which holds between a memory and the earlier experience of which it is the memory. However, although more could be said, these brief definitions are sufficient for us to get a hold of what is meant by connectedness. I discuss problems with this idea in the next section.

The second element of Relation R is continuity. To get from connectedness to continuity, Parfit introduces the idea of overlapping chains of connectedness. We can understand this by considering this schematic diagram:

E^1————*M*———*M*| *B*———————————| *B*——————————...

B————————————————| E^3—————————*M*————*M*——————...

I———*A* E^2——————*M*———*M*| *I*——————*A* *E*———————————...

Here, the horizontal scale is time, and the lines represent psychological connection. A vertical line indicates that connectedness has ceased. For

[138] ibid, p205

example, *E* represents an experience and *M* the memory of that experience. The last time any experience is remembered,[139] it can be said that there is no longer psychological connectedness with the original experience, and so the chain comes to an end. *B* represents a belief, and the vertical line shows when that belief is no longer held. *I* represents an intention and *A* the later action which is the fulfilment of that intention. In the diagram, a small fraction of our mental lives is represented, and in a crude, simplified form. However, what it does illustrate is how, although most connections may only hold over a fraction of the time scale, they can overlap with one another to form a continuous chain of connections. In this portion of the chain, there are no connections that hold over the whole time period. But there is still continuity. At the end of this period, the person no longer remembers E^1 or E^2, but the chains of connection that still hold at the end of the period overlap with other chains, which overlap with still others that stretch back to E^1 and E^2. To adapt a more concrete, famous example, there was a boy who stole an apple, who grew up to be an officer who did a brave deed, who went on to become an old general.[140] The old general may not remember stealing the apple, but he remembers carrying out the brave deed, and the person who did the brave deed remembered stealing the apple. So even though there is no direct memory connection between the general and the boy, as there are such connections between the general and the officer, and the officer and the boy, so there is continuity of memory between the general and the boy.

Parfit claims that both connectedness and continuity are required for a unified mental life. We can explain this in a schematised and simplified form by returning to the psychological spectrum. Imagine the case where Smith's psychology is

[139] It is perhaps more accurate to say that the chain of connection ends the last time a person has the capacity to remember *E*, in which case the diagram would read "*E*-------*M*------|", but nothing hinges upon which of these two ways we interpret what a chain's ending means.

[140] The original example came from Reid [1941], reprinted in Perry [1975] p114.

replaced by Napoleon's, but let's imagine this occurs in 100 phases, 1% per day for 100 days. We can represent this diagrammatically using the same symbols as above, with the difference that the horizontal lines represent not single connections but chains of related connections:

```
      1      2      3      4      5   ...
A ——| ———————————————————————————...
B —————————| ———————————————————————...
C ———————————————| ———————————————...
D —————————————————————| —————————...
E ———————————————————————————| ——...  And so on.
```

Chain A, for example, could refer to memories of childhood, which we carry throughout our lives. Although there is almost certainly no distinct set of memory traces in the brain that corresponds to this group, we know that there is a fairly distinct set of memories that corresponds to it. As we are concerned with the psychology rather than neurology of consciousness, this lack of analogue in the brain doesn't matter.[141] Chain B could be fears and phobias, C aesthetic preferences, D political beliefs and so on. On day 1 Smith's childhood memories are replaced with those of Napoleon. After day 1 it is those memories that continue into the future, although, of course, they will gradually fade, as all memories do. On day 2 Smith loses his fear of spiders and gains one of heights. Again it is the vertigo that will persist in the future, continuing that chain of connection. And so on. It is clear that by day 100, no chain of connection will stretch back earlier than day 1. The resulting person, call him Smith–Bonaparte will then have no psychological connection with Smith prior to day 1. This means

[141] It could matter to the practical possibility of ever wiping away a distinct set of memories only, but my discussion in §3.2 about the importance of possibility should make it clear that this is not an important problem here. Mental holists would claim there is a more fundamental conceptual problem in thinking of thoughts as discreet in the way in which the thought experiment requires.

Smith–Bonaparte will have no memories, residual traits or beliefs from the period of time before the experiment. Given this total transformation, it could then be said that the original unified mental life has ended and has been replaced by another. But Smith-Bonaparte will be psychologically continuous with Smith, because the chains of connection all overlap. There is no sudden rupture with the past. Compare this with the case below, where the machine in the psychological spectrum case changes Smith's psychological features all at once:

A————————————————| ————————————...

B————————————————| ————————————...

C————————————————| ————————————...

D————————————————| ————————————...

...

In this example there is neither connectedness nor continuity. At a certain point all psychological connections terminate and are replaced by a new set. Because in our first example the chains overlap there is continuity. On each day of the experiment the subject will be continuous with the person of the day before. Only one aspect of their psychology will change. But we would still consider that at the end of the 100 days, the thoughts, memories and other psychological features of Smith were part of a different unified mental life than the one now lived by Smith-Bonaparte. One unified mental life will have been superseded by another. They would, however overlap: one extends from birth to day 99 of the experiment and the other from day 1 to death, the overlap thus occurring during the period of the experiment. In the second case, we again have two unified mental lives, but no overlap. There is neither connectedness nor continuity. In

However, my aim in this chapter is to defeat relation R on its own terms, and the way in which I subsequently revise relation R is not inconsistent with mental holism.

order for there to be one single unified mental life throughout all the subject's life, there would have to be both connectedness and continuity.[142]

We have seen what connectedness and continuity are and seen why both are required for a unified mental life. There is still the third element to consider: the right kind of cause. Without knowledge of the causes our account is incomplete. It would be a good idea to start by considering various different types of cause that could produce Relation R. The normal cause is the continued survival of the animal, and especially the brain, which is the bearer of the mental states. Of course, the explanation is not complete until we have a good explanation of how the brain does manage to produce memories and retain personality traits and so on. But this is a project for neurology, not philosophy. If reincarnation were possible, then transmigration of souls might sometimes be a cause of Relation R. If King Arthur were indeed to be reincarnated in the body of a twentieth century taxi driver, along with his heroic character and memories, then Relation R would hold between King Arthur and that taxi driver. A science fiction case would be the reading of information from the brain and the re-writing of this information back into the same or a similar brain, as described in the spectra thought experiments.

Before deciding which of these are 'the right kind of causes', I should stress that causation is a necessary part of Relation R. Relation R is connectedness and continuity of mental states and events and as such requires a causal relation between those states and events. It is not just a similarity relation. For example, let's say that I have a doppelgänger in a parallel universe. If I were to be run over and killed and just seconds later my doppelgänger stumbled upon a way into our universe at the exact spot where I had been killed, then there would be someone

[142] Over a very long period of time, lack of connectedness between the two times may not threaten a unified mental life, as long as there is enough continuity. For example, a person of seventy-five may share no connections with their childhood self. Whether this means that the child

just as capable of carrying on with my life as I would have been, had I survived the accident. This is very interesting, if implausible, but what should be clear is that Relation R does not hold between myself and my doppelgänger. There is no causal connectedness between my mental states and his, because my mental states have played no part in forming his. We do share type identical mental states, but Relation R is not just about type identical mental states but the relations between mental states.

Having established that causal connection is required for Relation R, we now need to ask whether any cause is acceptable or whether Relation R can only hold where there are certain particular causal relations. Parfit claims that Relation R with any cause is what matters. The problem for this view is best brought out by a thought experiment described by Douglas Ehring.[143] We are to imagine a device similar to the spectra machine, which is capable of recording the information about a person's brain states and then transferring these brain states into another person's brain. Ehring imagines this as a rather old-fashioned device where the brain state information is stored on tape. In the thought experiment, an extraordinary accident occurs:

> After encoding information on a tape from A after A's destruction, the tape is accidentally dropped onto another encoding machine. As a result the tape is damaged beyond repair. However, the tape's hitting the machine causes the machine to malfunction with the highly improbable outcome that this machine produces another tape exactly similar to the original.[144]

and the elderly person do not share in the same unified mental life is a matter of conjecture. See Lewis [1976].

[143] Ehring, p52

[144] ibid

If this tape is used to “programme” the brain of B, thus making B type identical to A, should we now say that Relation R holds between person A before and person B after?

The problem here is that although there is a causal chain between the original person and the person who has the taped information copied into their brain, this is a chance event, as is the person from a parallel universe appearing in the same spot you are knocked down dead on. The link between the knocked-down dead me and my doppelgänger is as weak, it seems, as the link between the original person and the person created by the transfer of taped information. In both cases, the nature of the later person is not determined by the nature of the previous person. Their qualitative identity is just a matter of chance.

Certainly it is a chance event, indeed almost a miraculous one. Obviously it is also so preposterously improbable that it would surely never happen. It seems, though, that there is no reason in this case to suppose that Relation R does not hold between the two persons before and after the tape. The chance element in the story doesn’t destroy the connection between the two. Think of other chance events. Imagine a surgeon attempting a tricky operation, slipping, and thus making an incision in just the right place. Or a painter after a particular aesthetic effect who falls onto his painting, smudges the paint and thus achieves serendipitously the effect he was after. In both these cases, a chance, unplanned, possibly unrepeatable event is the cause of a particular sought after effect. We might say such events are almost miraculous. But in both cases an earlier cause (the doctor, the painter) are linked to a later effect (the success of the operation, the creation of a great painting) via a chance happening. As Parfit points out, the reliability of a cause is irrelevant if the cause has the desired effect:

> Suppose that there is an unreliable treatment for some disease. In most cases the treatment achieves nothing. But in a few cases it

> completely cures this disease. In these few cases, only the effect matters. The effect is just as good, even though its cause was unreliable.[145]

In Ehring's case we have a cause which is extremely unreliable: it only works once in a universe's lifetime. But that one time it does work, the effect is the same as that when there is a normal reliable cause.

It certainly doesn't seem that the reliability of any cause gives grounds for distinguishing the right kinds of causes. But there is another point to Ehring's story, aside from the issue of chance. What seems to be the problem is that the person who results from the accident with the tapes could have come into existence whatever the nature of the original person. Just as the colour of an axe is irrelevant to the cut it produces, the data on the tape is irrelevant to the effect it has in the other tape. The nature of the resulting person is thus totally independent of the original person's, and so it seems ridiculous to suppose that those two people have the same kind of connection as holds between a normal person over time.

The problem with this objection is that if we accept that the nature of the two persons are quite independent of one another then it undermines the central premise of the argument: that the original person's brain-states caused the latter person's brain states. If it is irrelevant to the causal story that the tape which fell contained person X's brain state information, then on any account of causation, X's brain states are not a cause of Y's. But if, on the other hand, it is important to the story, then it is no longer true that the nature of X's and Y's brain states are independent. Then we are back to the situation whereby we simply have a very unreliable cause.

[145] Parfit, p287

Parfit thus seems right not to place any restrictions on which material causes are the right ones for Relation R. As long as the cause performs the right function, the material basis for it is irrelevant. This then leads to another problem which is, what is meant by the right function? What Parfit says in the quotation above is that what is important is the effect. Not that this takes us very far. What that cashes out as is the need for the cause to produce psychological connectedness and continuity. As long as memories are causally connected to earlier events, actions to intentions and beliefs and other psychological features persist, and that these connections overlap into continuous chains, then whatever the cause, that cause is of the right kind. So to answer the question of what is the right function, we need to examine more closely what is meant by connectedness, which I turn to in the next section.

For Parfit then, Relation R will hold with any material cause, provided that the material cause is between mental states. However, although Relation R holds whatever the cause, it still remains to be seen whether or not Relation R can matter whatever the cause. As I shall be rejecting Relation R in its current form, I do not need to answer this question. However,it will need to be asked of my replacement for Relation R.

2. Problems With Relation R.

If we were to ask the question, 'How are the experiences that form an R-related sequence connected?' we wouldn't find much of an answer in Parfit. The 'how' is presumably a question for neuroscience, if we believe that it is due to brain processes that we think, remember and so on. There is still another question though that does need to be answered, which is 'what does it mean to talk about mental experiences being connected?' As we saw in the last section, Parfit offers examples of mental connectedness – between an event and a later memory; intention and action; persistence of mental feature – but doesn't say what it is that makes these connected. We have already seen how they must stand in a certain causal relation to each other. But without more detail this will not do. A Jehovah's Witness' belief that they know the meaning in life stands in a causal relation to many people's beliefs that Jehovah's Witnesses are pests, but these are inter-personal relations and are thus not part of the same unified consciousness. The answer we may instinctively want to give is that the mental events are connected if they are part of the same unified consciousness. But this answer cannot be given, as Parfit is trying to explain why we have a unified consciousness solely in terms of the relation between the mental events. Having claimed that our mental lives are reducible to the mental events that compose them, to then say that mental events are connected because they are part of the same consciousness would be viciously circular.

Parfit's ideas on what R-relatedness entails actually takes us far closer to the Jehovah's Witness scenario than perhaps we'd think. Parfit seems to think that R-relatedness is simply a matter of thoughts causing other thoughts, memories of earlier experiences, fulfilling of intentions and so on. That this entails the possibility, even reality, of R-relations holding inter-personally, rather like in the case of the Jehovah's Witness, is something Parfit seems more than willing to accept. Consider this quote:

> My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less.[146]

[146] Parfit, p281

The image of the glass tunnel is certainly an effective metaphor for the the way in which our lives have a sense of being essentially separate from the lives of others. Despite the philosophical attack on the reality of an ‘inner’ private life that stretches from Wittgenstein and Ryle to contemporary eliminative materialism, most people still have a sense of their mental lives as something to which only they have direct access. Conrad’s line in Heart of Darkness, “We live, as we dream, alone,” still strikes a chord. If we consider death as The End, within these invisible walls that separate us from others, life does indeed seem to be a one-way ticket to darkness on an ever accelerating express train.

But Parfit suggests that, if his view of what people are is correct, this glass wall can be lifted. We will no longer feel as essentially separate from others as we may otherwise do. This is because for Parfit the relations that hold for himself over time may not just hold between moments of his own life, but between himself and others as well. We need to see how this is possible to properly appreciate how Parfit’s position is flawed.

First off, we must remember what a person consists of, on the Parfitian view. A human person consists of a body, a brain and a set of experiences which are R-related to one another. But it is in virtue of Relation R, not the body or the brain, that various different experiences are united in the life of a single person. Indeed, a body and a brain without Relation R cannot be a person, as there would not be a conscious being with a sense of its own past and future. That I am R-related to a past and a future person is then the most important thing about my existence as a particular person. Thoughts, beliefs and other mental states and events are R-related if they are connected and continuous, and have the right kind of cause. Mental events and their connections are thus the fundamentals from which Relation R is built. Intrapersonally, such a connection could be, for example, an experience causing a memory, the memory persisting and going on to effect decisions, beliefs and feelings. This is one chain of psychological connections. But there is no prima facie reason why such relations cannot hold interpersonally. And that there are in fact such relations is vital for Parfit to lift the ‘glass tunnel’. Consider how Parfit follows up his discussion of the glass tunnel:

> When I believed the non-reductionist view, I also cared more about my inevitable death. After my death, there will be no one living who will be me. I can now redescribe this fact. Though there will later be many experiences, none of these experiences will be connected to my present experiences by chains of such direct connections as those involved in experience-memory, or in the carrying out of an earlier intention. Some of these future experiences may be related to my present experiences in less direct ways. There will later be some memories about my life. And there may later be thoughts that are influenced by mine, or things done as a result of my advice. My death will break the more direct relations between my present experiences and future experiences, but it will not break various other relations.[147]

This is a very telling passage. It reveals just how broad the idea of mental connection and relatedness is for Parfit and how he explicitly does believe that some such relations hold interpersonally. Unless the relations that hold across his own life are of the same kind as those that hold between himself and other people, how can the glass tunnel be lifted? The relations which bind together experiences into a unified mental life differ from such interpersonal ones only in that they are more direct than those that hold interpersonally. Once more, Parfit is vague with his terminology. Although we can get a feeling for what is meant by saying that experience memory is a more direct relation than, for example, the relation between an idea and its influence on other people, Parfit offers no definition or criteria of directness. Are the connections that hold between the person who steps into and the person that steps out of the teletransporter direct? As Parfit claims Relation R holds in this case, we must assume that the answer is yes. But there is certainly nothing direct about the causal route such connections would take: from brain to scanner to radio wave to scanner and to brain again. Parfit seems to be helping himself to a question begging concept. Direct seems to mean something like ‘intrapersonal’, even though this cannot be an adequate definition, as Parfit is supposed to be explaining the personal solely in terms of mental events and their relations. If these relations were defined in terms of the personal, this would defeat the project. This illustrates the same point made about the right kind of cause. ‘Directness’ rather like ‘the right kind of cause’ must be unrestricted materially if Parfit’s position is to hold. The medium of causation or connection is irrelevant to their rightness or directness respectively. Nor are there any functional criteria offered for directness, and the only implied criterion, that directness means something like intrapersonal, would undermine the claim that relation R explains what it means for mental events to be of one person. The concept of directness is therefore rather empty. This is not so say that an alternative Parfitian position could not give substance to the notion of directness. Indeed, this is part of what I attempt to do in the next chapter. But suffice it to say for now that Parfit’s own account, which is the one currently under scrutiny, fails to breath life into this concept.

Psychological connections are distinguished solely in virtue of their being direct or indirect, strong or weak, where ‘strong’ means “there are enough direct connections,” i.e. “at least half the number of direct connections that hold, over every day, in the lives of nearly every actual person.”[148] On such an account, it is

[147] ibid

[148] Parfit, p206

clear that there is nothing to make intrapersonal relations intrinsically different to interpersonal relations, save in directness and quantity. And Parfit welcomes this, as it enables him to see his death as the end only of some of the mental connections he values. What I shall now do is offer some examples of what on Parfit's view must be seen as interpersonal R Relations, real and imaginary, and explain why Parfit's view entails them. Then we shall be able to see how Parfit's position is flawed.

One case can be taken directly from Parfit, from his discussion of quasi-memory (or 'q-memory'). The concept of q-memory[149] was employed as a response to the objection that as "it is part of our concept of memory that we can remember only our own experiences,"[150] memory cannot be part of a criteria of personal identity:

> I have an accurate quasi-memory of a past experience if:
>
> (1) I seem to remember having an experience.
>
> (2) Someone did have this experience.
>
> (3) My apparent memory is causally dependent, in the right kind of way, on that past experience.[151]

What was said earlier about the right kind of cause for Relation R also applies to 'the right kind of way' in (3). Let us consider a case of q-memory. In §3.1 we looked at the case where the normal neural cause of psychological connectedness and continuity was interrupted by a machine that recorded the neural states of the brain and then re-wrote these states back into the brain later on. Let us imagine that the tapes got accidentally spliced and, whilst sticking the tapes back together, a short section of one tape was mistakenly inserted into the wrong tape. After the brain had been 'reprogrammed', a person thus wakes up to

[149] This concept was first formulated by Shoemaker [1970].

[150] Parfit, p220

find that, although largely the same, he seems to remember going to Kew Gardens, even though he had never been there. He checks this out and finds that the person with whose tape his got confused has forgotten his visit to Kew. He should conclude that, by accident, he has a q-memory of another person's experience and is thus psychologically connected with this other person to a very small degree, but in exactly the same way as he is psychologically connected to himself at an earlier time. That is to say, the q-memories were continued first by normal neural means, then briefly by tape and then by neural means again. This story could be told with any other of the connections which form Relation R. The person could acquire a belief, and intention or a personality trait. Thus, on Parfit's own account, the connections which form Relation R can hold interpersonally.

This is a piece of science fiction. But there are good reasons for saying that if Parfit is right, some of these connections hold interpersonally as a matter of fact. When people have a conversation, watch people performing, listen to them talking or even read their words then what is going on in the mind of one is affecting what is going on in the mind of the other. When I talk about what I believe, through the faculty of hearing, you are affected by these beliefs which will in turn alter your own beliefs, even if it is only to give you the new belief that I believe X. The cause of the connection here is not just neural firing but speech and sense perception. But it is still true that a psychological event is connected with another, only now the two psychological events are in different people. And on Parfit's view, any cause will do. If two or more people live very close lives, these connections will form overlapping chains. Parfit wants to say that, although this is the case, we are not R-related to other people. This is because, firstly, these are not direct connections, and secondly, there would not be strong connectedness. However, it should be clear that these are not important

[151] Ibid

qualitative differences, but matters of quantity and directness, whatever that is supposed to mean. The relations that hold interpersonally are of the same kind as those that hold for Relation R.

So far, we have seen how the connections which form Relation R can hold interpersonally. But if Parfit is right there is at least one possible case where we can have full-blown Relation R interpersonally. Shoemaker considers the fission case and notes that here "there are two later persons who are psychologically continuous with the owner of the original brain."[152] But as we saw in chapter 3, and Shoemaker agrees, neither of these resultant persons can be considered the same as the original person. Parfit too acknowledges that “the best description [of the fission case] is that neither of the resulting people will be me.”[153] Therefore we straightforwardly have a case of psychological connectedness and continuity – Relation R – interpersonally, because the pre-fission person is R-related to two future, but different, persons.

We are now in a position to formulate three claims that Parfit’s view rests on. They are:

(1) Mental events are independent. (See chapter four)

(2) There are no intrinsic differences between intrapersonal and interpersonal psychological connections. The differences are solely in directness and the strength of chains of connectedness, which is determined by number of direct connections.

(3) A unified mental life just consists of enough directly connected mental events forming overlapping chains.

It should be clear that Parfit’s idea of Relation R depends upon claim (1) being true. In chapter four, I argued that it was not. Let us suppose that there I was

[152] Shoemaker/Swinburne, p120
[153] Parfit, p26

wrong. Parfit’s view still depends on claims (2) and (3). The weakness of (2) is that there is no clear meaning to the word ‘direct’. Parfit clearly doesn’t mean by this that the connections should be neural, for example, as many of his thought experiments utilise non-neural connections. Nor is any functional definition offered, and it is hard to see how one could be offered without circularity and/or artificial distinctions being made. Furthermore, connectedness is something which on Parfit’s own account can hold interpersonally. There is nothing intrinsically intrapersonal about psychological connectedness. What we are left with is then essentially the claim that there are such things as psychological connections and that if enough overlap they will form continuous strong chains of connectedness. And by claim (3), we can see that this will result in a unified mental life. Does this claim stand up?

Certainly claim (3) doesn’t follow from claims (1) and (2). However, unless something important about the nature of mental events and/or the connections between them is missed out by claims (1) and (2), (3) looks like being the only claim compatible with them. If there are just mental events and connections between them, and there are unified mental lives, then these lives must just be a product of the events and their connections. The problem is that in (1) and (2) there is nothing in the mental events and their connections that makes them necessarily of a single unified mental life, or even, if we take Parfit’s claim about the independence of mental events seriously, of any subject. We are then to suppose that it is simply by their being bundled in sufficiently large numbers that a single subject emerges. There is such a gap here between what goes into the mix – events and their connections – and what comes out – unified mental lives – that it is hard to make the position seem plausible. We have mental events that can exist unowned and connections that can be interpersonal and simply by putting them together we are to believe that they become owned and intrapersonal.

Relation R is supposed to explain what is required for a unified mental life. But as the connections which form Relation R can hold between patently different unified mental lives, and the only difference between these and intrapersonal ones is directness – which is an obscure concept – and quantity, Relation R fails to explain how a unified mental life does emerge. And that it should explain how this is possible is one of the requirements we can justly put on reductionism, as I explained in §4.1.

Unless we think this really is the best account we can give, claims (1) or (2) must be wrong. I have already argued that (1) is wrong. I could have made a mistake. But even if (1) is defensible, (2) is not. There must be a better account of connectedness if we are to get from unowned experiences to unified mental lives. Most importantly, we must put flesh on the ides of direct connectedness. Only then will we be able to explain how it is that some connections are interpersonal and others intrapersonal, and then there may be some hope of explaining how connectedness helps constitute a unified mental life.

The conclusions of this and the last chapter seem fatal for the Parfitian. However, although it does mean Parfit’s precise position cannot be accepted, I believe a broadly Parfitian account can still be salvaged. There are two reasons for saying this. Firstly, I have suggested already how it could be possible to offer a Parfitian (in the precise sense defined) account of personhood without undertaking Parfit’s final reduction of a mental life to the thoughts and feelings that comprise it. Secondly, as I pointed out earlier, it appears to me that the fundamental instinct of the psychological reductionist is that it is our mental life which really counts. If this instinct is a genuine insight, then it can clearly be preserved if the details of Parfit’s account need to be changed.

What then needs to be done to salvage the Parfitian position? We need an alternative to Relation R that is up to the job of explaining what is required for a unified mental life in terms of psychological connectedness and continuity. But the relata of such an account must not be independent thoughts and experiences and the notion of connectedness must be adequately explained. Such a relation must also retain the features of Parfitianism. It is to the task of describing such a relation that I now turn.

Chapter Six

The Parfitian Conception Revised

So far I have criticised Parfit's account of psychological connectedness and continuity whilst maintaining that, once these errors have been rectified, a distinctly Parfitian account of personal survival can be retained. In this chapter I attempt to show how that can be done. To begin with, I will outline the Parfitian features I believe can be retained and the state of my argument so far, and then indicate how I intend to proceed.

Parfit's position has three key features. The first of these is an explanation of what is required for a unified mental life over time in terms of psychological connectedness and continuity. Although I reject Parfit's own version of what this explanation consists of – Relation R – as long as it can be replaced by another explanation which performs the same function, then this distinctly Parfitian feature will remain. Secondly, this relation is distinct from the identity relation. This means that the fact that there is a unified mental life that connects person X at T^1 and person Y at T^2 does not entail that person X is person Y. Thirdly, it is claimed that it is the relation which supports the unified mental life which matters in survival, not identity. I shall return to precisely what is meant by "what matters" in section five.

I claim that these three features of Parfit's account can be retained, even though I have rejected the details of Parfit's account of Relation R. I have outlined two key failings of Parfit's position. Parfit argues that, in my terminology, a unified mental life is a function of mental connectedness and continuity, where the relata which are connected are individual mental events. The first problem is the account given of the relata. Thoughts and mental events are supposed to be independent of the person who has them. However, if Kant is right, then there is a formal requirement for all thoughts to have a subject. Furthermore, if Cassam is right, then at a time, I-thoughts actually entail a substantial subject. I believe these views to be more credible than Parfit's. The second problem is the account of connectedness. This notion is simply not properly explained by Parfit. If connectedness is to be the force which binds together a unified mental life, it must be more thoroughly explained.

It is clear that neither of these two errors rule out a position which includes the three Parfitian features outlined above. The fact that psychological connectedness and continuity is not satisfactorily explained by Parfit's account does not rule out the possibility that it can be satisfactorily explained and can be a relation distinct from the identity relation. My aim in this chapter is to offer such an explanation. The key to my account is the rejection in chapter four of the idea that we can account for a unified mental life over time in terms of the relations between thoughts. Rather, we must consider relations between persons-at-a-time. Consider a person X at T^1 and a person Y at T^2. There are many ways in which they may be related. Person X may be identical to person Y. Person X may be psychologically connected to person Y, but not identical to Y. The nature of this psychological connection can take many forms. The connection may simply be that the thoughts of X have influenced the thoughts of Y, via conversations between them, or by a book X has written, for example. A science fiction example is that Y may have had some of X's memories transferred into her brain. Finally, Y may be psychologically connected and continuous with X. Parfit's thought is that the latter is possible, even if Y is not identical with X, and that if this is the case, the lack of identity does not prevent it being true that Y is a survivor of X. What I attempt to do in this chapter is argue for a way in which this can be true, without relying on the problematic account of psychological connectedness and continuity given by Parfit and hopefully without making Parfit's mistakes. By making my account one concerning the relations between persons-at-a-time, rather than between thoughts and other mental events, I aim to avoid the unwanted mental atomism of Parfit's account.

It is worth outlining now what I see this replacement relation to be. I propose a replacement of Relation R which I call the I* relation[154]:

> The I* relation holds where there is (a) direct phenomenological connection over time between persons; (b) functional indexicality between persons over time; (c) such connections are held both forwardly and retrospectively, forming a coherent internal narrative.

In outline, direct phenomenological connectedness over time is a memory relation, whereby the person at a later time remembers the experiences of the person at the earlier time as if they were her own memories. By functional indexicality I mean that all thoughts of the I* related persons are causally efficacious with regards to future actions, in the right kind of way. Functional indexicality and direct phenomenological connectedness are forms of direct psychological connection over time. These connections can hold to a lesser or greater degree. We know if they hold to the correct degree if the third condition of the I* relation is met. i.e. if there is a coherent internal narrative, by which I mean that each I* related person can make sense of all the experiences of the I* related persons as being part of a single, continuing life.

Firstly, I will discuss each part of the I* relation in turn, explaining why they are necessary and how it avoids the problems of Parfit's account. I will then consider

[154] This terminology is borrowed from Rovane [1990], who introduces the pronoun I* to "range over persons with whom I am psychologically connected...and would refer to them without implying they are identical with me." (p368) My account of the I* relation is in many ways an expansion of Rovane's concept of I*, which she herself leaves undeveloped.

how this account offers certain improvements over Relation R and will then turn to the question of whether the I* relation is “what matters in survival”.

1. Direct Phenomenological Connectedness.

Direct phenomenological connectedness is concerned with the form in which we have memories of past experiences. On my definition, X is directly phenomenologically connected with Y iff X remembers Y's experiences as her own.[155] Although diachronically this is a memory relation, to make clear what is meant by having an experience as one's own, it is a good idea to first consider some synchronic examples. If I stub my toe, I experience pain. It seems to add nothing to the description if I say that I experience that pain *as my own*. But now consider if you stub your toe, and I experience that pain. There are many ways in which this could be true. I could have a kind of secondary, empathetic pain. This is often the case when someone suffers a heavy blow, and a witness says something like, "I really felt that." It could mean that I mysteriously feel a pain in my toe at the same time as you feel pain in yours. But it could be the case that I actually experience the pain in *your* toe. We experience pain in our own bodies as pains coming from these parts of our bodies. But our own proprioceptive awareness can be altered. Amputees often feel pain in a part of their body which is no longer there. Conversely, people with artificial limbs can come to feel sensations in these limbs. So there is nothing logically impossible in my proprioceptive awareness encompassing, for a while at least, your toe. In such a case, however, even though the pain is in your toe, I would experience that pain as my own. It would be as if I had stubbed my own toe, but that my toe was for some reason, several feet away from the rest of my body. In such a case, I would be, momentarily, directly phenomenologically connected with you.

[155] This phrase is used by Braude [1991] who writes, "State x is autobiographical for S' =df 'S experiences x as its own." (pp87-88)

Consider next a case in which I see through your eyes.[156] I may be in London, while you are in Paris, looking at the Mona Lisa. Although it is not I who is looking at the painting, I will experience looking at the painting as my own experience. It will be as if I were looking at the painting myself.[157] Again, that would make me directly phenomenologically connected with you.

In these examples it is clear that there is direct phenomenological connection to a limited degree, as X does not experience all of Y's experiences as her own. However, the examples are merely intended to show what is meant by experiencing as one's own and to illustrate how one could conceivably experience someone else's experiences as one's own. Now we must turn to the more important diachronic case of memory. The archetype of direct phenomenological connectedness *over time* is quasi-memory, or q-memory. Consider Parfit's q-memory example:

> *Venetian memories*. Jane has agreed to have copied in her brain some of Paul's memory traces. After she recovers consciousness in the post-surgery room, she has a new set of vivid apparent memories. [...] One apparent memory is very clear. She seems to remember looking across the water to an island, where a white Palladian church stood out brilliantly against a dark thundercloud.[158]

In such a case, Jane is said to experience the memories as her own. This does not just mean that the experience of remembering is experienced as her own, but that she recalls the experience remembered, the sight of the Palladian

[156] Strawson [1959] discusses a similar thought experiment, pp90-91.

[157] One difference will be that I would not be able to control the direction of my gaze. But this is unimportant. If I were strapped to a wheel chair, with my head constrained and eyes propped open, and then taken around the Louvre against my will, that would not prevent my visual experiences being experienced as my own, and in this case, actually being my own.

[158] Parfit, p220. It has been argued, by Wiggins [1992] and Hughes [1975] for example, that such a story is incoherent. However, these criticisms are usually against the possibility of discreet memories being transferred, or of the subject being able to have memories not considered to be

church, as if it were her own experience. This is a purely phenomenological claim. Parfit, borrowing Peacocke's expression, says that the quasi-memory comes in the "first-person mode of presentation".[159] That is to say, the quasi-memory is had from "the seer's point of view"[160]. Although the quasi-memory may not be of something the person themselves experienced, it is presented to them from the viewpoint of the person who did have those experiences. In the same passage, Parfit gives as another example the occurrence in dreams of seeing oneself from the outside. Here, one has a point of view which is not that of oneself, but is nonetheless first-personal in nature. In such dreams we see ourselves as if we were someone else looking at ourselves.

For Jane to q-remember the experience as her own does not require her to believe it was her own. Jane could think, "I q-remember seeing Venice, but it was Paul, not I, who saw Venice". It is enough for her to remember seeing Venice as if it were her who actually did see Venice for her to be directly phenomenologically connected with Paul.

If q-memory were possible, then the person who q-remembers X and the person who experienced X would be directly phenomenologically connected. There is one respect in which q-memory differs from the synchronic examples given above. With q-memory, I could say, "I remember seeing X, but it was not I who saw X." But with the Louvre and toe examples, it is not obvious that the experiences one has as one's own are not also actually one's own. There's an oddness to saying, "My experience of seeing the Mona Lisa is as if I am in front of the Mona Lisa, but it is not I seeing the Mona Lisa," which is absent from the q-memory case. But the difference is not important. The intention is to demonstrate

their own. Because my final position does not require the denial of either of these two criticisms, I need not respond to them.

[159] Parfit, p221; Peacocke [1983].

[160] Parfit, *ibid*

what it means to experience something as one's own experience, *whether or not* it is actually one's own experience. The synchronic examples offered are at the very least problematic, in that it is debatable whether we should claim that I am seeing the Mona Lisa or that the pain in the toe is my pain.[161] But these examples are only illustrative, and in the crucial diachronic case of q-memory, it is very clear that the experience Jane q-remembers is not her own experience.

This definition of direct phenomenological connectedness is incomplete as it stands, as it will run into familiar problems with the possibility of deviant causal chains. What if, for example, your verbal recollections of your holiday are so vivid that it comes to seem to me that I remember the holiday as if it were I who went there, and that my apparent recollections correspond to what you did actually do and see? Having seen your photos and heard your descriptions, I could have a memory-like experience qualitatively identical to that of yours. Would that mean we have become directly phenomenologically connected? The problems here are familiar. There is a need to make the memory dependent upon the original experience in such a way as to rule out such possibilities without making the definition circular or arbitrary. I cannot offer such a definition here. But however difficult formulating such an account may be, I do not think it unreasonable to suppose that a significant distinction can be made between a memory-type experience that is built up from third-party information and a memory-transfer. Specifying exactly what that difference is I must leave to others.[162]

[161] Peacocke's [1979] (p99ff) discussion of similar problems of perception, such as whether or not I see a car in my mirror, or a person on television, should make clear to those unconvinced that this is a problematic area.

[162] Peacocke has formulated conditions for a causal chain to be non-deviant. The main condition is the requirement for the effect to be differentially explained by the cause, where "*x*'s being ø differentially explains *y*'s being ψ iff *x*'s being ø is a non-redundant part of the explanation of *y*'s being ψ, and according to the principles of explanation (laws) invoked in this explanation, there are functions... specified in these laws such that *y*'s being ψ is fixed by these functions from *x*'s being ø." (Peacocke [1979], p66. The second main condition is for "stepwise recoverability" (ibid p80). I believe that on this account, causal chains of psychological connection are non-deviant.

What should be clear at this stage is that direct phenomenological connectedness on its own is not nearly enough for a unified mental life. No matter how many experiences we q-remember as our own, if we don't think of them as our own experiences then the connectedness between ourselves and the actual experiencer is going to be very limited. Furthermore, there is much more to a unified mental life than retrospective memory. There are beliefs, desires, intentions and so on. We must see how direct phenomenological connectedness fits in with the other elements of the I* relation before we can fully appreciate its role.

Parfit, in his account of psychological connectedness, considered memory and then tried to fit other thoughts and mental contents, such as intentions, into the same framework. It is a criticism of Parfit that memories and intentions are disanalogous.[163] I believe memories to be in a category of their own, precisely because, diachronically, they have a phenomenological element lacking in other mental contents. I can remember something as if the original experience of that something were my own experience. But what could it mean, for example, to experience an earlier intention or belief as one's own, unless that simply means that we remember them? Memory provides the only phenomenological access to our pasts. The way in which intentions and beliefs unify our mental lives over time is, I contend, quite different from the case of memory. But intentions and beliefs do form part of a unified mental life, nonetheless. The second element of the I* relation, which I now turn to, explains how this can be the case.

[163] See Rovane [1990]

2. Functional Indexicality.

Consider once more the person who enters the teletransporter. The person who steps out on Mars will be directly phenomenologically connected with this person, because he will remember the experiences of that person as his own. This will be the case even if this person is convinced that teletransportation destroys people and creates new ones in their place. This belief cannot stop the way in which the memory experiences are had.

Intentions are not the same. If the person enters the teletransporter intending to watch a movie and the person steps out intending to watch the same movie, this person is not experiencing the intention simply *as* their own, it *is* their own. The cause of the intention may be the earlier person, but if the man on Mars intends to see a movie, that can only be his intention. In this respect, if the man on Mars believes that teletransportation kills people, he doesn't have to watch his language so closely when ascribing intentions to himself as he does when he recalls past experiences. While he would have to be vigilant to stop himself thinking that what he remembers is what he did, (which will become increasingly difficult after time, when he will not be able to automatically know which memories are pre-teletransportation and post-teletransportation), he has no such problem with intentions. Whatever he intends to do, he intends to do.

However, there is clearly a link between the intentions of the reluctant teletransporter user on Earth and the man on Mars which is utterly unlike the link between the intentions of two different people and quite a lot like the the link between a single person's intentions over time. To understand this link, let us return to Earth again. Our reluctant "traveller", let us call him Bob, has arrived to find his shuttle trip cancelled, and has been offered teletransportation instead. He is on a business trip, and his boss decides to take up the offer, but our man still refuses, believing that the machine will kill him. The boss begins to make threats.

Bob thinks to himself, “If he makes me use that machine, I’ll prosecute him for murder.” In his temper, Bob fails to see the lack of logic in this thought.[164] Bob’s beliefs entails that if he is teletransported, he will be killed, and so won’t be able to prosecute anyone. Perhaps at some level Bob does believe what is going to happen is going to very much like survival. His boss then punches him in the head, drags him into the teletransporter and the machine is turned on.

A little later on Mars, Bob (a different Bob?) walks out of the booth fuming. “Right, that’s it,” he screams, “I’m charging you with murder.” There are many reasons why he may change his mind. Why should he be concerned with the murder of someone else? Furthermore, if his prosecution is successful, that would establish that he isn’t Bob, which may leave him without any of Bob’s possessions, or maybe even the wife he loves, as much as Earth-Bob did. But the point of the thought experiment is much simpler. No matter what we believe about the identity of the person on Earth and the person on Mars, the intentional link is as if they were one person. Normally, if I want to do something, that I intend to do it, continue to intend to do it, and the circumstances arise when I can do it is enough to ensure that I do that something. My intentions are casually efficacious with regards to future actions. That is to say, *ceteris parabis* nothing other than my forming an intention to do something is required for me to actually go on and do that something. In this example, the same is true for Bob on Earth and Bob on Mars. Earth-Bob’s intention to prosecute is enough to lead Mars-Bob to prosecute, so long as no change of mind occurs, such as often happens with normal intentions.

[164] This lack of logic isn’t necessary to the thought experiment. He could equally think, “If he makes me use that machine, my replica will prosecute.” However, the logically inconsistent version makes the point rather nicely that no matter what we believe about this machine, we know that what comes out at the other end is psychologically continuous with us, in the same ways in which the person who wakes up each morning is with the person who went to sleep the night before.

Here then, the earlier intention is the cause of a later intention or action just as if the earlier and later intentions and actions were the same person's. This is not because of the phenomenology, as is the case with memory, but because of the way intentions function. If they function as if they were the intentions and actions of a single person, then we have what I call functional indexicality. This is why Earth-Bob's logical error is unimportant. He thought, "I will prosecute," and later Mars-Bob thought, "I will prosecute." The logic of identity requires that Earth Bob's "I" and Mars-Bob's "I" refer to different persons. But the "I"s here function just as though they do refer to the same person. There is the formal identity of the "I" explained by Kant.

Continuity of other psychological features can also be explained in terms of functional indexicality. Beliefs, dispositions, phobias and so on all lead one to behave and react in certain ways to certain situations. For example, if Earth-Bob believes Cliff Richard to be the greatest rock star in the world, and Mars-Bob meets Cliff Richard, barring any radical reappraisal of his view, Bob will be awed and amazed by this meeting. Beliefs form and develop over time. Mars-Bob's reaction will be just as Earth Bob's would have been, had his belief been allowed to develop and his life had not been cut off by the teletransporter. Again, because the belief functions in the same way, I say that there is functional indexicality between Earth-Bob and Mars-Bob.

Functional indexicality adds to direct psychological connectedness the other mental features of intention, belief, desire and so on which direct phenomenological connectedness leave out. What is now needed is an account of how these different factors must blend together, in order for there to be a unified mental life. This is explained in the third element of the I* relation.

3. The Internal Narrative.

Parfit's account included psychological connectedness and continuity. The first two elements of my own account have described two types of psychological connectedness. For Parfit, continuity was a product of enough overlapping chains of psychological connections. To a certain extent, this is true of the account being offered here. Consider again the Venetian memories case. Although Jane is directly phenomenologically connected with Paul, this connection is a small one. It only consists of a small set of memories. There is no way in which there is a unified mental life that connects Jane and Paul.[165] Even if Jane shared all Paul's memories, this would not be so. We would still require functional indexicality. If the relations of direct phenomenological connectedness and functional indexicality held between Jane and Paul to the same, or very nearly the same, degree, as is normally the case in a single person, then we could say there is a unified mental life. Jane would remember what Paul did as if she herself had done the things she remembers. Paul's intentions would be as causally efficacious to Jane's actions as they would have been for his own actions. What Paul's beliefs disposed him to do, they dispose Jane to do. Under such circumstances, we would have continuity as well as connectedness, and thus a unified mental life. But we must consider more carefully just what this continuity requires.

I want to claim that continuity is the existence of a coherent internal narrative. Although this is a product of direct phenomenological connectedness and functional indexicality, we need the concept of the internal narrative to decide whether these psychological connections hold to the correct degree. Consider first how it is a product of direct psychological connectedness. Where there is a high degree of direct phenomenological connectedness and functional

[165] I defined a unified mental life in §5.1 as "a temporally continuous, though not necessarily uninterrupted, series of mental events which are related to each other in the same important ways

indexicality, it seems there cannot but be a coherent internal narrative. It would help to consider what a normal internal narrative is like. I remember being born and being brought up on the Kent coast. I know that my choice of A level exam subjects was partly a result of my experiences in my O level exams, and that I went to a University which I choose for various reasons, and so on. All fairly humdrum stuff. But the notion of an internal narrative is a hum drum notion. It is the sense we make of our lives from the inside, and unless we lead exceptional lives, this is often, for better or for worse, a very hum drum affair. Now if all my memories fit together in this smooth way and my actions reflect a consistent, if evolving, set of intentions and beliefs, it is very probable that I will have such an internal narrative.

But sometimes we need to invoke the concept of the internal narrative to judge if there is the right amount of direct psychological connectedness. I have already pointed out how if Jane is only directly phenomenologically connected to Paul by a few memories, she will not be I* related to him. We need more or less as much connectedness as we normally have between a single person at different times. The requirement for an internal narrative is one way of helping us to understand just how much is enough. If Jane is directly phenomenologically connected and functionally indexical with Paul such that Jane has a coherent internal narrative that stretches back for the present back through Paul's life, the we will know that there are enough of these relations for us to say that there is a unified mental life that links Jane now and Paul in the past.

On the other hand, too much can also cause problems. If Jane does not lose her connections with her earlier self, then her internal narrative will go askew. She will remember, for example, not only visiting Venice in June, but visiting New York at the same time. She will in effect have two sets of conflicting memories. She will

as are our normal mental lives." See §5.1 for a full explanation of this.

also have conflicting desires and beliefs. So as well as requiring enough of the right connections with the person to whom the latter person is to be I* related, there also has to be very little or no such connections with any other earlier person. Specifying a quantity of connectedness is very difficult. What we can do is use the requirement for a coherent internal narrative as a kind of requirement for there being the right amount of connectedness between each directly psychologically connected person.

Given the circumstances of their story, whether Jane and Paul could ever be fully I* related is a difficult question. There are several obstacles to there being a coherent internal narrative that links their lives. Most obviously, she will remember being a man. She will remember living in a different house. How is it, she may ask herself, that I have been planning to be a rock star for all these years, and yet there isn't a single musical instrument in the house? Of course, what it will feel like for her is that she is Paul, waking up in the body of a woman in a strange place. The reason why this could threaten one's sense of identity is precisely, I claim, because it undermines the internal narrative we have of our own lives. Much will depend upon what Paul knew before the operation. If he knew his brain would be wiped and the states transferred into a woman, then maybe it wouldn't be such a problem. Whatever would happen in this case, it is clear that without the internal narrative, our mental lives cannot have the unity which they normally do.[166]

Briefly, a word about "coherent". Just how coherent does the internal narrative need to be? This is a question I feel unable to answer. What is clear to me is that without an internal narrative, there cannot be a unified mental life. But our

[166] The importance of a narrative to a subject's life history has been stressed most notably by Alisdair McIntyre [1985]: "The concepts of narrative, intelligibility and accountability presuppose the applicability of the concept of personal identity [...] The relationship is one of mutual presupposition. It does follow of course that all attempts to elucidate the notion of personal identity

narratives can be more or less structured and coherent. It is also probable that the amount of structure required varies from person to person. Enough is simply enough for the person involved to make sense of that narrative. Just as some people can make sense of the narrative of Leopold Bloom's life in Joyce's *Ulysses*, whilst others require the straightforward certainties of *David Copperfield*, so our abilities to make sense of our own internal narratives may depend upon our own capacities for self-interpretation. Ricoeur talked of life as "a story in search of a narrator".[167] How we "narrate" the stories of our own lives is almost certainly not the same from person to person.

Direct phenomenological connectedness, functional indexicality and a coherent internal narrative together seem enough to provide for a unified mental life. If X remembers Y's experiences as her own; Y's desires, intentions and beliefs develop and function as if they were X's; and X can make sense of the life which stretches back and connects with Y's life as a single life, then surely there is a unified mental life that links X and Y. In this chapter I have constantly used variations of the expression, "as if it were one's own". This expression is the essence of the I* relation, and the three parts of the I* relation are attempts to enumerate the factors that together give the impression of a unified mental life which appears as one's own. In teletransportation we have a clear example of how two numerically distinct persons, one on Mars and one on Earth, who can look backwards and forwards respectively to a numerically distinct person and yet see that person's life as their own. In such a case there cannot but be a unified mental life.

Now I would like to turn to how this account of a unified mental life improves on the Parfitian conception.

independently of and in isolation from the notions of narrative, intelligibility and accountability are doomed to fail. All such attempts have."(p218) See also Ricoeur [1975] and Glover [1988].

4. The I* Relation v. Relation R.

As I have explained, the need to formulate the I* relation as a replacement for Relation R was primarily motivated by the inadequacy of Relation R's account of connectedness, continuity and the relata of Relation R. These could be described as technical problems, in that they are not problems with the three basic features of Parfitian account I outlined, but with the details of how a diachronic mental life should actually be described. However, having made these alterations, we can see several other advantages the I* relation has over Relation R. In general, the I* relation explains better why it is that certain hypothetical cases are importantly different from cases of normal survival. In this section I explain why it is that the I* relation is a less liberal principle than Relation R, and why that illiberality should be welcomed.

Firstly, we should consider fission. Rovane[168] has pointed out how fission undermines some of the most basic features of our lives. Consider intentions. Though it seems possible for someone who knows that they are going to split to formulate intentions for each of the two fission products to fulfil, these intentions could not function like our normal ones. Firstly, our intentions assume that we will not split as they assume only one person will be there to carry them out. In certain cases, such as intending to be the first person to do something, it *requires* that only one person be there. Secondly, and more seriously, the way in which we fulfil our intentions in a fission situation would differ quite fundamentally from the way we fulfil normal intentions. For example, normally, if I intend to write a novel, in order to fulfil that intention it is enough that the intention continue and I act upon it. But if I know I am going to split, I may decide to form the quasi-intention that

[167] From the title of Ricoeur [1991].

[168] Rovane [1990]

only one of my fission products write that novel and that the other travel the world. In this case, acting upon the continued intention is not enough to fulfil it. This is because both fission products will wake up with exactly the same desires and wishes as the pre-fission person. It will then be necessary for them to work out, or decide, which of them should fulfil which part of the intention. It may be that the decision made was that the person who woke up on the left hand side of the other would write the novel and the person on the other side travel the Earth. Maybe a coin was to be tossed. But whatever the decision, the fission products will not know what was quasi-intended for them until they either toss the coin or see which side they are on. But no such procedure is required with normal intentions. The fact that the intention is remembered and continued is enough.

Parfit’s Relation R does not account for this difference, as it doesn’t explain what sort of connection must hold between an intention and a later action. However, the I* relation does. I maintain that the intention must be causally efficacious, in the same way as ordinary intentions are causally efficacious, if two persons are to be fully I* related and thus be part of a fully unified mental life. In the fission case, this condition is not met, as Rovane explains. What this means is that in the case of fission, there is rather less mental continuity than there is in normal cases, because there is rather less functional indexicality. This is one reason to suppose that fission is not like ordinary survival. Now, the importance of this difference is another matter. Even in ordinary survival, there can less continuity than is normal. An amnesiac, for example, may need to make notes for themselves to carry out their own intentions, for example. This does not mean that the amnesiac cannot enjoy a unified mental life, but it does mean that it is not as unified a mental life as is normally the case. In my opinion, the same is true in fission. We have rather less mental continuity than is normal, but enough for us to say that there is a unified mental life connecting the pre and post fission persons.

One reason for holding this is that there would still be a very clear internal narrative for both fission products. But however important we decide the difference is, there is a difference, and my account acknowledges this difference in a way in which Parfit's does not.

These considerations also bring out the point that all parts of the I* relation can hold to a lesser or greater degree. Neither functional indexicality nor direct psychological connectedness is an all or nothing relation. Consequently, the degree to which the internal narrative of our lives coheres can vary.

A second reason for favouring an account of the kind I have suggested is that, taking as it does its standards from normal cases of survival, it corresponds more closely with what we believe about real people.[169] The kinds of relations required for there to be an unified mental life that links two numerically distinct persons are the same as those required for a unified mental life in a single person. And the borderline cases in such cross-person cases are just like borderline single-person ones. For example, as discussed above, we may doubt that a severe amnesiac has a unified mental life, just as we can doubt if two people who are not completely I* related have a unified mental life. But the tests which we apply to try to answer those questions are the same. Does the amnesiac/later person remember earlier experiences as their own? Are the intentions of the amnesiac/earlier person causally efficacious as regards the actions of the amnesiac/later person? Is there a coherent internal narrative that allows the amnesiac/ later person to make sense of the past? Of course, sometimes these questions will have indeterminate or uncertain answers. But the uncertainty is the same in both cases. The borderline cases in our imagined examples are just like the borderline cases in ordinary, pathological cases.

[169] Wilkes [1993] has made the strongest case for the view that our accounts of persons must correspond to what actually is the case.

A third aspect of mental continuity which Parfit's account neglects is the need for both backward and forward looking connections.[170] Consider q-memory once more. If Jane q-remembers John's experiences, Parfit says Jane is phenomenologically connected to him. Relation R is simply a product of enough of these kinds of connections. I would not deny that under such circumstances Jane is psychologically connected with John, but no number of such similar connections could ever lead to the kind of psychological continuity Relation R should demand. The reason for this is that Jane will only be retrospectively connected with John. Her memories look back at John's experiences. But John never looks forward to Jane's future. He doesn't plan what Jane will do, nor does he anticipate having Jane's experiences. His plans, beliefs and desires will not effect Jane. This is different to the teletransportation case. Here, there is a pattern of planning, intention and memory. The beliefs, intentions and desires of the person on Earth will effect the person on Mars and the person on Mars can recall life back on Earth. In the Venetian memories case, Jane cannot look forward to the holiday in Venice. She can look forward to receiving Paul's memories on his return, but her desires, plans and wishes cannot effect what Paul does in Venice, just as Paul's desires, plans and wishes cannot effect what Jane does later.

Parfit does talk about other kinds of psychological connection, even though I have found his account wanting. My account, I hope, makes it clearer. The idea of the internal narrative is crucial to this. The internal narrative is not just a backward looking retrospective autobiography. The internal narrative is lived, day by day, and so is permanently being created by its author, the person. Sense cannot be made of this narrative unless one can look both forwards and backwards at whatever moment of the story one finds oneself.

[170] McInerney [1985 & 1991] first drew my attention to this failure in Parfit.

There are thus three respects in which I believe the I* relation is more precise than Relation R. Firstly, it accounts for the differences to psychological continuity pathological cases such as fission makes, because it does not explain direct psychological connectedness simply in terms of connections between particular thoughts and events, as if different kinds of thoughts all worked in the same way. Secondly, the I* relation corresponds more closely to what normally is involved in psychological continuity. And thirdly, the backward and forward elements in psychological continuity are more fully brought out by the I* relation.

One more point remains to be discussed. Parfit claimed that Relation R was what mattered in survival. Just what does this mean, and can the same claim be made of the I* relation?

5. “What Matters in Survival”.

To ask what matters in survival is a nebulous question. Interpreted one way, the question is very general. In the fission case, for example, what could matter to me is that after the split there will be two people instead of one whose lives are dedicated to the pursuit of an ideal I hold dear. In this case, what matters to me is served well by fission, even if one views fission as death. On this reading, “what matters” simply means “what is important for the particular person”. But this cannot be what Parfit means and it is not the conception I intend to use. There is clearly no way of explaining what is of value to persons in the future in the general terms required by a philosophical theory. Furthermore this question is clearly far removed from that of specifying the necessary and sufficient conditions for personal survival, which was the starting point, if not the end, of the present discussion.

We could interpret “what matters” in another way. We could say that what matters in survival is simply what is required if X is to be a survivor of Y. But this too is unsatisfactory, as it makes the question of what matters indistinguishable from the question of what is essential for personal survival. To talk of mattering would simply be to misleadingly add the suggestion of value to a straightforward enquiry into the necessary conditions of survival.

Neither of these two interpretations captures the sense of “what matters” in Parfit. What matters in survival is not just a question of what survival entails but nor is it a question purely of what we value. It is rather a question about what it is about survival which we value. It is a question about what is important to us, but it is specifically a question about what is important in survival itself. It seems to me that this is as far as Parfit really gets with the notion. His use of the expression clearly combines the idea of value with that of the requirements of survival, without actually explaining what this combination entails. In clarifying what is meant by “what matters”, therefore, it is possible that I will be moving on from Parfit’s own view. This is inevitable as there simply isn’t a sharp conception of this term in Parfit. However, I think an explanation is possible which is both sharp and which captures something essential to the Parfitian view which Parfit himself overlooks. I believe that the correct interpretation of the phrase “what matters in survival” is “what is required for survival from the first person point of view”. If this is the correct interpretation, it should be clear that in answering the question “what matters in survival” we would meet the relevance requirement. i.e. we would have addressed the first personal question of survival.[171] I shall now explain what my interpretation of “what matters” means and why I think it is the correct interpretation.

What is it that makes the question of personal survival different from that of, say, rabbit survival? Many of the puzzle cases used in the personal identity debate, such as those of fission and teletransportation, could be equally applied to rabbits. And even the thought experiments where the mind is effected could probably apply to dogs and ‘higher’ mammals. One factor which could make a difference is that persons, and perhaps certain other ‘higher’ animals, have a first-person perspective on the world. This perspective need not prevent us from approaching the question of survival in a completely third personal way. Even when talking about dolphins or pigs, the fact that these animals do have some sort of first-personal perspective is not in itself a reason to stop us approaching the issue in a third personal way. One can simply treat the organism as one other object in the world and attempt to specify the conditions of survival or identity. The same can be, and has been, done for persons. The result has been the search for the necessary and sufficient conditions of personal identity, what I

[171] See chapter two for a more detailed account of the Relevance Requirement.

have called the factual question of identity.[172] But as I explained in chapter two, there are questions about our survival for which an account of the factual question of identity cannot provide answers. One of these is whether we should have any special concern for the person who steps out of the teletransporter on Mars who is a replica of, but clearly not numerically identical to, me.

For this reason, I argued that there is a second question: the first-person question of survival. In cases such as teletransportation, are there reasons why I should consider the person who walks out of the booth on Mars as my survivor, who I should care about just as much as if he were me, even though we are not numerically identical? Here, the first-personal perspective I and my replica have seem to be in tension with the third-personal facts. There appears to be a clash between the way we view our survival from the first personal viewpoint and the way in which it is viewed from a third personal viewpoint. This is the most puzzling feature of teletransportation. Whilst we know that, from the first personal point of view, teletransportation would seem like a means of transport, from the third personal point of view it is destruction and replication of a person. Of course, our first personal judgments can be effected by our third personal ones and vice versa, but the gap between the two nonetheless remains.

On my view, however, the factual question of personal identity over time should be taken as quite distinct from the first person question of survival. The two questions are not in conflict, they are simply different. The possible area of disagreement is over the importance we give to each question, and even here, how important the questions are depends on what our interests are. If we are concerned with the issue of personal identity, period, then the first person question of survival is irrelevant. But if we are concerned with the importance of personal survival, then neither question can simply be ignored.

[172] See §2.1

On this point, I fundamentally disagree with Parfit. Parfit believes that teletransportation is a problem for identity theorists. We can see this by the use he makes of the concept of an “empty question”.[173] An empty question is a question where the answer (i) is not required to understand fully the situation referred to and (ii) cannot be said to be right or wrong. Parfit argues that some questions of identity are empty. For example, in the case of teletransportation we can know all the facts concerning the situation. He claims there is no further fact of identity to be discovered. There is no correct answer to the question, “Is the person on Mars identical with the person on Earth?” We thus have simply to choose between two descriptions of the events. We could say the person survives teletransportation or we could say a person is destroyed and replicated. These two descriptions do not describe two different states of affairs, for each of which only one of the descriptions can be right. Rather, they both describe the same situation and we need not decide which description is right. In order to decide which is the *better* description, we have to consider “what matters in survival”.

On my view Parfit is wrong. There is a fact of identity is this case. The replica is not identical with me. The question of identity is therefore not empty at all. But there is another set of considerations in this case. That is, the person who steps out of the teletransporter will be related in a certain way to the person who walks in. Namely, they will be I* related. On my view, this is enough for the Mars person to be a survivor of the Earth-person. Even though he is not identical with me, I have reason to care about my survivor as much as if he were me. In fact, I have more reason to care about him than I do about myself in the future, if I were not I* related to my future self, because of severe brain damage, for example. This is

[173] Parfit 213-214

because the I* relation states what is important to the first-person question of survival, whereas identity conditions only satisfy the factual question of identity.

Notice that I have not distinguished the factual question of identity and the first personal question of survival in terms of objectivity and subjectivity. The I* relation is, I believe, best described as an objective account of the first-personal conditions for survival. It is not ‘subjective’ because it is a set of criteria over which we can agree and which we can apply to persons other than ourselves. It is also not ‘subjective’ because we can use it to overrule the subjective judgments of persons concerning their identity. We can say that someone who claims to be the survivor of a dead person is deluded because there is no evidence to believe the two are I* related. We recognise that others too have this first personal perspective. We also acknowledge that any creature, or indeed automaton, that fulfils something like Locke’s criteria of personhood[174] also has this viewpoint. This means that when considering the conditions of survival for beings other than ourselves, we can use this understanding to allow for the special nature of the first personal viewpoint.

So the question of what matters in survival becomes for us now a choice: For a person to consider a future person as her survivor, someone she should plan for, save for, and whose interests she should have a special concern for, should that future person be numerically identical with her or is it enough that she be I* related to her? In short, is it identity or psychological connectedness and continuity of a certain sort which is the basis of our concern for future selves? The answer, I claim, is the latter one, and I shall defend this claim in the final chapter. Indeed, it is the very existence of this view which makes the question of personal survival distinct from animal, vegetable or mineral survival. The existence of the

[174] See section 1 of this chapte

first personal perspective and the fact that we occupy this perspective makes the question of our survival crucially different to the question of an animal's survival.

6. Conclusion.

As I have already stated, because my position is a development of Parfit's, it is not an argument from scratch. I have not offered strong arguments as to why it is psychological connectedness and continuity which matter in survival. What I have done is tried to improve on the account of what psychological connectedness and continuity is and explained what is meant by "what matters". What I hope to have done is to have improved on the Parfitian account which we had many reasons to accept, but with which we also had severe problems. In the final chapter I shall turn to an appraisal of the view proposed, in which I hope the attractions of the view become more apparent.

Chapter Seven

Appraisal

Parfit argued that what matters in survival is Relation R – psychological connectedness and continuity. I have argued that his Relation R is inadequate and have replaced it with what I have called the I* Relation. The I* Relation performs the same basic function as Relation R and it retains the distinctive features of the Parfitian account, whilst avoiding what I have argued are the main errors of Parfit. This revision forms the main part of my critical development of psychological reductionism. The I* relation will, of course, need to come under scrutiny itself, but it does seem to indicate a promising direction in which to develop Parfitianism. In this chapter, I shall begin this scrutiny. In particular, we need to ask if the I* relation is what is required to answer the first-person question of survival. There are also the requirements of chapters one and two to be considered. Firstly, does my account meet the relevance requirement, i.e. does it address the first personal question of survival? What we will find is that the questions, "Does the I* relation capture what matters in survival?" and "Does this revised Parfitian view meet the relevance requirement?" are very closely related.

In chapter one, as part of my examination of the *raison d'être* of philosophical anthropology, I formulated the Kierkegaardian requirement. i.e. the requirement to explain how it is we can view ourselves both as beings trapped in the here and now, and as beings whose existence extends backwards and forwards over time. We need to ask whether this requirement has been met. I consider this at the end of this chapter.

This thesis is an investigation into persons. It cannot be ended without an account of what persons are. This account has had to come at the end, and not at the start of the thesis, because my conception of what persons are follows from what has been argued concerning their survival over time. Now that the account of survival has been completed, what this entails for a conception of what persons are needs to be considered. This is done in the first section of this chapter.

Considering these questions is a means of appraising the view I have argued for. I have already tried to show how my view avoids the main errors of Parfit, but this in itself is not a demonstration of its worth. We also have to consider if the theory achieves what it is required to achieve. The aim of Part One was to make clear just what we can expect a philosophy of persons to achieve. I now need to show how the conception outlined here meets these expectations.

However, there are limitations on how thorough this appraisal can be. This thesis is a critical development of psychological reductionism and does not argue for such a position from first principles. Psychological reductionists such as Parfit have already discussed in great detail many arguments in favour of the view that psychological connectedness and continuity is the basis for personal survival. I see little value in going over these arguments again.

The chapter is divided up as follows. Firstly I consider the question, “What is a person?” and describe the answer I believe my position entails. In sections two and three, I defend the view that the I relation is what is required to answer the first-person question of survival against some apparent counter-examples. Sections four and five then turn to the relevance and Kierkegaardian requirements respectively and show how they been met by my account. Establishing these requirements and showing their importance is actually a lengthier task than seeing if they are met. By the time we see if they are met, my conception of persons should be sufficiently clear for the task to be quite straightforward. The importance of meeting these requirements is, however, great, as I hope the first two chapters have shown.

1. What is a Person?

The psychological reductionist view I have been developing entails a revision of what we mean by the term ‘person’. My preliminary definition was Locke’s: “A thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places”.[175] The problem with this definition is the use of terms such as “being” and “thing”. How can it be that a person is a thing if the conditions of the survival of a person do not include the survival of any substantial thing? Locke himself seems to recognise this difficulty and later adds a second definition:

> So that *self* is not determined by identity or diversity of substance, which it cannot be sure of, but only by identity of consciousness. Person, as I take it, is the name for this *self*.[176]

Here Locke tries to separate the identity of the person from the identity of a substantial being, locating the former in the identity of consciousness. What Locke calls “identity of consciousness” I would prefer to call “continuity of consciousness”, for I have argued that the relation which sustains a continuity of conscious is not necessarily an identity relation. The problem for Locke is, how can we individuate a consciousness over time if it is not any substance or entity? And if a consciousness is not a thing, then what is it? Can we even talk of *a* consciousness?

To raise such questions requires us to start thinking in terms of individuable consciousnesses. But what we should not be tempted to do to resolve this

[175] Bk II, Chapter XXVII, Para.9

[176] Bk II, Chapter XXVII, Para. 25/26

dilemma is to attempt to reify consciousness in some way. One of Locke's central points, reaffirmed by his successors and in this thesis, is that personal survival does not depend upon the survival of any *thing* at all. Hence to try and make a thing out of consciousness would simply be to fall back into the error Locke revealed. Also we do not want to talk of persons as ephemeral "consciousnesses" as this would be to erect a new mystery as puzzling as any of the dilemmas surrounding persons which we may have dissolved thus far.

Psychological reductionists wanted to be able to claim that personal survival does not require the survival of a substantial thing but that is not to say that persons are nothings or ephemeral somethings. One way of making these two claims compatible is to hold that persons are particular beings but personal survival does not require survival of that particular being. At first sight, this may seem an absurd proposition. It seems to ignore the principle that if any F is an X, then for F to survive it must remain an X. For example, if the *Mona Lisa* is a painting, for it to survive it must remain a painting. However, if one is to accept that a person is a particular living being, but that the person's survival does not require this living being's survival, something along these lines would have to be accepted.

However, I believe there is a better solution. This relies upon the specific notion of survival described in chapter two. Survival may not require a token-identical continuer. A type identical whole or part continuer may be enough. There are at least two reasons why this conception of survival offers a better solution.

Firstly, in chapter two, I considered a possible counter-argument against psychological reductionism:

(1) Persons are of type X. (e.g. human beings)

(2) Relation R can survive transformation of type X to type Y

□ (3) Relation R can survive transformations which human beings cannot.

□ (4) Personal identity cannot consist in Relation R.[177]

But, I argued, to say a person survives a certain transformation may simply mean that there is a type identity of certain vital parts or features of that person before and after the transformation, rather than that there is strictly survival of the whole person. This is enough to neutralise the counter-argument. Although a person can be a human being, it is not the identity of the human being which is required for personal survival. This is because it is the consciousness of the human being which is of relevance to the first-person question of survival. It is argued by psychological reductionists that consciousness can continue across different token individuals, as would be the case in teletransportation. In other words, type-identical parts or features of persons are enough to sustain continuity of consciousness and hence facilitate personal survival.[178]

In this way we can see how it is possible that, although persons are beings, survival of the person does not require survival of that being. In my view, what Locke did was to demonstrate that this was true, without showing how it could be true. In this thesis, I hope to have offered some reasons as to how it is true. A second reason comes from the Kantian considerations of this thesis. To sum these up, a unified mental life requires the formal identity of the first-personal subject, not the actual identity of a physical subject. Because when we consider our survival from the first-person point of view it is this unified mental life that we are concerned with, it is this formal identity which determines personal survival. If this is the case, it is quite clear that personal survival does not require survival of the particular being. But that is not to deny that a person is at any one time a particular being.

[177] §2.1

[178] This possibility was argued for in more detail in §2.1

Interestingly, although Parfit calls himself a constitutive reductionist, he still prefers to say that a person *has* a body and a brain rather than *is* a body and a brain.[179] Personally, as long as "is" is understood in the constitutive sense, I can't see how Parfit's description can be better, particularly as the verb 'to have' implies there is something which is doing the having, which simply begs the question once more as to what these persons are which have bodies.

The advantage of this conception of survival is that it accepts what many animalists insist upon, namely that if a person's body is destroyed, that person is also destroyed. But as I have explained, the I* relation is a relation that holds between persons-at-a-time. If I am teletransported then I, this particular person, ceases to exist. A replica is built in my place. But there is a conception of survival on which the person on Mars is my survivor. The first-person question of survival is, "Should I, now, consider person X at a later time to be my survivor, who I have reason to care about just as much if he were me, regardless of whether we are numerically identical?" If I am I* related to that person, then the answer is yes. And that entitles me to call that person my survivor.

Locke's first definition of a person is therefore still quite acceptable. What we have done is understood a little more about how a person can be a "thinking being" without the survival of the person consisting in identity of this being. This is certainly a significant change to the traditional psychological reductionist position. But what it does is enable us to admit that after teletransportation, for example, there is no longer the same person, whilst retaining the fundamental psychological reductionist claim that psychological continuity and connectedness is what is important to my survival.

2. Why Is The First Person Point of View Important?

[179] Parfit reaffirmed this point in his notes to the 1993 Jacobsen lecture

I have claimed that the first personal point of view is the important one when we consider personal survival. Over the next few sections, I shall subject this view to scrutiny by considering potential counter evidence. I start by looking at William James' work on the self. James' work on personal identity in the *Principles of Psychology* is usually divided by critics into his interesting discussion on Hume and Kant, and his more extraordinary ideas such as those about the importance of clothes for personal identity and the role of the glottis in assenting and negating. But it is within the more unusual sections that I think we find some of the most interesting ideas. We know James holds to an idiosyncratic view the moment he describes not one, but four different selves: the material, social and spiritual selves plus the pure ego. Although all these "selves" are interesting, I am only concerned with one: the social self.[180]

For James, the social self is itself a multifarious thing. Many, if not most, people behave, appear talk and even feel differently depending on whom they are talking to and where they are. It would seem wrong to attribute all of these differences to putting on an act and to say that only one of these represents the 'real self'. What would be more accurate would be to say that we display different facets of ourselves on different occasions. Think of the stereotypical mafia man, who is all love and kindness to his family and a cold killer at work. It seems an over-simplification to say that this man's true self is only manifest in one half of this double life.

In his discussion of the social self in particular, James gets us thinking about the importance of factors other than a unified mental life, even though he, rightly it seems, does not believe the social self determines personal identity. We may concede that our social and material selves are important factors in what makes

[180] James [1890], pp293-296.

us who we are, but we would not immediately consider discontinuity of these selves to threaten our identity over time.

I would now like to bring this together with my argument that Parfit's claim about what matters in survival is best interpreted as meaning what is required for survival from the first person point of view. The question suggested by considerations such as those of James is, why should it be the first person point of view which settles the question of survival? Considerations of James can help us to push this question harder. Imagine that all the social interactions you currently engage in come to an end, and that these are very important to you. Your family no longer treat you like a family member, you no longer enjoy the position you have at work, if you are famous, you are no longer sought after or recognised. This is the death of what James calls the social self. Loss of the social self is in effect society treating you as if you have died and that someone else now occupies your body. Given that we are social animals, and despite the continuity of consciousness, I would say that for many people, such a fate would be seen, maybe not immediately, as death, in as strong a sense as global amnesia may be for many other people. Thus being treated as a different person could make the person feel that they were a different person from their first personal point of view too. This raises two questions. Firstly, in such a case, is it not the case that it is the third personal, and not the first personal view of survival which is important? Secondly, is not the first personal judgment informed by third personal judgments, so that to analyse personal survival purely in the first personal terms as I have done is mistaken?

In this case, the I* relation is still very important. To be able to think that one is not a survivor in such a case actually requires the I* relation. In order to feel the break with the past one actually has to have a fundamental awareness of the continuity with the past, or the break would not be apparent. The I* relation therefore provides the fundamental unity against which a judgment of disunity can take place. In our actual society, such a person would not be considered as having died. Legally, socially and morally, we would consider that the person has survived the trauma altered rather than has altered so as not to have survived. But it is possible to imagine a society which places more importance on social position. In this case it is possible that both the subject and the society in general could view the person as having died.[181] Once again, it is important that the I* relation must be underpinning all these changes, because it is only because there is a unified mental life linking the person before and after the loss of the social self that one can conceive of a change in social position being equivalent to a change of person. My first point is therefore that the crucial factor in survival is the I* relation, even in those societies which decide to view other changes as resulting in a change of person. Such societies can only form these conventions if the more basic personal unity provided by the I* relation is already in place, at least for the vast majority of its members.

But this is not yet enough. It could be objected that in such a society, the I* relation would simply be one necessary condition of personal survival, but not a sufficient one. The reason for this is quite simple. In such a society the criteria for personal survival are third personal, not first personal. Anyone who grew up in such a society would therefore apply the third personal criteria to themselves and would not take the continuity of their own mental life as evidence for personal survival. What this means is that the claim that it is survival from the first person point of view which is important could be a culturally relative one. And yet, it is supposed to be a claim about persons in all possible worlds. The I* relation would thus seem to fall short of its goal.

[181] I understand this is the case in certain Melanesian societies where one’s identity is taken to change when one’s role in society changes.

However, I believe this failure in the I* relation is only an apparent one. We need to answer the question, “In a society where the social self determines personal survival, are there persons, no persons, or are persons something different?” Consider the first possibility, that there are persons in that society, beings who fulfil Locke’s criteria of personhood, just as there are persons in our society. That there are such persons in such a society seems to me to be indubitable. The day before someone takes on a new role, for example, there is nothing to stop that person looking forward to that new role and making plans for it. Similarly, when in that new role, memories of and connections to life before are not lost. In this sense, there is still the unity of personal life that we have in our society. Consider also people who join cults and gain a new identity. Just because that community treats its members, and its members treat themselves, as new persons, doesn’t mean that there is not, before and after the conversion, what we would call one person.

So, despite the differences, there seems no reason to conclude that there are not persons, in our sense of the word, in that society. The third possibility is that there are persons, but that this means something different. In some sense, this must be true. But just so long as there are persons in our sense of the word – i.e. beings who meet something like Locke’s conditions of personhood – whether or not there are also persons in some other sense is irrelevant. What I have said about personal survival applies in all cases to what we call persons. It can hardly be a counter argument to this to claim that in some cultures this concept of a person doesn’t operate. A person is a person whether they recognise that description of themselves or not.

There is a dilemma facing any claim that a concept is culturally relative. To establish that a concept is culturally relative, it is not enough to show that the concept is applied differently or has a different meaning, as this really boils down to the claim that it is just not the same concept. The claim must be a stronger one, that the concept has no application in that society. If it could be shown that what I have called a person is not a concept with any application in some other societies, then it would be a genuinely culturally relative term. But I can’t see how such a claim could be made. I shall further argue for this view in the next section.

Adapting Parfit, I have claimed that it is the first personal point of view which is the important one. I still believe this is true. Without the I* relation, there cannot be persons such as the persons we are. As I have argued, without the I* relation, alternative conceptions of personal identity that preserve a recognisable sense of “person” are not possible. We need the underlying unity of the I* relation. It is in this sense that the first-personal view of survival is the important one. Without it, we could not even think of plausible alternatives. Expressed in another way, this means that we must have persons, in the sense of the word I described in section 1, before we can have other, culturally relative, variants on the concept of personhood.[182]

So it is not true that the possible importance the social self could have in other societies supports third-personal rather than first-personal views of personal survival. What of my second worry, that it suggests third-personal views influence our first personal judgments more than my account allows? It is a truism that beliefs we have about our culture and about the attitudes of others effect the way we think about ourselves. The beliefs we have about ourselves can clearly be influenced very strongly by the societies we live in. But my account is not about the beliefs we have about ourselves, but the way we are, as persons. My point in this section has been to show that we are persons, in the sense I have described,

[182] We even have variants on the term person in our own society, such as the legal concept of a person. However, nobody would believe that these variants threaten our core conception of what a person is.

whatever other beliefs we may come to form about ourselves. Therefore, the effect of these beliefs is not central to the enterprise of this thesis.

To strengthen this argument, in the next section, I consider some of the views about persons held in various cultures that, *prima facie*, most clearly conflict with my views. I will then show how these differences do not in fact threaten my own position.

3. The Self and Society.

I have argued that the fundamental unity provided by the I* relation can underlie different views of what persons are, but that in each such case, persons, in the sense I have defined, must exist in order for ‘persons’ in some other sense to exist. If this is a case, we should expect to find societies somewhere in the world where there are significant variations on the concept of a person, but where nonetheless we can still apply the conception I have described. Anthropology shows us that this is in fact true. I am cautious about using anthropological examples as I am cautious about using data from any discipline of which I am largely ignorant. But I do not believe that any of the examples I use rely on anything other than accurate ethnographies and data. Indeed, in one case, the example is familiar to us all. Also, I am more interested in the alternative possibilities suggested by these examples than I am with the actual societies themselves. I shall argue that the cases I discuss which may seem to undermine my claims about persons are actually entirely compatible with them. I consider three such views, which I have called the functional self, the manifold self and the human self.

(i) The Functional Self.

Rom Harré considers cultures where different linguistic use of personal pronouns points to a different conception of the self. I will look at two of these.

The first is the culture of traditional Eskimos. In their language, Inuit, “personal reference is accomplished with only two suffixes, ‘-ik’ and ‘-tok’, the former indexical of the speaker, the latter referring to any other person or group of persons”.[183] Harré argues that this suggests the Eskimo sense of self is weaker than it is for us, who have many ways of self-reference. Ethnographic evidence supports this view:

> Eskimo emotional states appear to be much more socially dependent than ours. Isolated Eskimos, in so far as they can be observed, seem to be stolid, neither cheerful nor depressed. But once they become part of a community (a family, say) they quickly take on the emotional tone of the community, whether they are intimately bound up with its concerns or not. [...] It is said that all moral issues are referable only to relationships of the individual within a family group. The active, decision-making unit is not the individual human being, but the family.[184]

It is clear that this is not just a matter of different conceptions. The individual, it is supposed, *feels* differently about his or herself than a European does. Self consideration becomes a very much rarer thing. A person is therefore defined much more by their social situation than their sphere of consciousness. It could be argued that in such a society, a lack of a unified mental life would not present a problem for personal identity. As long as the individual continued to fill their space in the family and community, there could be no problem of personal identity. If this is how Eskimos view each other, it would be hard to imagine why they would view themselves any differently. In the various puzzle cases discussed by philosophers, so the argument would run, for the Eskimo to be able to answer they would have to know if the resulting person continued to fulfil their social role,

[183] Harré, p107

[184] ibid

not whether they psychologically connected and/or continuous with the earlier person.

I dub this conception the “Functional Self” because it is the view that *a* is the same person as *b* iff *a* performs the same social function as *b*. This conception is worth considering whether or not it is the correct interpretation of the Eskimo view of the self.

I believe this conception is not tenable. What reasons are there for supposing that in such a society, someone who was purely functionally continuous with an earlier person would be considered to be the same as that person? Perhaps it is because, as long as this were the case, everyone would consider it to be the same person. But we need to push this further. There are two possibilities. Either everyone knows that this is not a numerically identical human being or they don’t. If they don’t, then we can infer little from their beliefs. Consider Searle’s Chinese room argument[185]. Here, someone in a room passes out correct Chinese responses to Chinese questions, but without knowing a word of Chinese, thanks to a very helpful manual. Despite this making her indistinguishable from a Chinese speaker, Searle argues that anyone who knew what is going on would not think that the person in the room could speak Chinese. In our case, as long as the later human being responded in the correct way, they would be indistinguishable from the earlier human being whose life they were ostensibly continuing. But this does not show either that the later human actually is a continuer of the original person or that her contemporaries would recognise her as so, if they were in possession of all the facts.

What, then, if everyone knew that this human being was not numerically identical with the earlier one? Again, there are two possibilities. It could be the case that nobody minds because the important thing for this society is that social roles are performed. But this would only show that what is important is not the survival of the individual rather than that the I* relation is not what is required for the survival of the individual. There would be no conflict between this and my view. The remaining possibility is the one which seems truly to exemplify the functional self: everyone is aware that the later human is not numerically identical with the first but nobody minds because survival of the individual just is continuation of someone who performs the same social role. Here, conditions of personal survival would be third personal, not first personal, and functional, rather than a matter of psychological connectedness and continuity. Could such a conception ever be applied?

When we consider what such a society would be like, it becomes increasingly difficult to make sense of it as being a society of persons at all. There is a further fork in the thought experiment to consider. On the one hand, this could be a society without I* related persons at all. There are no beings who have a forward and backward looking relationship with themselves in the way described earlier.[186] This would be a society, not only where there was a different view of personal survival, but with a different sort of inhabitant altogether. If we try to think ourselves into such a society, we find that doing so is virtually impossible. Beings without the first personal sense of their past and futures are simply too alien for us to be considered persons at all. In order for the thought experiment to have any bite, therefore, we need to imagine that, though these beings are I* related to the future and the past, they simply don't use this as their criterion for survival, but use functional considerations instead. This seems to me to be deeply implausible. How could beings with a sense of their own pasts and futures, with the powers of memory and intention, divorce all these factors from their

[185] Searle [1980]

[186] Chapter six, especially §6.5

conception of personal survival and replace it with a functional view which is entirely third-personal? For example, they would seem at least to need two different personal pronouns: one “I” which is the I of a unified mental life, and one functional “I”. It would thus be possible for someone, in the functional sense, to do things which, in the other sense, they would have no possibility of remembering. Users would certainly have to be aware of the differences. And this would mean that, in effect, they were employing two parallel conceptions of personal continuity over time. This, once more, would collapse into the view that we had a society where personal survival is determined by I* relatedness, but that importance is placed on something other than personal survival, which is based on functionality.

The dilemma is this. If we don’t allow the beings in this thought experiment to have I* relations, then they become so alien that they cannot be considered as persons at all, and thus it ceases to become a counter-example to my conception of persons. But if we do allow them these relations, the idea that they could take a *purely* functional view of personal identity just becomes implausible. What happens is that if we try to imagine what it would be like to be in such a society, we find that we already presuppose the conception of persons I have argued for. In order to be able to think of identity in functional terms, one already requires the characteristics of persons I have described. At best, they have two different views of personal survival, which still means my account is applicable to that society.[187] The characterisation of the Eskimo view of personal identity being functional must then be false, and the idea that the functional self can exist without the I* related self is unconvincing. The Eskimo view of the self is therefore a variant which operates within the limits laid down by the I* relation. Certainly, the group may be more important for them and may be crucial in making a person what they are.

[187] As, I have argued, we too have two views: one based on the third person conception of identity, the other on the first person conception of survival.

But if we are to suppose that this entails a breakdown of the sense of a person being a thinking being with a sense of its own future and past, we are surely mistaken. In short, when we really consider what the functional self would entail, it only becomes plausible in beings rather different to persons.

(ii) The Manifold Self.

Harré's other example comes from Japan. In Japanese, the way one refers to oneself and others always depends upon who one is with. It is impossible to speak to anyone without the use of vocabulary indicating the social distance between you and them. This is part of a culture where much depends upon its social place. In morality, for example, what is important in one area of life may not matter in another:

> The shame that diffuses that part of the psyche which has to do with others in the world of work cannot diffuse into the system within which a Japanese manages his home life.[188]

The self, therefore, is not immutable and absolute but is something which is defined relationally. Again, we would therefore expect that personal continuity depends upon the maintaining of these relations, not on psychological unity, except in the derivative sense in which psychological disunity could interfere with these relations.

What we seem to have here is a case where different sections of society place value only on those parts of the lives of persons which are relevant to those sections. This is really only a more extreme form of what occurs in any society where a person has a variety of roles. In any one day, a person can be, for example, a parent, a partner, a boss, an employee, a team member and so on. The fact that the person acts and is treated differently in each of these cases doesn't threaten the fundamental unity which underlies those diverse roles. It could be argued that without an awareness of the different roles we play, we would be unable to fulfil them properly.

There is, however, another view not dissimilar to the Japanese one, which believes there is an increasing fracturing of the self, which some see as part of the postmodern condition. The postmodernist claim is that we are witnessing a fracturing of our traditional views of the subject as a unified entity. As part of a broader fracturing of thought and society, we will come to see persons, not as single uniform entities, but as beings without any essential unity whatsoever. Each person will have a variety of personae, and each of these can be seen as much as individual entities as the whole human being. We will consequently have to revise our views concerning the fundamental unity of the individual. This can be seen as a version of the view that for the Japanese it is incorrect to think of a single individual with various social facets but rather a collection of different persons which occupy different stages of the human being's life. I call this conception "The Manifold Self" as it denies the psychological unity entailed by my own view and instead postulates the existence of a variety of selves within the single organism.

There are two points I would like to make about such a conception. Firstly, such a radical fracturing of the individual appears extremely unlikely. It would seem to make the unity we currently enjoy a kind of social construct. Certainly, the way in which society is organised could make us divide up our lives into more discrete compartments, but the unity which we experience is surely not something that can be swept away so easily. Even if we did make greater distinctions between our various public and private faces, these would themselves be unified by the thread of the I* relation. It is inconceivable that, for example, the bank

[188] Harré, p109

manager would not be able to plan what they would do that evening after work or remember what they had done before, for example.

But secondly, even if the postmodernists are right, this doesn't threaten my view. However fractured the individual human being is, in order for there to be any unity among the various characters, or whatever they are to be called, they must be I* related. What I suppose we are being asked to accept is the possibility of a number of I* related 'threads' running along the life of the human being, with each one taking it in turns to sit out while one continues its development. Within the single human life, each person-at-a-time would only be I* related to certain other persons-at-a-time. This is certainly compatible with my view, even if it does sound more like a wild thought experiment than an actual possibility.

So it seems that either way, the postmodernist view of the manifold self is no threat. Either there is a more fundamental unity which links together the distinct personae of the human being or the various personae themselves have their own internal unity, which can be fully explained by the I* relation. The fracturing of the self has to have limits if the result is to be persons at all.

(ii) The Human Self.

Finally, I would like to offer an example from our own society. When someone's relative suffers terrible brain damage, so that they are almost certainly not I* related to their former self, one does not always consider that person to have died. Whatever happens, we are still the children of our parents, the siblings of our siblings. We can view this sentimentally or as the cold result of our genetic programming. But whatever the cause, it happens, and these relations do help define who we are. Even stronger than this, it could be claimed that this shows there is something immutable about our identity. We cannot entirely cease to be what we once were, no matter what changes occur. Such a view would support the animalist view of identity, as it seems that it is the human being in this case which is always considered to be the same person. It could be claimed that if we reduce personal identity merely to psychological unity, something which we value about ourselves has been left out. And all through this thesis I have been arguing that what we value about ourselves has to be included in the final analysis.

Of course, this is most apparent when we consider how we view others. When we think about ourselves going senile, for example, we tend to take the view that ‘it won’t be me there’ quite easily. That’s not such an easy position to take up with someone else. Many people say of senile people that the person they once knew is already dead. But if deeds speak louder than words, then we do not view the ending of a loved one’s unified mental life as the end of that person, merely the end of that person as we knew them. We still feel for them when they suffer and grieve for them when they die.

What is also apparent is that we cannot neatly distinguish the way we view ourselves and the way we are viewed. Our culture at least partly defines what we consider ourselves to be. It seems impossible for anyone to hold a vastly different notion of what it means for other people to be individual persons to the one they hold for themselves.

Cases like this demonstrate the difficulties of theory meeting practice. I believe that if we review the arguments, we will find psychological reductionism to be superior to animalism. All we can learn from cases of amnesia and terrible brain damage is that certain intuitive responses cannot be overruled by philosophical reflection. Hume found that, however bothered he was when he thought about the existence of real bodies in his philosophical mode, playing billiards his doubts soon melted away. Equally, it seems to me that however much we are convinced by psychological reductionism, our being accustomed to thinking of persons as individual human beings makes a complete transfer of theory into practice almost impossible. Parfit also candidly admits the same thing:

> I can believe this view at the intellectual or reflective level. I am convinced by the arguments in favour of this view. But I think it likely that, at some other level, I shall always have doubts. [...] Something similar is true when I look through a window at the top of a sky scraper. I know I am in no danger. But, looking down from this dizzying height, I am afraid.[189]

As human persons, it seems our humanity is important to us. This may be hard to defend rationally. Indeed, it could be characterised as an example of a pernicious speciesism[190]. Alternatively, our attachments to human beings regardless of what has happened to their minds may be sentimental, explained by the fact that we cannot but associate the unified mental life we cherish with the body which constituted or realised it. Whatever the explanation, to be consistent the psychological reductionist must argue that such body-based concerns are misplaced. The difficulty of putting this belief into practice is not an argument against it truth.

[189] Parfit p279

[190] As defined by Singer [1979], chapter 3. Speceisism is to deny equality to beings solely on the grounds of their specie.

4. The Relevance Requirement.

The relevance requirement is the requirement for our account to address the first-personal question of survival. As this question is partly concerned with our interests and concerns in persons as persons ourselves, there can be no acid test for whether or not a theory meets the relevance requirement. We can ask, “does this really capture what is required for my survival, from my point of view?” and “does this theory leave anything important out?” but at the end of the day, such questions are not definitive and their answers are more judgmental than factual.

The relevance requirement has been used throughout this thesis in the development of the final conception. That the requirement has been met should therefore be clear from the way in which the whole thesis has developed in the light of the requirement. Throughout this thesis I have endeavoured to stray as little as possible from our fundamental concerns in persons. Crucial in formulating The Kierkegaardian Requirement in chapter one were considerations of how we view ourselves and how these views affect the way in which we live our lives. Kierkegaard stresses the existential as well as philosophical importance of self-understanding, which is a good deal of what philosophical anthropology is. The motivation for fulfilling the Kierkegaardian requirement is thus not the simple motivation for solving a philosophically vexing problem, but is the motivation to understand ourselves so that we may know better what we are. Thus at the very start of my thesis, the need to address the concerns of actual persons is given a central role.

Chapter three could well appear to be a fairly dry piece of methodological and terminological analysis. But even here, the imperative of the relevance requirement was never far away. Part of my defence of the term ‘person’ was that this term captures what we value about ourselves and our survival in a way in which more precise but more limited terms such as ‘homo sapiens’ do not. I also defended thought experiments by showing how, even if impossible, they help us to understand the relative importance to our survival of uniqueness, and bodily and mental continuity.

Hence in Part One, I attempted to found my enquiry on pillars which meet the relevance requirement. The requirements I set for a philosophy of persons and the methods and concepts employed in such an enquiry were all justified at least partly on their meeting the relevance requirement. In Part Two, I turned to the Parfitian conception. These chapters were concerned with the details of Parfit's reductionism, and how it could be revised to rid it of its more problematic features. Parfit's position is notable for its stress on what matters in survival. As I have explained, the idea of considering "what matters" is to enable us to adopt the perspective on personal survival which is most important to us: the first-personal perspective. Because of this, we approach the question of personal survival from the point of view which is most relevant to our own concerns. So although the relevance requirement slipped into the background in part two, it was there, forming part of the reason for our interest in Parfit's position in the first place. As I have said, it would be too lengthy a task to argue for Parfit's basic position here and it would require much raking over old ground.

The foundations of my enquiry meet the relevance requirement. The relevance requirement is also apparent in the Parfitian position I have developed. In this and the previous chapter, the question of relevance has also come to the fore. Particularly, I have shown why it is that the first-personal perspective on survival is so important. It is important because as persons, it is the perspective we occupy on our own lives, and because it is the existence of this perspective which makes personal survival different from other questions of identity over time. Without the first personal perspective, we would not be persons, as trying to imagine a society where personal identity was decided on a merely functional basis showed. These reasons are also reasons why the approach adopted in this thesis keeps the issue of persons relevant to our first-personal concerns, addressing our concerns not just as philosophers but as persons.

5. The Kierkegaardian Requirement.

In chapter one I formulated what I called the Kierkegaardian requirement. This is the requirement for any philosophical description of persons to be able to account for both the aesthetic and ethical character of persons. As regards her aesthetic character, a person can be seen as a succession of moments, a being who is tied to the present with no existence beyond the now. As regards her ethical character, a person can view themselves and be viewed as a single subject with a continuous existence that stretches backwards and forwards in time. It is not obvious that these two ways of viewing the self are incompatible, but there is certainly a tension between them. Can the psychological reductionist account I have developed resolve this tension? I believe that it can. To show how, I will try to say a little about how my account can explain both the aesthetic and ethical views of the self, and about how the two can happily co-exist.

(i) Aesthetic Persons.

The fundamental characteristic of the aesthetic person is that they are inescapably bound to the moment. The psychological reality of this is not something that requires any demonstration. The fact that we are only conscious in the here and now, the “specious present”, is well known and is the cause of puzzles of its own. The fact that we are nothing more than individual human beings, as finite and bound to the present as any other animal, is also confirmed by the current thesis, but is, again, not a novel or illuminating explanation of our aesthetic character. What we need to consider here is rather the ways in which my conception of persons reinforces the sense in which we are bound to the moment.

Consider first ways in which we can understand something as *not* bound to the moment. The easiest way of doing this is to consider how a being or object which perdures or endures exists at different points of time. If X at T^1 = Y at T^2 then X exists not only in the here and now but at various points of time. There are two ways in which this is more complicated in the case of persons. Firstly, on my account, there is no *thing* the endurance or perdurance of which is required for personal survival. A person's existence requires a brain and a body, but a person's survival does not depend on the continued existence of either of these particular things. At any particular time, a person is a whole human being, but the person's survival does not necessarily require the continued existence of that human being. As a person's survival is not a matter of the survival of a body, this makes a person an aesthetic being, as there is no *thing* the continued existence of which constitutes personal survival. Personal survival requires certain psychological relations between persons-at-a-time, not identity between persons-at-a-time.

A person-at-a-time is, however, a being. As I argued in section 1, the force of Locke's point is that though a person is a living being, survival of the person is not survival of that being. There is a sense then in which a person is not what they will necessarily become. That I am this particular human being now does not entail that I will survive as this particular human being later. The continued existence of the human being is just not a part of what my survival consists in. As Kant argued, the unity of our mental lives over time entails no identity of substance. We can trace the identity over time of any living thing and by doing so trace out a four-dimensional being not entirely bound to the moment. We could do the same with persons. But because survival does not consist in identity, the four-dimensional being we are tracing does not necessarily support the unified mental life which constitutes our survival. The fact of personal identity over time is thus not of importance to our first-person considerations of survival. Therefore, facts of identity cannot be invoked to relieve our sense of being bound to the moment, as facts of identity are not relevant to the idea of our survival.

So unlike other objects or beings, our survival does not entail identity at all. I have argued that the relation which unifies the life of a person over time, the I* relation, is not a relation of identity. This is one of the features of Parfitianism which my account has retained. This means if X at T^1 is I* related to Y at T^2, although Y would be seen as a survivor of X we cannot deduce that Y is X. There is always the logical possibility that fission has occurred, for example. There is nothing in the I* relation itself which rules out such possibilities. In this way, though a person can think of themselves and be thought of as the same person over time, this does not entail that there is actual personal identity over time, as identity is a one-one relation and is not entailed by the I* relation. Judgments of personal identity therefore have to be third-personal and empirical, based on the continued existence of the human being. So the fact that I have survived entails nothing about the continued existence of any being. This again shows how one cannot even consider the continued existence of the physical human being as evidence for our transcending of the moment, as it is simply not part of what personal survival entails.

What we can see here is that the aesthetic view of the self is tied up with the negative aspects of my conception: that survival is not survival of any substance and that survival does not entail identity. Both these factors together suggest that, even though beings exist beyond the here and now because they endure or perdure, this fact is not relevant to human survival. Therefore, it is not a fact that can remove our sense of being bound to the moment. Rather, because personal survival does not require identity of a being, the sense in which our existence is dependent on the being-at-a-time, the human being in the here and now, is accentuated. For Kierkegaard, this was the cause of the despair which is characteristic of the aesthetic. But when we consider how my conception fits in with the ethical view, we shall see that this despair is not inevitable.

(ii) Ethical Persons.

If the aesthetic character of persons is reinforced by the negative aspects of my conception, then the ethical characteristics emerge from the positive points. An ethical person is not bound to the moment, but has a continuous existence over time. It would appear that if my account makes persons aesthetic then they cannot also be ethical. At the end of chapter one, I made it clear that making the ethical and the aesthetic compatible with each other was only one possibility. We could also explain the appearance of one in terms of the other. This is what I shall now do.

The appearance of the enduring self has already been explained in Kant's paralogism. Despite the fact that 'I' entails no substantial identity, the formal identity this entails leads us to think of of ourselves as substantially identical over time. We could then say that the ethical view we have of ourselves as single entities existing over time is an illusion. What I have made clear is that this illusion of substantial identity is not anything we should feel cheated by. I have argued that our survival just doesn't entail identity, and if this is the case, we should accept Kant's point with ease. As A.J.Ayer once put it, "it is perverse to see tragedy in what could not conceivably be otherwise."[191] What I believe is the correct view of what our survival consists in preserves all that was important when we thought of ourselves as substantially identical over time. The point of the relevance requirement was to make sure that our real interest in our survival, the first person question of identity, is addressed. Having addressed this, in the process we have found that in fact, the first-person question of identity can be answered without the concept of identity. Hence the important issue of our survival just doesn't involve the substantial identity of the self we perhaps thought it did.

The I* relation is thus the way in which we transcend the moment. By being capable of thinking of ourselves as ourselves over time we are able to live our lives outside of the present. Strictly speaking, we could call this an illusion. But to call it an illusion has certain negative connotations. It suggests we have been fooled, that there is no reality in what we perceive. But this needn't be the case. The I* relation is real and we needn't be fooled into thinking that it provides us with identity. Once we accept that we are beings for whom identity has no importance in survival, we can transcend the moment by thinking of survival in terms other than those of identity.

This is how I believe the Kierkegaardian tension can be resolved. The ethical self is in a sense an illusion, because there is no being whose endurance or perdurance is required for the survival of a person. The illusion of this is explained by Kant's third paralogism. What I hope I have explained throughout this thesis is that when we consider what it means to survive and what it is that gives unity to our lives, we can see that these do not require an enduring or perduring self.

Kierkegaard wrote, "finitude's despair is to lack infinitude."[192] I believe this despair can be dispelled by a proper understanding of what survival entails. As what matters is not the persistence of any entity, the lack of infinitude in this sense is irrelevant. What does matter is continuity and unity of mental life. This can and does transcend the moment and enables our lives to be lived in more

[191] A.J.Ayer [1956],*The Problem of Knowledge*, London: Penguin. p41

[192] Kierkegaard, [1849], p60 & 63. See §1.4. for an explanation of these expressions.

than just the moment. So in a sense, Kierkegaard was wrong. The aesthetic person is not “ensnared by immediacy”[193]. Rather, she can escape immediacy by the ethical character of the I* relation.

The Kierkegaardian requirement had force because it demanded that in explaining what persons are, we must not neglect either the ethical or the aesthetic. My conception of persons and their survival over time pays heed to both the ethical and the aesthetic. It brings together the lack of permanence in the self and the unity over time in the self. Because of this, I believe it truly meets the Kierkegaardian requirement.

6. Conclusion.

There are four main conclusions of this thesis. (1) There are two distinct questions in philosophical anthropology. One is the factual question of identity, the other is the first person question of survival. (2) Survival in the latter question should be understood as psychological connectedness and continuity. Parfit argued that psychological connectedness and continuity does not imply identity and is what matters in survival. Parfit’s account is flawed in several respects. His conception of “what matters” is incomplete, his account of thoughts as independent is mistaken and his account of psychological connectedness and continuity is inadequate. However, I have argued that (3) the flaws can be removed from Parfit’s account whilst retaining the key Parfitian claims that psychological connectedness and continuity is what is required for survival, and not identity. The I* relation is an attempt to show how this could be done. I have also argued that (4) this conception of persons meets two requirements that any philosophy of persons should meet, what I have called the relevance requirement and the Kierkegaardian requirement.

[193] Hannay, p5

The project of the thesis can be seen as a compatabalist one. I have tried to reconcile the animalist’s claim that a person is a particular being with the psychological reductionist’s claim that a person can survive the destruction of her body. I have also tried to reconcile the aesthetic and ethical views of persons. How successful these attempts have been I must leave to others to judge. But I hope that at the very least, potentially fruitful directions of research have been suggested and started on.

BIBLIOGRAPHY

BELL, David [1991].*Husserl,* New York: Routledge

BENNETT, Johnathan [1966]. *Kant's Analytic* (Chapter 8), Cambridge: Cambridge University Press.

BENNETT, Johnathan [1974]. *Kant's Dialectic* (Chapters 4-6), Cambridge: Cambridge University Press.

BRAUDE, Stephen [1991]. *First Person Plural*, London: Routledge.

BRENNAN, Andrew [1987]. “Discontinuity and Identity,” *Noûs*, v.21, p241-260.

BROOK, J.A [1975]. “Imagination, Possibility and Personal Identity,” *American Philosophical Quarterly*, v.12, p185-198

BURMUDEZ, José Luis [1994]. “The Unity of Apperception in the Critique of Pure Reason,” *European Journal of Philosophy*, v.2, no3, pp213-240

CAMPBELL, John [1992]. “The First Person: The Reductionist View of the Self,” in Charles & Lennon.

CARRUTHERS, Peter [1986]. *Introducing Persons,* London: Croom Helm.

CASSAM, Quassim [1989]. “Kant and Reductionism,” *Review of Metaphysics*, v.43, 169, pp72-106.

CASSAM, Quassim [1992]. “Reductionism and First-Person Thinking,” in Charles and Lennon.

CASSAM, Quassim [1993]. “Parfit on Persons,” *Proceedings of the Aristotelian Society*, v.93, pp17-37.

CHARLES, David and LENNON, Kathleen (eds.) [1992]. *Reduction, Explanation and Realism* Oxford: Clarendon Press.

COCKBURN, David (ed.) [1991]. *Human Beings,* Cambridge: Cambridge University Press.

COLLINS, James [1954]. *The Mind of Kierkegaard*, London: Secker & Warburg.

COLLINS, Steven [1982]. *Selfless Persons: Imagery and Thought in Theraváda Buddhism*, Cambridge: Cambridge University Press.

CURZER, Howard [1991]. “An Ambiguity in Parfit's Theory of Personal Identity,” *Ratio*, pp16-24.

DAWKIN, Richard [1989]. *The Selfish Gene* (2nd Edition), Oxford: Oxford Univeristy Press.

DENNETT, Daniel [1992]. “Conditions of Personhood,” in A. Rorty.

DESCARTES, René [1641], *Meditations on First Philosophy*. Translated [1986] by J. Cottingham, Cambridge: Cambridge University Press.

ECCLES, J C [1965]. *The Brain and The Unity of Conscious Experience,* Cambridge: Cambridge University Press.

EHRING, Douglas [1987]. “Survival and Trivial Facts,” *Analysis*, v.47, No 1, p50-54.

EVANS, Gareth [1982]. *The Varieties of Reference* (chapter seven) (ed. John McDowell), Oxford: Clarendon Press.

GARDENER, Patrick [1988]. *Kierkegard* , Oxford: Oxford University Press.

GARRETT, Brian [1992]. “Reasons and Values,” *Philosophical Quarterly*, v.42, (168), p337-344.

GLOVER, Johnathan [1988]. *I: The Philosophy and Psychology of Personal Identity.* London: Allen Lane.

HARRE, Rom [1987]. “Persons and Selves,” in Peacocke & Gillet

HALL, Harrison [1984]. “Love and Death: Kierkegaard and Heidegger on Authentic and Inauthentic Human Existence,” *Inquiry*, v.27, p179-197.

HALL, Ronald L [1993]. *Word and Spirirt: A Kierkegaardian Critique of the Modern Age,* Bloomington: Indiana University Press.

HANNAY, Alistair [1982]. *Kierkegaard*, London: Routledge & Kegan Paul

HOBBES, Thomas [1839]. *Hobbes’ English Works* (Volume One), ed. Sir W.Molesworth, London: John Bohn.

HUGHES, R W [1975]. “Personal Identity: A Defence of Locke,” Philosophy 50, pp169-187.

HUME, David [1739]. *A Treatise of Human Nature* (Book I, part IV), reprinted [1962] by Fontana.

JAMES, William [1890]. *The Principles of Psychology* (Vol 1, Chapter X), London: Macmillan

JOHNSTON, Mark [1987]. “Human Beings,” *Journal of Philosophy*, 84, pp59-83

JOHNSTON, Mark [1992]. “Reason and Reductionism,” *Philosophical Review*, 101 (3), pp589-618.

KANT, Immanuel [1781]. *The Critique of Pure Reason*. Translated [1933] by Norman Kemp Smith, Macmillan.

KIERKEGAARD, Søren [1843a]. *Either/Or.* Translated by A.Hannay, Penguin

KIERKEGAARD, Søren [1843b]. *Fear and Trembling/Repetition.* Translated [1983] by H.V. & E.H. Hong), Princeton University Press.

KIERKEGAARD, Søren [1844]. *The Concept of Dread* . Translated [1946] by W.Lowrie, Princeton University Press.

KIERKEGAARD, Søren [1845]. *Stages on Life's Way.* Translated [1940] by W.Lowrie, Oxford: Oxford University Press.

KIERKEGAARD, Søren [1849]. *The Sickness Unto Death* . Translated [1989] by A.Hannay, Penguin.

KITCHER, Patricia [1984]. “Kant’s Real Self,” in Allen W.Wood (ed.), *Self and Nature in Kant’s Philosophy*, Ithaca: Cornell University Press.

KRIPKE, Saul [1971]. “Identity and Necessity,” in Milton K. Munitz (ed.), *Identity and Individuation*, New York: New York University Press.

LEWIS, David [1976]. “Survival and Identity,” in Rorty. Reprinted in Lewis, David [1983]. *Philosophical Papers, Vol 1*, Oxford: Oxford University Press.

LEWIS, David [1986]. *On The Plurality of Worlds* (Chapter Four), Oxford: Basil Blackwell.

LOCKE, John [1694]. *An Essay Concerning Human Understanding* (Book II, Chapter 27), reprinted [1947] by Everyman.

LOCKWOOD, M [1989]. *Mind, Brain and Quantum*, Oxford: Basil Blackwell.

LOWE, E.J [1989]. *Kinds of Being*, Oxford: Basil Blackwell.

MACINTYRE, Alasdair [1985]. *After Virtue* (Second edition), London: Duckworth.

MADDY, Penelope [1974]. “Is the Importance of Identity Derivative?” *Philosophical Studies*, v.35, p151-170.

MARTIN, Raymond [1987]. “Memory, Connecting, and What Matters In Survival,” *Australasian Journal of Philosophy*, v.65. No 1, p82.

McIINERNEY, Peter K [1985]. “Person Stages and Unity of Consciousness,” *American Philosophical Quarterly*, v.22, No.3, p197-207.

McIINERNEY, Peter K [1991]. “The Nature of a Person-Stage,” *American Philosophical Quarterly*, v.28, No.3, p227-235.

MILLS, Eugene [1993]. “Dividing Without Reducing. Bodily Fission and Personal Identity,” *Mind,* 102 (405), pp37-51

NAGEL, Thomas [1971]. “Brain Bisection and the Unity of Consciousness,” *Synthese* 22, p396-413. Reprinted in Perry.

NAGEL, Thomas [1986]. The View From Nowhere, Oxford: Oxford University Press.

NOONAN, Harold.W [1985]. “The Closest Continuer Theory of Identity,” *Inquiry*, v.28, p195-230.

NOONAN, Harold.W [1989]. *Personal Identity*, London: Routledge.

NOZICK, Robert [1981]. *Philosophical Explanations* (Chapter One), Oxford: Clarendon Press.

OAKLANDER, L.Nathan [1987]. “Parfit, Circularity and the Unity of Consciousness,” *Mind*, v.96, p525-529.

PARFIT, Derek [1971]. “Personal Identity,” *Philosophical Review,* 80, pp3-27

PARFIT, Derek [1976]. “Lewis, Perry and What Matters,” in A. Rorty (ed).

PARFIT, Derek [1984]. *Reasons and Persons*, Oxford: Oxford University Press.

PARFIT, Derek [1995]. “An Interview with Derek Parfit,” *Cogito*

PEACOCKE, Arthur & GILLET, Grant (eds.) [1987]. *Persons and Personality*, Oxford: Basil Blackwell.

PEACOCKE, Christopher [1979]. *Holistic Explanations,* Oxford: Clarendon Press.

PEACOCKE, Christopher [1983]. *Sense and Content,* Oxford: Clarendon Press.

PERRY, John (ed.) [1975]. *Personal Identity*, Berkeley: University of California Press.

PERSSON, Ingmar [1992]. “The Indeterminacy and Insignificance of Personal Identity,” *Inquiry*, v.35 (2), p249-269.

PLATTS, Mark de B [1981-82]. “Natural Kind Terms and ‘Rigid Designators’,” *Proceedings of the Aristotelian Society*, 82, 103-114.

PORN, Ingmar [1984]. “Kierkegaard and the Study of the Self,” *Inquiry*, v.27, p199-205.

PUTNAM, Hilary [1975]. “The Meaning of ‘Meaning’,” in *Mind, Language and Reality*, Cambridge: Cambridge University Press. Reprinted from K.Gundersan (ed.) [1975], *Language, Mind and Knowledge*, Minneapolis: University of Minesotta Press.

REID, Thomas [1941]. “Of Memory”, from *Essays on the Intellectual Powers of Man*, London: MacMillan. Reprinted in Perry.

RICOEUR, Paul [1991]. “Life: A Story in Search of a Narrator,” in Mario J.Valdés (ed.), *A Ricoeur Reader*, Harvester Wheatsheaf.

ROBINSON, John [1988]. “Personal Identity and Survival,” *Journal of Philosophy,* 85. pp319-328.

RORTY, Amelie (ed.) [1976]. *The Identities of Persons,* Berkeley: University of California Press.

RORTY, Richard [1980]. *Philosophy and the Mirror of Nature,* Oxford: Basil Blackwell.

ROSE, Steven [1992]. *The Making of Memory*, Bantam Press.

ROVANE, Carol [1990]. “Branching Self-Consciousness,” *Philosophical Review*, v.99 (3), p355-395.

SCHECTMAN, Marya [1990]. “Personhood and Personal Identity,” *Journal of Philosophy*, v.87, p71-92.

SEARLE, John [1980]. “Minds, Brains and Programs,” *The Behavioural and Brain Sciences* III, 3.

SHOEMAKER, Sydney [1963]. *Self Knowledge and Self-identity,* Ithaca, NY: Cornell University Press.

SHOEMAKER, Sydney [1968]. “Self-reference and self-awareness,” *Journal of Philosophy*, 65, pp555-567. Reprinted in Shoemaker [1984].

SHOEMAKER, Sydney [1970]. “Persons and their pasts,” *American Philosophical Quarterly*, 7, pp269-285. Reprined in Shoemaker [1984].

SHOEMAKER, Sydney [1984]. *Identity, Cause and Mind,* Cambridge: Cambridge University Press.

SHOEMAKER, Sydney and SWINBURNE, Richard [1984]. *Personal Identity*, Oxford: Blackwell.

SIDERTIS, Mark [1988]. “Ehring on Parfit's Relation R,” *Analysis*, 48, p29-32.

SINGER, Peter [1979]. *Practical Ethics,* Cambridge: Cambridge University Press.

SNOWDON, P.F [1991]. “Personal Identity and Brain Transplants,” in Cockburn (ed).

SNOWDON, P.F [1994]. “Persons and Personal Identity,” MS.

SOSA, Ernest [1990]. “Surviving Matters,” *Noûs*, v.24, No.2, p297-322.

SPERRY, R W [1966]. “Brain Bisection and Mechanisms of Consciousness,” in J.C.Eccles (ed.) [1966], *Brain and Conscious Experience*, Berlin: Springer-Verlag.

STRAWSON, Peter [1966]. *The Bounds of Sense* (part 2, sec.2; part 3, sec.2), London: Methuen.

STRAWSON, Peter [1959]. *Individuals* (Part One), London: Methuen.

TAYLOR, Charles [1989]. *Sources of the Self*, Cambridge: Cambridge University Press.

UNGER, Peter [1990]. *Identity, Consciousness and Value*, Oxford: Oxford University Press.

WIGGINS, David [1971]. *Identity and Spatio-Temporal Continuity*, Oxford: basil Blackwell.

WIGGINS, David [1976]. “Locke, Butler and the Stream of Consciousness; and Men as Natural Kinds,” in A. Rorty.

WIGGINS, David [1987]. “The Person as Object of Science, as Subject of Experience and as Locus of Value,” in Peacocke & Gillet.

WIGGINS, David [1988]. *Sameness and Substance*, Oxford: Basil Blackwell.

WIGGINS, David [1992]. “Remembering Directly,” in J Hopkins and A Savile (eds.), *Psychoanalysis, Mind and Art*, Oxford: Blackwell.

WILKES, Kathleen [1993]. *Real people.*, Oxford: Clarendon Press.

WILLIAMS, Bernard [1973]. *Problems of the Self,* Cambridge: Cambridge University Press.

WITTGENSTEIN, Ludwig [1953]. *Philosophical Investigations,* Translated [1958] by G.E.M. Anscombe, Oxford: Basil Blackwell. Second Edition [1958].

Acknowledgements

Many thanks to my various supervisors at University College London, Lucy O'Brien, Jerry Valberg, Tim Crane and Ted Honderich.

Thanks also to Reading University's "Famous Five", John Cottingham, T E Burke, Michael Proudfoot, Harry Parkinson and Jim Tiles for the initial inspiration.

Lastly, to my family for their unerring support and belief.