SENTIENT BEINGS AND PERSONS: A NOVEL THEORY OF PERSONAL SURVIVAL

By

WAYNE HENRY

B.A., The University of Victoria, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

PHILOSOPHY

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1987

© Wayne Ian Henry, 1987
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy
The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3
Date Aug 10/87
Abstract

This thesis is concerned with the philosophical problems of personal identity and personal survival. In the first case, we are concerned to establish what our identity as persons consists in at any instant. In the second case we are concerned to establish what our survival as the self-same (i.e.: numerically identical) person consists in. That is, we wish to know what it means to say that one is the same person now which one was ten years ago or will be ten years from now.

I first introduce the distinction between the epistemological question of how we can know these things and the metaphysical question of what these notions actually consist in and claim that much confusion has resulted from the conflation of these two. Further, I explicitly claim that this thesis is intended as a solution to the latter only. I then move on to an historical survey of the major theories of personal identity that have been held since the time of Descartes. After demonstrating how these theories are inadequate, I introduce and explicate the theory defended here which, it is claimed, is a novel one. This novelty consists in the following two distinctions: Firstly, that between persons and sentient beings and, secondly, that between qualitative and non-qualitative psychological relations. It is claimed that sentient beings incorporate the latter in a way which makes them immune to the sorts of contextual problems that typically affect theories of personal identity. Having already
established that we are all sentient beings as well as persons, I then claim that the former concept is the fundamental notion of any theory of personal identity and survival.

I subsequently consider reductionism, which currently prevails in the field, and conclude that such a position inevitably leads to many counterintuitive results. I then compare reductionism with the theory defended here and conclude that the latter is preferable, since it allows us to explain personal identity without abandoning our intuitions regarding what is involved in these matters.
# Table of Contents

Abstract ........................................................................................................... ii

Acknowledgements ......................................................................................... vii

Chapter 1: Introduction ..................................................................................... 1

PART 1: An Historical Survey

Chapter 2: Pure Ego Theories ......................................................................... 11
  §1 The theory as Descartes formulated it ...................................................... 12
  §2 Problems with the theory ........................................................................ 21

Chapter 3: Empirical Theories ....................................................................... 32
  §1 Psychological continuity theories ............................................................. 33
    1. Lockean memory theory ....................................................................... 33
    2. Qualitative psychological continuity .................................................... 43
  §2 Physical continuity theories ..................................................................... 46
    1. The basic theory .................................................................................... 46
    2. Combined continuity theories ............................................................... 55
  §3 The essential inadequacy of empirical theories ..................................... 57
  §4 Hume’s solution ....................................................................................... 60

PART 2: Sentient Being Theories

Chapter 4: Sentient Beings: A Further Fact .................................................. 66
  §1 The basic distinction ............................................................................... 66
  §2 p-space and p-locations ......................................................................... 69
  §3 The properties of sentient beings ............................................................ 73
  §4 The relevance of sentient beings in a discussion of personal identity .... 74
Chapter 5: Constraints for a Theory of Personal Survival

§1 The constraints .................................................. 80
§2 Additional constraints .............................................. 92
§3 Why these constraints are necessary .............................. 94
§4 A discussion of the theories of Chap's. 2 - 5 in terms of these constraints .......................................................... 96

Chapter 6: Sentient Being Theories: Two Alternatives ........ 98

§1 A reiteration of the basic details ...................... 98
§2 Qualitative and non-qualitative psychological relations .......................................................... 99
§3 Two alternatives
1. p-locations as self-persisting entities ..................... 104
2. Gap-crossing relational theory ............................... 114

Chapter 7: The Minimal Neural Activity Theory ............ 120

§1 A review of the problems of the alternatives considered Chapter 7 .................................................. 121
§2 The minimal neural activity theory ............................ 122
§3 The problem of breaks in consciousness .................... 129
§4 A comparison of all 3 alternatives ............................. 131

PART 3: Reductionist Theories

Chapter 8: Reductionist Theories .................. 138

§1 A reconsideration of Hume .................................. 138
§2 The connection with Hume ................................. 142
§3 The truth of reductionism .................................... 145
§4 The theory ..................................................... 150
Chapter 9: Reductionism and the M.N.A. Theory:
A Comparison ................................................................. 159

§1 General problems ....................................................... 159

§2 Reductionism and the M.N.A. theory ............................... 162

Chapter 10: Conclusion ................................................... 166

Bibliography .................................................................... 174

Index ............................................................................. 177
Acknowledgements

I would like to express my gratitude to everyone in the department for their instruction and encouragement, in particular Professor Ed Levy whose patience and advice were always appreciated. I would also like to thank those individuals in particular who read and commented on portions of the manuscript: my fellow student Nick Sleigh, Professor Paul Russell and especially professor Phil Hanson who graciously gave much of his time.

But especially, I would like to thank Professor Dick Sikora for his patience and generosity, which were unflagging. The sentient being theory explicated here is largely his own, my contributions being relatively inconsequential.
CHAPTER 1
Introduction

Let us begin by noting that there has been a considerable degree of confusion in the philosophical discussion of personal identity. Following Swinburne\(^1\) we must recognize that there are two distinct questions concerning personal identity which have typically been conflated. There is the metaphysical question of what it is that constitutes personal identity. This is reflected in such questions as, "What does it mean to say that I am the same person now that I was 10 years ago, or will be 10 years from now?" The second is the epistemological question of what constitutes evidence for the claim that I am the same person now that I was 10 years ago.

There are at least two factors that account for the conflation of these distinct philosophical questions. Firstly, most theories of personal identity since the time of Locke have been empirical in nature. That is, these theories purport to explain what our personal identity is constituted of in terms of characteristics that are, in principle at least, observable. Essentially, such theories have been of two sorts. There are those which explain personal identity in terms of the continuity of certain psychological characteristics such as memory,

character, beliefs, behavioural dispositions and so forth. The other sort explain personal identity in terms of physical continuity, usually of some functioning part of the brain. Additionally, there are other versions which have combined the criteria of psychological and physical continuity. Secondly, it will be shown later that the tendency to explain personal identity in terms of these characteristics is an intuitively attractive one and even has some plausibility. Nevertheless, there are insurmountable problems. The most significant is that all of these empirical theories are guilty of the same flaw: they mistake the criteria for justifying the claim that X is the same person as Y with what it is that constitutes personal identity. In short, they mistake epistemology for metaphysics.

There are some philosophers who would reply that the metaphysical question is really a non-sequitur. The criteria employed to verify or answer such questions constitute all we can know about the matter. The metaphysical question is, properly speaking, not even intelligible. But surely one could only believe this if one also accepts verificationism. I don’t and I believe that I am entitled to an argument before accepting this equivalence. To quote Swinburne,

In general there is plenty of evidence, normally overwhelming evidence, of bodily continuity, memory and character, as to whether or not two persons are the same, which gives very clear verdicts in the overwhelming majority of cases. Yet, while evidence of
continuity of body, memory, and character is evidence of personal identity, personal identity is not constituted by continuity of body, memory, and character. Hence the evidence may on occasion mislead, and two persons be the same [sic], although the best evidence which we have shows they are not, and conversely. Also on occasion the evidence of observable characteristics may give no clear verdict as to whether \( P_m \) is the same person as \( P_i \); but that does not mean that there is no clear answer to the question, merely that we do not know and cannot even make a reasonable guess at what it is.\(^2\)

This thesis is concerned with the metaphysics of what constitutes personal identity at a given instant and personal survival across time. The approach that will be taken is a novel one, as the title suggests, for at least two reasons. Firstly, personal survival is here explicated in terms of a distinction between persons and sentient beings hitherto unexplored by philosophers discussing personal identity. And secondly, this theory relies heavily on the notion of non-qualitative psychological relations.

Regarding the first point above, the distinction between persons and sentient beings is crucial, since the concept of sentient being is the fundamental one on which this theory is grounded. As such then, this is really a theory of survival as the same sentient being. While this may initially seem to be avoiding the crucial philosophical issue of what exactly is involved in "personal" survival, it really is not. It will emerge

\(^2\) Swinburne 240 - 241.
later that, in many cases, the concept of person is open-textured. Indeed, Margalit argues that many of our concepts are open-textured because the rules of our language, which determine how these concepts are to be used, have evolved to accommodate their use in ordinary circumstances. However, there occasionally arise circumstances in which these rules are inadequate, to the extent that we are unable to decide if the concept applies in the given circumstances. With respect to the concept of person, it is in some cases undecidable whether or not someone, qua person, has survived. But, as it will later be argued, the concept of sentient being is immune to these contextual problems. In this way, it is always possible to decide whether or not someone, qua sentient being, has survived. It is because survival as the same sentient being is an all-or-nothing matter in this way that this concept must be regarded as the fundamental notion of any discussion about personal survival.

However, it will later be shown that the concept of sentient being is immune to the sorts of contextual problems described above precisely because it is characterized in terms of non-qualitative psychological relations. This underscores the need for the second distinction noted above, viz, that between qualitative and non-qualitative psychological relations. When we are attempting to establish the continuity of the psychological

---

3 Refer to the article, "Open Texture", by Avishai Margalit in the anthology which she edited, Meaning and Use, (Dordrecht: Reidel, 1979) 141 - 152.
characteristics of persons, we rely on their qualitative similarity across time. Since it is easy to imagine cases of extreme qualitative change of all or part of the psychological characteristics that make up a person, we can find ourselves unsure of whether or not we still have the same person. But the concept of sentient being relies on non-qualitative psychological relations in a way that makes it immune, in principle, to these sorts of contextual problems.

I must here pause to mention that there are several sorts of non-qualitative psychological relations to be discussed here and that one of these in particular distinguishes my own position from that of Professor Sikora's. While this theory is in substance almost entirely his, it was my reluctance to accept his notion of a "non-qualitative, gap-crossing relation" (which is hypothesized as the necessary link that bridges gaps in consciousness in the lives of sentient beings) that led me to formulate the theory expounded in Chapter 7, which I call the "minimal neural activity" theory (hereafter, M.N.A. theory). My own contribution to the details of this theory are virtually insignificant however and I take this opportunity to acknowledge my indebtedness to Dr. Sikora, whose help and guidance have gone far beyond that of merely allowing me to hack his theory apart in this manuscript. Although he will not agree with all the details herein, I hope that I have not misconstrued his own position too badly.
The manuscript itself falls naturally into three parts. Part 1, consisting of Chapters 2 and 3, constitutes an historical survey of the more significant theories of personal identity which have held currency in the history of modern philosophy. I begin in chapter 2, as most investigations of the recent (i.e.: modern) history of philosophy do, with an examination of what Descartes had to say on the matter. His notion of a unitary, indivisible mental substance as the foundation of personal identity is still held by many able philosophers today. After discussing the insurmountable empirical problem of split-brain operations, I move on, in Chapter 3 to an examination of empirical theories that purport to explain personal identity in terms that are observable. As already mentioned, these are basically of two types: those that rely on psychological continuity of one form or another and those that rely on physical continuity of some part, or all, of the body. These are treated together because, it will be argued, all empirical theories fall by the same sword.

There is a dilemma that emerges at this point, which provides thematic continuity throughout the rest of the manuscript. The dilemma is, that there are only two strategies one may take to explain personal identity. Firstly, one may

attempt to explain it in terms of some "further fact", such as a soul or a Cartesian ego, or, secondly, in terms of some observable characteristics (i.e.:an empirical theory). It will have already been demonstrated that both attempts fail and the only other alternative is, following Hume, to reject the notion of personal identity altogether.

The stage will have been set for the introduction, in Part 2, of sentient being theories. Such theories are "further fact" theories but, unlike Descartes', this further fact is nothing like an immaterial ego. Rather, it amounts to the claim that any complete account of persons must also make essential reference to the fact that they are also sentient beings. The basic details of such a theory will be developed in Chapters 4 and 5 and in Chapter 6 two versions of such a theory will be examined and rejected. Then, in Chapter 7, the version to be defended in this thesis (i.e.:the M.N.A. theory) is presented.

We will take up Hume again in Part 3 when discussing Parfit's reductionism. It will there be argued that Parfit has taken Hume's gambit of rejecting the notion of personal identity

---

The "further fact" classification is relative to reductionist theories like Derek Parfit's (to be discussed at length in Part 3) which claim that persons are nothing more than a particular brain and body and an interconnected series of mental and physical events. On Descartes' theory the immaterial ego is conceived of as a separately existing entity, which is a further fact to be accounted for in addition to one's body and such a series of interrelated physical and mental events.
as it is normally construed. In fact, Parfit treats personhood as a scaling concept by attempting to show that personal survival is really a matter degree. In Chapter 9, I will compare the M.N.A. theory developed in Chapter 7 with Parfit’s reductionist theory, arguing that his treatment inevitably leads to highly counterintuitive results. I will further argue that these results are a consequence of treating the concept of person as a scaling concept, which is antithetical to its treatment in ordinary language. By comparison I will argue that the M.N.A. theory will allow us to have our cake and eat it too. Specifically, this theory allows us to explain what is involved in personal survival in a way that is commensurate with our intuitions on these matters, while at the same time allowing us to avoid the uncomfortable consequences of reductionist theories such as Parfit’s. In particular, it allows us to explain how our survival is an all-or-nothing matter.

In Chapter 10 we will again take up the dilemma posed above and conclude that the concept of sentient being must function as the fundamental notion in any discussion of personal survival. The reasoning will be that, only such a further fact can adequately explain what is involved in our intuitions regarding these matters. Additionally, such a theory allows us to capitulate to the intuitions of the empiricist who argues, quite rightly, that personhood is at least partly a matter of the continuity of certain psychological characteristics such as
personality and so forth. Yet this is achieved without accepting the empiricist's gambit and we are thereby able to avoid the conclusion accepted by Parfit, that personal survival simpliciter is a matter of degree.
PART 1

AN HISTORICAL SURVEY
CHAPTER 2
Pure Ego Theories

There is good reason for beginning our historical survey with Descartes, not only because his view, or at least something roughly like it, is probably the one most commonly believed by persons untrained in philosophical thinking but also because it is still considered to be a live option by many philosophers to this day. Actually Descartes himself never considered the problem of personal identity (at least not to the best of my knowledge), although what he would have said on the matter, and what his followers have in fact said, is fairly easy to reconstruct. The reason Descartes ignored this problem is that he took the existence of the ego, whose essence is to think, as his epistemological starting point. That is, the existence of a concrete, indivisible entity with which the true self is to be identified, was his fundamental premise. As the foundation for the refutation of his own thorough-going skepticism, he took the notion of the ego (in the sense noted above) as axiomatic and was not concerned with (in fact, we may say, not even aware of) the philosophical problem of personal identity.

I will begin with a brief sketch of the particulars of Descartes' metaphysical theory and will then consider the arguments he offered in defence. Along the way I will interject the implications of this position for our own problems of
personal identity at an instant and personal survival across time.

51

Essentially, Descartes maintained that the true self is to be identified with an indivisible, indestructible and active entity which he calls the ego. Now this ego is "made of" an immaterial (i.e.: unextended) substance which he calls "mental substance" and the activity, or the essence as Descartes puts it, of this substance is thought. Further, this immaterial ego is capable of existing independently of one's physical body and, even when associated with a body, the connection is construed to be rather like a driver at the wheel of his car. This independence of the ego is grounded on Descartes' radical dualism which postulates the existence of two entirely distinct types of substance: physical substance whose essence is extension (i.e.: filling space) and mental substance whose essence is thought. In terms of personal identity then, this is a "further fact" theory, since any account of what it is to be a person on

---

"That is, thought is the essential attribute of mental substance in the sense that that is what it does when it is active, although it need not be doing this all the time. Further, as we shall see later, Descartes had a notion of the mind as constituted of several different faculties such as thinking, willing, sensing and so forth. But all of these faculties are conceived of as thought to the extent that they are all part of the activity of mental substance."
such a theory must make essential reference to the further fact of this immaterial ego."

Now Descartes would claim that, in fact, that is all there is to our personhood; we are, in so much as we are persons, just these egos. Since the essence of this ego is to think and since, for Descartes, our essence as persons is to think, it seems to follow quite neatly that what we are really referring to when we say "I" is just this ego which is the seat of all the characteristics which make us distinctively human, "Just because I know with certitude that I exist, and because, in the meantime, I do not observe that aught necessarily pertains to my nature or essence, beyond my being a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing." And, on a personal level, it is this ego which makes us distinct from all other persons; it is the locus of all our distinctive character traits and so forth, "When I consider the mind, that is to say, myself inasmuch as I am only a thinking thing, I cannot distinguish in myself any parts, but apprehend myself to be clearly one and entire." And, regarding the functions of the ego,

---

7 Recall from Chapter 1 that further fact theories are contrasted with reductionist theories of personal identity.

8 Descartes, Meditations 270 - 271.

9 Descartes, Meditations 276.

10 Descartes uses the terms "ego", "mind" and "soul" as synonyms.

13
After having thus considered all the functions which pertain to the body alone, it is easy to recognize that there is nothing in us which we ought to attribute to our souls excepting our thoughts, which are mainly of two sorts, the one being actions of the soul, and the other its passions. Those which I call its actions are all our desires, because we find only by experience that they proceed directly from our soul, and appear to depend on it alone.\footnote{\textit{Descartes, Passions} 370.}

Later, in the same passage, he goes on to describe the sorts of activities which should be regarded as "actions" of the soul. These include willing, which moves the body accordingly to satisfy its desires, and thought, which is meant to include reflection, recall and imagination.

Descartes regarded all other animals as simply "automatic machines"; quite literally, all other animals are organic machines subject only to the causal laws of mechanics. As such they are "brutes" without will, thought or feeling. In fact, this conception of his was so thorough-going that he maintained that animals did not even feel pain (or experience any other sensations for that matter) but, due to perfectly natural causal laws they conveniently behaved "as though" they were in pain by writhing, screaming and so forth.\footnote{Convenient in the sense that God, because of his beneficent regard for man, has so constructed the natural order of things that animals behave as we expect them to when injured and so forth.}
However, the interesting point to emerge from all this is that it is the possession of this immaterial ego which sets us apart from the animals in all respects. In terms of his dualism, this means that animals are purely physical beings whereas people have a dual nature. Our bodies are physical entities but the essential part of ourselves are these immaterial egos which persist after the death of our physical bodies, "I assume that the body is nothing else than a statue or machine made of clay which God forms expressly to make it as nearly like as possible to ourselves, so that not only does he give it externally the colour and the form of all our members, but also he puts within it all the parts necessary to make it walk, eat, breathe and, in fine, imitate all those of our functions which may be supposed to proceed entirely from matter and to depend merely on the arrangement of organs". As such, the ego is the seat of our sensations and all our mental faculties, including the will which, for Descartes, was radically free. In this way we are free of the determined causal order of things. As moral agents we are, ourselves, origins of causal processes in the world. Unlike brute animals which are merely jerked about higgledy-piggedly by causal determinants, humans are radically self-determining in the sense that it is this ego which "steers" the body as it were. And, when in pain we are "truly" in pain, precisely in the sense that

---

13 Descartes, Treatise 350.

14 The exact nature of this connection between the body and the ego which "steers" it will be dealt with at greater length below.
our behaviour is accompanied by a very real sensation. More than this, because the mental substance of which these egos are made is indestructible, our egos are everlasting. Furthermore, since our egos are indivisible and unitary (i.e.: distinct from all other egos) and incorruptible (not morally but, rather, in the sense that they are immune to dissipation or fusion with other egos or impurities), we enjoy an everlasting life as "ourselves". This reinforces what has already been mentioned, viz, that Descartes regarded the ego as the seat of all our personally distinctive character traits, beliefs, behavioural dispositions and so forth. In terms of personal survival, this theory cashes out as the view that we persist as the self-same (i.e.: numerically identical) ego. Although we may rely on the evidence of character traits and so forth to determine whether or not someone is who we think he is, it remains certain that one remains the self-same person, regardless of whether or not we can empirically verify the fact,

[Although I certainly do possess a body with which I am very closely conjoined; nevertheless, because, on the one hand, I have a clear and distinct idea of myself, in as far as I am only a thinking and unextended thing, and as, on the other hand, I possess a distinct idea of a body, in as far as it is only an extended and unthinking thing, it is certain that I, [that is, my mind, by which I am what I am], is entirely and truly distinct from my body, and may exist without it.]

---

15 Descartes, Meditations 270.
We may summarize these results as follows. Descartes conceived of the ego as a kind of substratum composed of immaterial substance. As such, it is identified with the mind which is in turn identified with the "I" which is one's true self. This is presumed to follow from the "fact" that thought is our true essence in so much as we are persons, and thought is also the essential activity of mind. Further, it is capable of persisting after the death of our physical bodies. Within this mind, one may identify several different faculties, such as thought, will, sensation and so on. It is claimed that these are all faculties of the one immaterial substance; they are all activities of mind.

Descartes of course was concerned to make this position philosophically respectable, crucial as it was to his entire philosophical enterprise. In fact, his arguments in defense of this theory were quite ingenious, which goes a long way towards explaining the longevity of the theory itself despite its significant problems. In explicating his arguments I will be concerned to establish two things. First, how Descartes thought he had proven that our essence is thinking and, secondly, how he thought he had proven that this essence of ours exists as an immaterial entity which is, in fact, independent of our

---

physical bodies. It will be seen that the two work together to establish the existence of the ego as described above.

Initially the move from the claim that, since I think, I am certain that I exist to the claim that therefore my essential nature is thought, may seem puzzling. Certainly Hobbes thought so when he claimed that it was similar to arguing that, "I am walking, hence I am the walking". However, Malcolm argues persuasively that Descartes, having established the certainty of his own existence, sets out to discover what his own nature is with a criterion in hand. The criterion is a two-part test such that, 1) if in being aware of myself I am necessarily aware of a particular attribute and 2) if in being aware of a particular attribute I am necessarily aware of myself, then I will have discovered my essential nature. This attribute will be my "essence" precisely in the sense that it is a property which is constitutive of myself; the two are necessarily conceived of as inseparable. Malcolm states this criterion as follows,

\[ G: x \text{ is my essence if it is the case that if I am aware of } x \text{ then (necessarily) I am aware of myself and if I am aware of myself then (necessarily) I am aware of } x. \]

Now for Descartes, the only attribute that satisfies this criterion is thought. Firstly, it should be recalled that

---

17 Malcolm 312.

18 Malcolm 316.
Descartes uses the term "thought" very broadly to include any mental activity. Secondly, Descartes holds that we are necessarily aware of every thought we have; for Descartes, to think is to be aware of thinking. Given these points, it follows that, "for Descartes, if I am aware of anything then I am thinking, and so, if I am aware of thinking then I am thinking; and if I am thinking I am aware of thinking". Now since he also holds that, whenever we think we are also aware of ourselves (this was precisely the insight implied by the "cogito"), it follows that the attribute of thought passes the first test. That is, if I am thinking then I am necessarily aware of myself.

Given this, it follows easily that, if I am aware of myself then I am necessarily aware of thinking and so thought also passes the second test. For consider that, if I am aware of anything then I am thinking and since I am necessarily aware of every thought I have, I will necessarily be aware that I am thinking. Now since awareness of myself is just a particular instance of thinking, it follows that whenever I am aware of myself I will necessarily be aware of thinking. For Descartes, we are aware of ourselves iff we are aware of thinking.

Regarding the second point, we shall now see how Descartes thought he had established that minds are radically independent of bodies. The proof itself turns on conceivability. Essentially

17 Malcolm 318.
Descartes argues that, since in conceiving of himself he necessarily conceives of thought (as we have just seen) but does not similarly conceive of corporeality (i.e., extension), it follows that thought is essential to him but corporeality is not. Certainly awareness of one’s body will not pass the second test of Descartes criterion. It will pass the first of course, since in being aware of anything I will necessarily be aware of myself. But it should be obvious that it is false that in being aware of myself I am necessarily aware of my body and since extension is not essential to my nature, it follows that "I", qua immaterial ego, am independent of my body. To quote Descartes,

In the next place, I attentively examined what I was, and as I observed that I could suppose that I had no body, and that there was no world nor any place in which I might be; but that I could not therefore suppose that I was not; and that, on the contrary, from the very circumstance that I thought to doubt of the truth of all things, it most clearly and certainly followed that I was; while, on the other hand, if I had only ceased to think, although all the other objects which I had ever imagined had been in reality existent, I would have had no reason to believe that I existed; I thence concluded that I was a substance whose whole essence or nature consists only in thinking, and which, that it may exist, has need of no place, nor is dependent on any material thing; so that "I", that is to say, the mind by which I am what I am, is wholly distinct from the body, and is even more easily known than the latter, and is such that, although the latter were not, it would still continue to be all that it is.\(^{20}\)

\(^{20}\) Descartes, Discourse 171 - 172.
In this passage then, we are witness to the seamlessness of Descartes' thoughts on all the matters that concern us here; that his essence is thought, that this essence of his exists independently of his physical body and that it is his true self.

§2

Now the obvious problem with this conception is that it fails to explain how the interaction between the mental and the physical is supposed to take place. For if the two kinds of substance are radically independent in the way claimed, then how can the ego "steer" the body? How can something which is completely immaterial and unextended interact with a material, extended object? Descartes himself was aware of this shortcoming and he (perhaps jokingly) referred to the pineal gland as the locus of interaction. But, of course, to indicate a locus of interaction is not to explain how the interaction actually occurs. In defence of Descartes however, I think it is probably fair to say that his thoughts on the matter were that there is indeed some particular locus of interaction, probably in the brain, where the two substances interact in some way perfectly compatible with the normal causal order.\footnote{Compatible in the sense that the ego interacts with the physical order by initiating a physical event, vis a vis the locus of interaction, which subsequently sets in motion a series of physical events that follow each other according to normal laws of motion.} It is a testament to his buoyant optimism in the "new" scientific world-view then emerging, that Descartes thought that the explication of this
interaction was a straightforward scientific problem that would be resolved in good time.

We now know that the problem is a good deal more resistant to solution than he had anticipated, but there is another route that someone wishing to defend a Cartesian ego theory could take. Recall that Descartes' defence of the radical independence of mental and physical substance turned on the fact that he could conceive of the two as independent in this way without contradiction. Therefore, it would be a mistake to categorize minds as physical entities. Yet it seems to be an empirical fact that minds are always associated with bodies or, more specifically, with brains. In fact, although we don't know all the details, there is good reason to suppose that brains "cause" minds. But it is open to the sophisticated neo-Cartesian to concede this and claim that what Descartes actually proved was that there is some possible world where minds exist independently of bodies. In this world, the actual world, it is an empirical fact that brains cause minds, but it is a logical possibility that they should exist independently. Since I find no evidence that Descartes himself distinguished between physical and logical possibility, I make the claim that he mistakenly took his proof of logical possibility to apply to the actual world.

On this construal it may seem that the neo-Cartesian's position scarcely differs from standard dualistic theories of
mind but this is mistaken, for this neo-Cartesian still maintains that in addition to the mind there is also this further fact of an indivisible, unitary ego with which the "I" is to be identified. The ego is no longer identified with the mind in its entirety but, rather, only with that portion of it which constitutes our self-reflective awareness. It is like a spectator to the events of our mind from a privileged, first-person perspective. That is, the ego is now conceived of as that portion of our mind to which we attribute self-consciousness. To borrow Sartre's terminology, the ego is the transparent "I" which necessarily accompanies all our thoughts. Of course this is also reminiscent of Kant's way of speaking about the matter.

Let us then, following Grice, refer to this type of theory, as distinct from Descartes' conception, as a "pure ego" theory. On such a theory, the ego is conceived of as an ultimate particular, which is to say that, "The self has qualities and stands in relations, without either being or containing qualities or relations", and that this definition, "is meant to exclude from the class of ultimate particulars all entities which are complex in the way in which the fact that "This is red" or the event consisting in, "That noise being heard" are complex; for such entities, though they may be particulars, contain qualities

---


23 Refer to the article, "Personal Identity" in the Perry anthology.
or relations as elements, and are not, therefore, ultimate particulars'.

Given the identification of the ego and our self-conscious awareness then, it is this ego which explains the unity of our lives as persons. That is, we can know, beyond any shadow of a doubt, that we remain the same person throughout our lives because the ego, the "I" which we are directly aware of when we introspect, is immutable and incorruptible and, further, it is this "I" which is the true self. For consider that, unless it were connected with publicly observable or introspectible facts, one could never verify, even in principle, that one had survived. This is not merely crude verificationism either, but rather a case of refusing to admit an entity into our ontology which explains nothing.

Now it seems that there are two alternatives possible at this point. Regarding personal survival, one could maintain either that the ego, as conceived above, dies with us when we die or that it continues to exist independently after our physical bodies expire. The latter view might be the one taken by some Christians where it is supposed that, upon conception, God infuses the foetus with a unique and everlasting soul. In this way our faculties of sensation and emotion are still explained in the usual way as the causal effects of the brain but the ego,

--- Grice 76.
which still must be conceived of as the locus of our self-reflective awareness (and our distinctive character traits, since we are guaranteed "personal" survival), persists after our death. Alternatively, one could maintain that the ego is also a causal effect of the brain and expires with us when we die. Both of these are committed to the claim that any account of our survival as the "same" person must make essential reference to this ego though, to be a Pure Ego theory at all.

Now initially, such a theory as the last one considered above might seem plausible and perhaps even convincing. In fact I think many persons actually believe something roughly like this. In substance then, this theory commits one to the view that, although the mind is a natural causal result of the neurochemical processes of the brain, there is still something left out if this is proposed as a complete description of our lives as persons. What is left out is precisely the notion of this ego which, although distinct from the mind, is a concomitant effect of the same brain processes which cause the mind. It is this transparent, indivisible and incorruptible entity which is the self-reflective part of our consciousness; our self-consciousness if you will.

However, I am not actually familiar with anyone holding such a view.
The intuition underlying the plausibility of this view was certainly convincing to Descartes. Consider that when you introspect you are immediately aware that you are aware. More importantly, the first-person intimacy of this awareness indicates that it is the "I" introspecting oneself and that this awareness is private from everyone else. As Descartes himself put it, even when I doubt, I cannot doubt that it is I who doubts. From this it seems to require a small step to conclude that, in some sense, I am more than simply a mind. I am also this self-reflective awareness, which is the "I" whom I carry on my personal dialogue with and which is what I associate what is distinctively myself with.

Let us review the three alternative theories just considered. Firstly, there is Descartes' own version. On this theory, the immaterial ego is identified with the mind in its entirety; all of our mental faculties inhere in the immaterial substance of which this ego is composed. Further, this ego is immutable and everlasting. Secondly, there are two versions of "pure ego" theories. On both of these, the ego is identified only with our self-reflective awareness, all our other mental faculties being explained in the normal way as the causal effects of the brain. The difference between the two is that, the first conceives of this self-reflective awareness as an everlasting entity akin to a soul, while the second conceives of it as contemporaneous with the lives of our physical bodies.
Presumably, the motivation to the second of these is to dispense with any notion of an ego conceived of as a "ghost in the machine". That is, on this theory, the ego is also a causal result of the brain, but is retained as the only "reasonable" explanation of what we are aware of when we introspect; it is the subject of our experiences and the "person" with whom we carry on our personal dialogue.

Unfortunately, the results of split-brain research have disastrous consequences for anyone holding any sort of ego theory at all. It was discovered, several decades ago that severing the corpus callosum could often relieve the symptoms of epilepsy. The corpus callosum is the sinewy matrix of fibers which connect the two halves of the cerebral hemispheres. Severing these fibers in effect cuts the brain in half.\[^{22}\] Subsequent research has also discovered that persons who have had such an operation behave as though they were two different persons occupying the same body. Interpretations of these results are divided though. R.W. Sperry, a prominent neorophysician and seminal figure in this field, argues for the strong claim that we now have two persons where, formerly, we had one. Of course both persons are continuous with the single individual who is their progenitor but neither is numerically identical with him. He maintains that the evidence clearly indicates that there are two

\[^{22}\] Actually the lower and mid-brain are left intact but, since these don't contribute to the higher brain functions, they are irrelevant in this context.

27
persons with very different mental capabilities (that reflect the differential capabilities of their respective hemispheres) but, more importantly, they can also have very different thoughts, desires and so forth which remain hidden from each other. There are abundant examples such as the man who reached out for his wife with one hand only to push her away with the other, or the man who picked up a newspaper only to have his other hand grab it and throw it down. Further, there is no communication between the two hemispheres at all.

One of the more general and also more interesting and striking features of this syndrome may be summarized as an apparent doubling in most of the realms of conscious awareness. Instead of the normally unified single stream of consciousness, these patients behave in many ways as if they have two independent streams of conscious awareness, one in each hemisphere, each of which is cut off from and out of contact with the mental experiences of the other. In other words, each hemisphere seems to have its own separate and private sensations; its own perceptions; its own concepts; and its own impulses to act, with related volitional, cognitive, and learning experiences. Following the surgery, each hemisphere also has thereafter its own separate chain of memories that are rendered inaccessible to the recall process of the other.\textsuperscript{7}

Thomas Nagel, on the other hand, argues that we don't actually have two minds but that, in fact, there is no determinate number of minds at all. Rather, he generalizes the results to claim that there is no unity to the mind, even when

\textsuperscript{7}Sperry 724.
undivided. His argument rests on the claim the mind is modular in its construction. Specifically, he claims that the mind is composed of facultative units, for example the faculties of speech, olfactory sensation, tactile sensation and so on. Each of these faculties is located in a particular area of the brain (although there is some redundancy of function) and, in the normal, undivided brain they communicate and cooperate with each other in a manner which gives the illusion that there is a single subject of awareness.

The illusion consists in projecting inward to the center of the mind the very subject whose unity we are trying to explain: the individual person with all his complexities. The ultimate account of the unity of what we call a single mind consists of an enumeration of the types of functional integration that typify it. We know that these can be eroded in different ways, and to different degrees. The belief that even in their complete version they can be explained by the presence of a numerically single subject is an illusion. Either this subject contains the mental life, in which case it is complex and its unity must be accounted for in terms of the unified operations of its components and functions, or else it is an extensionless point, in which case it explains nothing.

It is my conjecture that this "illusion" of unity (i.e.:that there is a single subject of awareness) is not really illusory at all. Rather, since all the various "modules" contribute, in a normal undivided brain, to a single stream of consciousness, there really is only a single subject of awareness after all (the details of this will be elaborated in Part 2 when dealing with the sentient being theories).

Nagel 242 - 243.
With respect to both of the neo-Cartesian positions sketched above, these results establish the same point however. The unity of consciousness cannot be maintained in the face of the results of split-brain research. Consequently one is forced to hold either that there is no such entity as an immaterial ego or that it cannot be identified with our self-reflective awareness. In the latter case, the ego cannot be used to explain anything which it has been typically employed for; it cannot be the subject of experiences which one refers to when using the personal pronoun "I", and it therefore cannot be what explains the unity of our lives as persons.

Now the neo-Cartesian might resist this conclusion by attempting to argue that the experimental results merely establish a division between the functions of the brain that constitute its non-reflective awareness but that the self-reflective ego is still privy to the goings-on in both hemispheres simultaneously, in a manner implying unity at the level of the ego. However, the evidence of Sperry and other's research contradicts this view. Information can be withheld from one hemisphere (for example showing a picture to one side only) and the other hemisphere is unable to answer questions regarding that material. This implies that there is no self-reflective ego which is simultaneously aware of the content of both streams of consciousness. They are truly isolated from each other and function as self-contained, autonomous individuals. But clearly
this is impossible if the ego is supposed to be indivisible and incorruptible. If it were maintained that the ego has been divided along with the mind, then it would no longer be a pure ego, for recall that we have already seen that such a pure ego, conceived of as an ultimate particular, cannot be composed of parts. Consequently, it could longer perform the function required of it in the context of a theory of personal identity; it would no longer constitute our self-reflective awareness which remains unchanged and unitary to individuate one from all other individuals. It would, in short, be a redundant entity that would explain nothing. For this reason it is here claimed that Cartesian ego theories are false as a matter of empirical fact.
CHAPTER 3

Empirical Theories

In this chapter we will discuss what are called, by Swinburne,30 empirical theories of personal identity. These differ dramatically from the "pure ego" theories considered in the previous chapter. Where these latter purport to explain the unity of one's life as a person vis a vis essential reference to a unique ego, empiricist theories explain this unity in terms that are, in principle, observable. Specifically these theories are of two sorts; those that rely on the continuity of psychological characteristics and those that rely on the continuity of physical characteristics. They are considered together here because they fall together by the same sword. To anticipate the results of our analysis, it will be shown that, if one takes the empirical method to its logical conclusion in the manner of Hume, one is forced to abandon the notion of personal identity. Consequently, Hume is also considered here and he will come up again in our discussion of Parfit's reductionist theory in Chapter 8. There it will be argued that Hume's rejection of personal identity is the logical progenitor of Parfit's reductionism.

In this section we will consider theories that rely on the continuity of psychological characteristics. All such theories derive ultimately from Locke so our analysis begins quite naturally with him.

1. Lockean memory theory:

As a theorist of personal identity, Locke is actually a transitional figure between Descartes and Hume. Like Descartes, he believes that each of us, in so much as we are persons, is possessed of an immaterial soul which is capable of surviving our bodily death. However, unlike Descartes, he did not think that our personal identity was constituted of this immaterial soul. Locke seems to retain the notion only as a convenient explanation of the distinction between living and inert matter (and also perhaps as a hold-over of certain religious convictions). His arguments against the Cartesian position are more than mere rhetoric however. In fact, they are very revealing of what he actually took our personal identity to be constituted of and so, must be considered here.

Firstly, he reasoned, if our identity as persons consisted in identity of the same immaterial soul and if this same soul were, as some suppose, reincarnated in different bodies, then it follows that these individuals would be the same person. But this is absurd, reasons Locke, for it is impossible that two
individuals with no similarity of memory, personality or temper should be the same. More absurdly, it is at least logically possible that such an immaterial soul could be reincarnated in other than a human body,

For, if the identity of soul alone makes the man, and there be nothing in the nature of matter why the same individual spirit may not be united to different bodies, it will be possible that different men living in distant ages, and of different tempers, may have been the same man; .... But yet I think nobody, could he be sure that the soul of Heliogabalus were in one of his hogs, would yet say that hog were a man or Heliogabalus.\footnote{Locke 444 - 445. I contend that, in this quote, Locke is using the term "man" as a synonym for the term "person".}

Secondly, he argued that it is incumbent on such a theorist to show why personal identity could not be preserved across changes of immaterial substance. Having already demonstrated that the identity of vegetables and animals is preserved despite the changes in the material substance of which they are constituted, he argues that the case of immaterial substance must be regarded as analogous in the absence of any argument to the contrary, "And therefore those who place thinking in an immaterial substance only, ... must show why personal identity cannot be preserved in the change of immaterial substances, or variety of particular immaterial substances, as well as animal identity is preserved in
the change of material substances, or variety of particular bodies".33

Thirdly, he argued from the logical possibility of the division of souls that, were this to occur, there would then be two immaterial souls that were, in fact, the same person. But this is absurd since it is impossible that two persons (or two items of any sort for that matter) should be numerically identical, "[I]t must be allowed, that, if the same consciousness (which, as has been shown, is quite a different thing from the same numerical figure or motion in body) can be transferred from one thinking substance to another, it will be possible that two thinking substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved."34

Lastly, he argues that, if reincarnation is presumed to occur, then it is evident that the soul carries no memory of its previous life with it. He takes this as an empirical fact. Given this, we have no reason to suppose that the two lives sharing the same soul do not constitute unique individuals,

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

33 Locke 453.
34 Locke 454.
him also suppose that it was the same soul that was in Nestor or Thersites, at the siege of Troy .... which it may have been, as well as it is now the soul of any other man: but he now having no consciousness of any of the actions either of Nestor or Thersites, does or can he conceive himself the same person with either of them? 34

Of these, only the second is at all convincing, for all the others seem to presuppose what Locke will later argue for; that our personal identity consists in our self-consciousness. But why should Locke have singled out self-consciousness as that which constitutes our self-identity? There are several clues in the quotes above. For instance, in the last quote, Locke seems to imply that, without consciousness (using "consciousness" here to mean memory of a previous conscious act) of the actions of either Nestor or Thersites, one could not even conceive of oneself as identical with either of these individuals. Now this may mean that he thought it was logically impossible that two persons could be the same without the same consciousness, but this would make his claim uninteresting. I think that what is really going on here is, firstly, Locke thought that in such a situation there would be no reliable empirical evidence to support the claim that the two individuals were the same person. This is further supported by the quote immediately preceding; he assumes there that two numerically distinct souls sharing the same consciousness must be the same person because, by all empirical

34 Locke 455 - 456.
evidence, they would be indistinguishable. But secondly, I think he also thought that consciousness and memory of our previous conscious acts carries with it the moral and legal obligations that are the lot of moral agents. It is inconceivable that I should regard myself as the same person as Nestor for then I should also be obliged to accept all rewards and punishments that may be due him, even though I should have no memory of the events themselves.

But let us follow Locke’s own reasoning by identifying the strategy he uses to establish the identity of anything. This strategy consists of two steps. Firstly, one must establish what it is that makes an object, X, at some point in time, T₁, identical with itself (i.e.: distinct from all else). This is what Locke refers to as the "Principium Individuationis". Secondly, one establishes that the object, Y, at time Tₛ, is the same as X at time T₁ by establishing a continuity between X and Y. In the case of inanimate objects this latter is a fairly straightforward task, for what identifies the object at time T₁ is existence itself, "[I]t is easy to discover what is so much inquired after, the principium individuationis; and that, it is plain, is existence itself, which determines a being of any sort to a particular time and place, incommunicable to two beings of the same sort". It is clear that Locke’s meaning here is that the existence of any object at a particular time and place excludes

---

[37] Locke 441.
all other objects of the same sort and so, we can be sure that it is itself and no other.

Regarding the continuity of an inanimate object with itself at a later time, Locke tells us that this consists in the continued existence of a similar lump of particles sharing the same configuration. But to get clear on this, we must establish what he means by the terms "similar" and "configuration". It was obvious to Locke that one could replace parts of an object over time without making it a numerically different object. The standard example is of a ship which, over the decades, has parts replaced during the course of normal repairs such that, it eventually no longer has any of the original parts. On the one hand, we should wish to say that it was still the same ship. Yet, on the other hand, it no longer contains any of the original materials. So how are we to explain our intuitions that it is still the same ship?

Locke's response is that, firstly, we still have a similar lump of particles. That is, all the parts which have been replaced have been replaced by particles that were qualitatively similar (i.e.: made of the same material, of the same dimensions, etc.). Further, there is a continuity linking the various stages of the ship such that, at no time have all the parts been replaced simultaneously. We may state this more exactly as follows:
If, at any time \( T_n \), object \( Y \) is the numerically same object as \( X \) at a previous time \( T_p \), then \( X \) and \( Y \) share some parts in common and there is an overlapping chain of such shared parts linking the two.

In this way, we may say that there is an overlapping chain of shared parts linking the ship throughout its duration. But Locke also goes on to say that the parts which constitute the object must also maintain the same configuration. From the context and also from what he says regarding plants and animals (i.e.: the parts constituting plants and animals continue to "participate in the same life"), I maintain that his intended meaning must be something roughly like what follows. All the parts which constitute any given object, are arranged in certain ways such that, to disturb this arrangement either destroys or alters the object in such a way that it becomes a numerically different object. An obvious example of the former is the complete disassembly of the ship. The complete disturbance of the arrangement of the parts has destroyed the object. Regarding the latter, consider this example. The ship is retired from active duty and bought by a wealthy merchant who has it reduced in size and drastically altered so that it now serves as a yacht. In this case, the arrangement of the parts has been so altered that we now have a numerically different object, although it has not been destroyed altogether. Given these observations, it seems that Locke's intended meaning of "configuration" could only have been something like, "the general arrangement of the parts and the
relations between them which, minimally, must be maintained for the object to remain numerically identical".

In the case of plants and animals, the case becomes somewhat more complicated since even the above notion of configuration may not be preserved throughout the life of the organism. A simple example is of the caterpillar which becomes a butterfly. While we can still identify the organism at any particular time, $T_1$, as self-identical by the criterion of existence, the continuity of a similar lump of stuff sharing the same configuration will not suffice as a criterion to identify it at a later time. In such cases, Locke claims that continuity is established by the continued participation of the constituent parts in the same life, "That being then one plant which has such an organization of parts in one coherent body, partaking of one common life, it continues to be the same plant as long as it partakes of the same life".\footnote{Locke 443.} It seems obvious that the notion of the "continued participation of the constituent parts in the same life" is intended to function analogously to the notion of "configuration" in the case of inanimate objects.

But the case of persons is even more complicated if we carefully analyze what is meant by the word "person". Locke maintains that there is a distinction between "man" and "person" such that the continued identity of a man can be established in

\footnote{Locke 443.}
the same way as that of plants or animals. That is, the term "man" refers to the living body of the organism alone. But this is not the case for persons on the accepted meaning of the word,

[W]e must consider what person stands for; which, I think, is a thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does by that consciousness which is inseparable from thinking, and, as it seems to me, is essential to it: it being impossible for anyone to perceive without perceiving that he perceives.

Further, he argues that this is the only explanation of our concern for our own happiness, success and so forth; it forms the basis of our moral notions of punishment and reward, "Person, as I take it, is the name for this self. Wherever a man finds what he calls himself there, I think, another may say is the same person. It is a forensic term, appropriating actions and their merit; and so belongs only to intelligent agents capable of a law, and happiness, and misery." 37

Having established self-consciousness as the principium.

33 Locke 448 - 449.

33' Locke 466 - 467.

40 Although Locke used the term "consciousness" in the quote above, I am certain that his intended meaning in that context is "self-consciousness". Locke’s use of the term consciousness was notoriously ambiguous and was freely used to refer to memory, self-consciousness, consciousness (as opposed to unconsciousness) and perhaps other meanings as well, as pointed out by Anthony Flew in, “Locke and the Problem of Personal Identity”, Locke and Berkeley: A Collection of Critical Essays, ed. C.B. Martin and D.M. Armstrong, (New York: Doubleday and Company, Inc., 1968), 41
individuation of persons, Locke’s next task was to establish the criterion of continuity for persons. He claims that this lies in the continuity of consciousness itself, which is the constant companion of all our thoughts,

For, since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things: in this alone consists personal identity, i.e., the sameness of a rational being; and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.41

Now in this case Locke is using consciousness to refer to memory when he talks of it being extended backwards to past actions and thoughts and, so, the criterion of continuity for persons is memory. This may be stated more clearly as follows:

\[ P_m \text{ at time } T_m \text{ is the same person as } P_1 \text{ at time } T_1 \text{ iff } P_1 \text{ and } P_m \text{ are both persons and } P_m \text{ contains an actual memory at } T_m \text{ of an experience had by } P_1 \text{ at } T_1 \]

155 - 178.

41 Locke 449.

42 The term "actual memory" is intended to distinguish those memories which are genuine in the sense that X has an actual memory of an experience which he himself had. This meets the counterexample that, since I remember an experience which happened to my brother because he related this to me, we must therefore be the same person. Such memories will be called "memories" simpliciter. The category of memories simpliciter is broader than that of actual memories and includes the latter within it. That is, actual memories are memories which are also genuine.
We shall consider objections to this theory in the course of examining the subsequent development of theories of psychological continuity in the following section.

2. Qualitative psychological continuity:

The most obvious flaw in Locke's conception is that it is logically impossible for someone to forget everything of his past and remain the same person. But there is a simple response that can be made. We simply alter the claim to read as follows:

\[ P_m \text{ at } T_m \text{ is the same person as } P_1 \text{ at } T_1 \text{ iff } P_1 \text{ and } P_m \text{ are both persons and } P_m \text{ contains or could (contingently) contain an actual memory of an experience had by } P_1. \]

But this formulation remains vulnerable to Reid's "Brave Officer Paradox". The General as a young boy was caught stealing apples and was flogged. Then later, as a young officer, he was decorated for his bravery in battle. Now the old general can remember his decoration as a young man but he can no longer remember the flogging as a young boy. But the young officer could. So, the paradox is that the young officer is the same person as the young boy and the old general is the same person as the young officer, but the general is not the same person as the young boy. Ordinarily we suppose that identity is a transitive notion such that, if \( A = B \) and \( B = C \), then \( A = C \). Reid's counterexample shows that the above identity relation for persons is not transitive.
Quinton was the first to formulate a version to account for Reid’s paradox by constructing the ancestral relation of the above.*" We divide a person’s life into episodes, each of which is an instantaneous time-slice. Then, we can say that $P_x$ is the same person as $P_s$ iff there is a sequence of such episodes such that $P_1$ is the first and $P_n$ is the last, and each contains, or could contain, a memory of an experience contained in the one immediately previous to it. But Grice countered with the "Senile General" paradox. This case is just like the one before except that the senile old general remembers the flogging as a young boy but he cannot remember the decoration as an officer nor anything since that time. Here the old general is the same person as the boy and the boy is the same person as the young officer, but the officer and the general are not the same person.

Grice’s solution was to disjoin the ancestral relation above with its converse and to add the relations of resemblance, in terms of personality, behavioural dispositions and so forth. He substitutes the term "total temporary state" (hereafter: t.t.s.) to designate this broader notion of resemblance that is intended to include all facets of our psychological life. Replacing our talk of episodes then, for talk of these t.t.s., we have the following:

*" This theory is set out in Quinton’s article, "The Soul", which is contained in the anthology edited by John Perry, *Personal Identity*, (Berkeley: U. of C. Press, 1975) 53 - 72.

44
P\textsubscript{m} is the same person as P\textsubscript{1} iff there is a sequence of t.t.s.'m linking the two such that each t.t.s. in the sequence contains, or could (contingently) contain, a memory of an experience in the t.t.s. immediately previous to it or vice versa and/or each t.t.s. in the sequence is linked by the relation of resemblance to its neighbors.

Ingenious as this example is in meeting counterexamples, it remains vulnerable to the objection originally raised by Bishop Butler to the effect that all such theories are circular, in so much as the notion of memory presupposes personal identity. Parfit has an ingenious reply to this however, which will be considered when discussing Parfit's reductionism in Chapter 8. Further, as argued by Flew, this definition is simultaneously too stringent and too lenient. One can imagine examples where we should wish to say that P\textsubscript{1} is the same person as P\textsubscript{m} but the definition would rule otherwise or vice versa. For example, consider the case of a man stricken with complete amnesia such that he can remember only how to speak, walk and so forth. Let us add the further detail that his ailment allows him to experience only the specious present. He retains no short term memories. In such a case, we would wish to say that he was still the same man, yet the theory claims otherwise. Conversely, imagine the case of a man who comes to resemble me in all psychological respects, while remaining physically quite distinct. We should not wish to say that he was the same person as myself, yet the theory claims that he is. A similar example, used by Parfit, is one where I have an identical replica made of myself. On Grice's theory, we
would be the same person, despite our intuitions on the matter. But let us consider theories that rely on the continuity of physical characteristics to see if they fare any better.

§2

We will examine the arguments for the priority of bodily continuity as presented by Bernard Williams in, "The Self and the Future"^{3} in section 1 before moving on to consider combined continuity theories in section 2.

1. The basic theory:

William's argument for the priority of bodily continuity is presumed to follow from these claims. Firstly, in so much as we are concerned for our own future, it can be shown that this concern must follow the body, since it is the normal cause of what we take ourselves to be. That is, the continuity of our mental lives is causally dependent on the continued existence of the same brain. Secondly, it can be shown that this causal dependence is violated by thought experiments which imagine cases of bodily exchange. Specifically, it can be shown that such cases assume that there is some critical level of change such that, before this level one is the same person and after this level, one becomes a numerically different person. But this supposition cannot be verified since, for one thing we could never discover, by any empirical means at our disposal, what this critical limit

^{3} Contained in the Perry anthology.
could be. And, for another thing, it can be shown that none of the levels of change introduce anything which, from the perspective of the victim, can be seen as relevant to his concerns for his own future.

Let us move on to consider the details of Williams' argument then. He presents his case by first having us consider the case of two men, call them A and B, who are trapped by a diabolical scientist who intends to exchange their bodies. Now, Williams first cautions us that it is difficult to describe this situation at all in a way that is not question-begging. We can only say that, after the process is completed, the utterances coming from each are expressive of the memories previously had by the other and similarly for other behaviours, such as movements, to the extent that these are expressive of character. But there is a further problem with the notion of memory, as it is normally construed. For it is a necessary condition of our being seriously prepared to say that we are now confronted with A, in the body of B, that we should take the utterances coming from that body as genuinely expressive of memories of A's past.

But memory is a causal notion; and as we actually use it, it seems a necessary condition on X's present knowledge of X's earlier experiences constituting memory of those experiences that the causal chain linking the experiences and the knowledge should not run outside X's body. Hence if utterances coming from a given body are to be taken as expressive of memories of the experiences of B, there should be some suitable causal link between the appropriate
state of that body and the original happening of those experiences to B.\(^\text{15}\)

Let us say then, that the method utilized is to extract the information from each brain and store this temporarily in a holding device, from which it is then transmitted to the other brain. We shall assume, for the time-being, that this constitutes an appropriate causal link and accept this as an accurate description of the event as exchanging bodies in non question-begging terms.

Now a further detail is added to the effect that the scientist tells each individual that, after the operation, one of them will be tortured and the other will be given $100,000. Each is then asked which body he would prefer to receive the torture and which the $100,000. If they accept the description of the operation as exchanging bodies, then it would be reasonable to suppose that A should say that he wished for the B-body person to receive the money and the A-body person the torture. Similarly, B should reply that he would like the A-body person to receive the money and the B-body person the torture. If the scientist should then reveal that he intends to give the money to the A-body person and to torture the B-body person, then A can rightly say that that is not the outcome he chose and B can rightly claim that that is the outcome he chose. We can further suppose that,

\[^{15}\] Williams 180.
after the experiment, the A-body person will express satisfaction with the results, while the B-body person will complain.

These observations seem to support the conclusion that this thought experiment does indeed describe a situation we should call an exchange of bodies. But now let us consider a slightly different case. In this case I am the prisoner, and I am told that I will be tortured tomorrow. Clearly, I have good reason to be concerned for my future welfare. But, I am also told that, shortly before the torture, my brain will be tampered with in such a way that I will remember nothing. Surely, argues Williams, I have just as much reason as before to be concerned since it will still be me who will suffer. I can easily imagine a case where I am in a serious car accident and suffer amnesia and then wake up in great agony. The mere fact that I would have also suffered amnesia is not sufficient reason to make me no longer fear the prospect of pain. In fact, argues Williams, it should cause me greater concern since I will also suffer psychological damage.

What if he now adds the detail that, in addition to complete amnesia of the details of my own life, I shall have a completely fabricated set of memories installed? Surely this would not cheer me up either for I can just as easily imagine a case where, previous to my accident, I go mad and come to believe that I am actually Napoleon, complete with apparent memories of my life as
the diminutive dictator. Again I have no reason to suppose that this should lessen my concern for it will still be me who shall suffer, despite my delusions.

In the last case, I am further told that the fabricated life history will be an actual replica of the life history of some actual, living person. We will be psychologically identical up to the time of the operation. Again, I have no reason to be less concerned for my welfare, for what should it matter to me that, previous to being tortured, I shall come to have the memories of someone else? Indeed, this should add to my horror.

Now Williams argues that we have, in the last case, redescribed the same events which we previously decided was a case of exchanging bodies. There are only two differences here which Williams dismisses as irrelevant. Firstly, in the second version, the torture is throughout represented as going to happen to me. This is irrelevant, according to Williams, on the grounds that, as long as I can follow what is being told to me, and if, on further reflection, I decide that I have good grounds for concern, then that is all that matters, "I could consider that behind my fears lies some such principle as this: that my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time, with the platitudinous exception of those psychological states which in themselves exclude experiencing pain, notably (if it is a
psychological state) unconsciousness". Secondly, in the second version, there is no essential reference made to any other person except in the incidental role of being, in some sense, the origin of my fabricated life history. This detail is also irrelevant though since my knowledge of any other person cannot in any way change the facts of my future; I will still have my psyche tampered with and then be tortured and, surely, nothing else can matter to me,

My selfish concern is to be told what is going to happen to me and now I know: torture, preceded by changes of character, brain operations, changes in impressions of the past. The knowledge that one other person, or none, or many will be similarly treated may affect me in other ways, of sympathy, greater horror at the power of this tyrant, and so forth; but surely it cannot affect my expectations of torture?" "

The gist of Williams argument then is this. There is a normal causal antecedent of our experiences, memories and so forth, and this is the brain. Further, we can see from examining the case which supposedly describes an exchange of bodies that it is question-begging. If suitably redescribed, the case shows that we must suppose that, somewhere along the continuum of changes perpetrated on the victim, we must reach some critical limit such that, further changes result in a different person. This supposition is counterintuitive according to Williams and,

" Williams 187.
" Williams 190.
furthermore, we have no means of determining what this limit could be. Consider again the continuum of cases described in the second version,

(i) A is subjected to an operation which causes total amnesia;
(ii) amnesia is produced in A, and other interference leads to certain changes in his character;
(iii) changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
(iv) the same as (iii) except that both the character traits and the memory impressions are designed to be appropriate to another actual person, B;
(v) the same as (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;
(vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction.\footnote{Williams 190.}

The case of (i) should be straightforward. Clearly, A has good reason, in this case, to be concerned for the promised torture. Similarly, his concern seems to extend quite naturally to both (ii) and (iii), since neither of these introduces anything new in principle. In fact, it would seem that A should have good reason to be even more concerned in these cases since he will not only be physically tortured, but shall be psychologically damaged as well. Although (iv) introduces another person, B, he does not seem to be introduced in any essential way. That is, from A's perspective, the existence of B can have
no material implications for his own welfare. Nor, in this case, is the causal condition for memories being actual memories satisfied. And even though this causal condition is perhaps satisfied in (v), it still seems reasonable to suppose that this would make no material difference to A's concern for his own welfare. All that can be said is that we now have an actual model for the memories induced in A, which is also their cause. Further, if A's concerns depend upon the issue of whether or not he actually is B, then this is clearly not satisfied in this case, since we have another individual who is undisputably B, in the normal sense.

We can only suppose then that the change occurs in (vi), yet there seems to be no good reason to suppose that this should be the case since,

if A's original fears could reach through the expected changes in (v), as they did in (iv) and (iii), then certainly they can reach through in (vi). Indeed, from the point of A's expectations and fears, there is less difference between (vi) and (v) than there is between (v) and (iv) or between (iv) and (iii). In those transitions, there were at least differences - though we could not see that they were really relevant differences - in the content and cause of what happened to him; in the present case there is absolutely no difference at all in what happens to him, the only difference being in what happens to someone else."^{49}

\(^{49}\) Williams 192.
Williams' claim then, is that continuity of the body must take precedence because it is the continued existence of the numerically same brain which is the normal cause of our memories, character traits and so forth. While his arguments have been persuasive, it is again a fairly easy matter to construct counterexamples. For instance, imagine the case of a man who steps into a machine which begins replacing all the cells of his body while he remains conscious. Upon completion, he is constructed of entirely new matter, yet there has been complete continuity, both psychologically and physically (in the sense that there has been continuous existence of a cause of his psychological characteristics). Now it may be objected that this is just the limiting case of the normal replacement of cells that occurs throughout our lives, but this is not so, for the cells of our central nervous system are never replaced. We have the same (i.e.: numerically identical) brain cells when we die as we were born with; they are never replaced and they never regenerate. Given this, what are we to say about this person, regarding the matter of his personal survival? If Williams' objections to the psychological criteria are correct then, despite his being psychologically continuous with his former self, he cannot have survived as the same person. On the other hand, if our counterexample works against Williams' arguments for the priority of bodily continuity, then it again seems that he has not survived as the same person, despite the qualitative physical similarity.
Perhaps the solution is to construct a theory that includes both the criteria of physical and psychological continuity. Let us consider such theories.

2. Combined continuity theories:

Since I am not directly familiar with anyone holding such a theory, I will, following Swinburne, discuss them in the most general terms as a theoretical possibility. Firstly, the motivation to such a combination theory arises from the following dilemma. Theories of psychological continuity rule it out as a logical possibility that a man should lose his memory and character and yet remain the same person, yet we ordinarily suppose this to be possible. On the other hand, theories of physical continuity rule it out as logically impossible that a man should move from one spatiotemporal location to another without tracing a continuous path between the two. Yet it is at least a logical possibility that a man should do so. For example, imagine the case of a man stepping into a teleporter, whereupon he is beamed to a remote location without passing through the intervening space intact as a human organism.

The obvious solution for an empirical theorist is to construct a compromise theory that combines the criteria of physical and psychological continuity as follows.

\[ P_2 \text{ at time } T_2 \text{ is the same person as } P_1 \text{ at time } T_1 \text{ iff both are persons and there is a continuous link of episodes such that, either:
}\]

a) each episode contains, or could
(contingently) contain, a memory of an experience in the episode immediately prior to it or vice versa, or; b) at each episode there is a body (or functioning brain) that is qualitatively similar to its neighbors.

It might be objected that there could be two or more persons who could satisfy such a criterion. In such a case a clause could be added to the effect that, "[N]either person at T₂ is the same person as P₁ at T₁, or that the person who satisfies the bodily criterion is the same person as P₁; or the theory will provide some other solution. Or the theory may provide some other detailed account of how such criteria as bodily continuity, similarity of memory or character are to be understood and balanced against each other." 49

Now there is an uncomfortable feeling that emerges from all this. We have begun from the premise that the characteristics which seem to be definitive of persons are also the key to what constitutes our personal identity, yet it has been notoriously difficult to construct a plausible and coherent account along these lines. Even the combination theory must, inevitably, rely on ad hoc restrictions to be able to meet obvious counterexamples, giving the entire edifice an air of contrivance. But, more damaging still is the problem of open-texture which was mentioned in Chapter 1 and is taken up again in the next section.

49 Swinburne 233.
I indicated in the introduction to this chapter that all empirical theories may be considered together because they are all guilty of the same flaw. All are based on the mistaken assumption that it is possible to provide a definition of personal identity, in terms of the continuity of observable characteristics, such that, if given all the relevant factual data, we shall be able to decide unambiguously, in all actual or imaginable cases, whether or not \( P_x \) is the same person as \( P_s \). To understand why this assumption is mistaken, consider the following.

Firstly, there is the problem of ambiguity. There will always be marginal cases where we simply do not know where to draw the line. Most referential terms of natural language are vague to the extent that they have a core meaning that is clearly enough understood but the periphery is not clearly demarcated. Such is the case with most of the crucial terms involved in the topic of personal identity such as "person", "man", "self" and so forth. Secondly, there is the problem, alluded to in Chapter 1, of the open-texturedness of many of our concepts. To quote Flew,

\[ \text{[S]ince our ordinary language, and the concepts of ordinary language, have been evolved or introduced to deal with the situations which are ordinarily met with, and not with the extraordinary, we may reasonably expect some failures of adaptation when new and unexpected situations arise. ...The old} \]

51 I draw on both Swinburne and Flew for this discussion.
conceptual machinery breaks down. The old terminological tools fail to cope with the new tasks.\textsuperscript{51}

The difference between these two is best illustrated by the use of an example. The term "man" can have several meanings depending on context and purpose. It might refer generally to the species or, if used in a court of law, it might be intended to distinguish a legally responsible adult from a boy. This is an example of ambiguity. But now recall the case of Locke's where he admits that, were the soul of Heliogabalus to be reincarnated into a hog, we should be reluctant to call him a man. But what if the pig were to behave precisely like Heliogabalus in every respect; he is even able to answer all questions put to him and so forth. Now what should we say? This latter is an example of open-texturedness. In the case of ambiguity, we can stipulate arbitrarily that a legally responsible "man" shall be an individual 18 years of age or more, but such arbitrary stipulation is not always appropriate nor even possible. Further, stipulation will not solve the problem of open-texture. We cannot provide any descriptive definition that will, a priori, take into account all possible puzzle cases such as the one of Heliogabalus and his hog.

Further, no such descriptive definition is possible because they all mistake the criteria for answering the question of

\textsuperscript{51} Flew 174.
whether $P_1$ is the same person as $P_{10}$ for what it is that constitutes personal identity; they mistake epistemology for metaphysics. Even if the concept were not open-textured, this problem would remain. While the evidence of psychological or bodily continuity may be the best we have to answer the above question, it will not answer the metaphysical question of what personal identity consists in,

Wherein does the identity of persons consist? The identity does not consist solely in the continuity of one or more observable characteristics, for empiricist theories took all of these into account. ... The only alternative is to say that personal identity is something ultimate. It is unanalyzable into conjunctions or disjunctions of observable properties.\textsuperscript{59}

So, our historical survey has revealed a dilemma. On the one hand, we may try to explain personal identity in terms of something ultimate, such as a pure ego theory of the sort considered in Chapter 2. Thus far, all such attempts have failed. On the other hand, there are good intuitive reasons to think that the key lies in the continuity of certain observable characteristics which we ordinarily take to be the distinctive traits of persons, yet this attempt has also failed. In the next section we will examine Hume's position. There it will be argued that he took the empiricist approach to its logical conclusion only to discover that our idea of personal identity is a fiction that cannot be justified.

\textsuperscript{59} Swinburne 240.
Hume begins by attacking Locke's principium individuationis as follows. A single perception of any object cannot, in itself, convey the idea of identity for this would imply that the proposition, "this object is self-identical", is analytic, which it surely isn't. The idea that is really conveyed in such perceptions is that of unity. Nor can we ground the idea of identity in the perceptions of a series of objects that resemble each other, for this conveys the idea of diversity. Now for Hume, the perception of the same object at different times is equivalent to such a series of perceptions of closely related objects. This follows from his phenomenalist theory of perception to the effect that what we directly perceive are our perceptions and not the objects themselves. Further, each perception exists as a logically distinct entity. In this way it is logically impossible that we should ever be able to simultaneously compare the distinct perceptions of a single object at different times to ensure that it was, in fact, self-identical.

From what impression could we obtain the idea of identity then? According to Hume, there is no such impression. Rather, the idea arises from the combination of the ideas of unity and time or, more properly, duration. That is, duration implies succession and when we combine this with the notion of unity, we conceive of the same object persisting unchanged. "Thus the principle of individuation is nothing but the invarableness or
uninterruptedness of any object, thro' a suppos'd variation of
time, by which the mind can trace it in the various periods of
its existence, without any break of the view, and without being
obliged to form the idea of multiplicity or number".54

But, he argues we have no such idea of the self, for the
self is not grounded in any one impression but is the entity to
which our perceptions are supposed to have reference. That is,
selves "have" perceptions. Further, if there were any one
impression that did give rise to the idea of the self as it is
commonly understood, then it would have to continue unchanged
throughout our lives. Hume here is maintaining that there must be
strict numerical identity of the self throughout our lives to
accurately reflect its usage in ordinary language. And he
emphatically maintains that there is no one impression that does
remain unchanged in this way. Rather, Hume likens the mind to a
theatre where the perceptions pass across the stage in rapid
succession and persons are nothing more than bundles of
perceptions,

I may venture to affirm of the rest of
mankind, that they are nothing but a bundle
or collection of different perceptions, which
succeed each other with an inconceivable
rapidity, and are in perpetual flux and
movement. .... There is properly no
simplicity in it at one time, nor identity at
different, whatever natural propension we may

54 Hume, Treatise 201.
have to imagine that simplicity and identity.

While it is one thing to show that our notion of the self is ill-founded, it is another to then explain where we obtain this idea. Hume's explanation is psychological, as hinted in the above quotation where he refers to our "natural propension" to form such an idea. Specifically, it results from a confusion of the ideas of: 1) identity; an object which persists without change and 2) diversity; that of a succession of related objects. The tendency to confuse the two results from the fact that a succession of similar objects is subjectively indistinguishable from imagining an object persisting unchanged. That is, although the mind is actually constituted of a succession of impressions that are related by causation and resemblance, we inaccurately substitute the idea of identity in its place. There are other factors which contribute to this propensity, for instance, the slow rate of change in persons from day to day. The fiction of the "self" has arisen specifically because philosophers have assumed that there must be some such entity to be the referent of our impressions, attributes and so forth.

Now while Hume derides the notion of a self as some sort of self-subsistent entity, he does have a theory of what persons actually are. This was hinted at above when mentioning his view

---

Hume, Treatise 252 - 253.
of the mind as a succession of impressions that are related by causation and resemblance. To return to the analogy of the mind as a theatre then, the impressions of any one bundle succeed each other according to these relations. Consequently, we may say that distinct impressions belong to the same person if they are linked by the relations of causation of resemblance. Of course Hume would never admit to this for he maintains that we do not persist as the same person. Rather, he offered this as what we have instead of personal identity.

There is a problem with Hume's analysis, for note that he claims that our propensity to attribute identity to persons is aided by the slow rate of change that can be observed. Further, he claims that persons are nothing more than bundles of perceptions, which seems to imply that one is nothing more than what is contained in one's stream of consciousness. So, the first claim seems to rely on observations of a person's physical characteristics, while the second implies that persons are purely mental entities. When we observe others we do not observe their minds but their bodies and these physical characteristics are what change gradually. And even when we introspect, we cannot observe any such gradual change if his theory of the mind as a kind of theatre is right. Rather we observe a rapid succession of constantly varying impressions. He was also wrong to insist on strict numerical identity in the case of persons and we shall discuss this further in Chapter 8. But he is right in so far as
he maintained that no empirical theory of personal identity could be justified. We shall also see that he was at least partly correct in characterizing persons as bundles of perceptions that are linked by the relations of causation and resemblance. In the next part of the manuscript we shall consider theories that rely on the distinction between persons and sentient beings. Such theories regard sentient beings as ultimate entities, but of a sort entirely different from Cartesian egos. Then, in Part 3 we will consider Parfit's reductionist theory, which takes Hume's gambit arguing that our standard notion of personal identity cannot be justified.
PART 2

SENTIENT BEING THEORIES
CHAPTER 4

Sentient Beings: A Further Fact

In this chapter it is my intention to introduce the basic distinction between persons and sentient beings and to explain why sentient beings are relevant in a discussion purportedly about "personal" survival.

§1

As a person one is also a sentient being. This fact, which is established below, is true of all persons. It is also true of many organisms which are not persons. Specifically, most animals are sentient beings but they will not, like ourselves, also be persons. It is even possible for some individuals to cease being persons at some point in their lives (as a result of illness, senility or whatever) while continuing to be sentient beings. It should be clear then that the concept of "sentient being" is of wider scope than that of "person" and is meant to include, in addition to the example of sentient animals already given, any possible inhabitants of other planets that would be sentient (of course, some of these might also qualify as persons).

\[\text{(Footnote: For our purposes it is uninteresting to decide exactly where, in the phylogenetic scale, to draw the line between organisms that are sentient and those that aren't. )}\]
What exactly is meant then by the term sentient being? At this point I define a sentient being as a "haver-of-experiences" or, equivalently, as a subject of experiences. To be a subject of experience, one must be conscious. Consciousness is here intended to mean the opposite of anaesthesia; to be aware. Now to be a person, one must, minimally, be conscious and, therefore, a subject of experiences. This establishes the claim made above that all persons are also sentient beings. For the moment though, let us ponder this notion of "havers-of-experiences". Firstly, it is intended to express the essential property of sentient beings. That is, all sentient beings are havers-of-experiences. Secondly, the experiences themselves need only be of the most rudimentary sort. There is no requirement that the experiences "had" by any sentient being must be of the sort which one might call refined in any way, for example the experience of aesthetic pleasure one might feel upon listening to Bach. In fact, if one were barely conscious and was experiencing only a dull throb of pain, that would be sufficient to be the sort of experience appropriate to a sentient being. That is why the concept of sentient being is of much wider scope than that of person. This also points to what it is that further distinguishes persons from all other sentient beings; it will be at least partly a matter of the quality of the experiences. This underscores the observation above that some individuals can cease to be persons at some point in their lives as a result of illness or whatever. Clearly, such cases will result from a decline in the quality of experiences "had" by that
individual. The example of the man who has declined to a state of bare consciousness and who experiences only a dull throb of pain is a case in point.

As a subject of experience then, we may say that a sentient being is a stream of consciousness where the occurrent experiences may be of even the most rudimentary sort. But where are all of these various streams of consciousness to be found? The reply is, in psychological space (hereafter p-space). All streams of consciousness then, occur in p-space and each individual stream of consciousness will occupy its' own particular location within p-space, known as its' own p-location. The picture that begins to emerge then is as follows. Sentient beings are conscious entities which, in virtue of their consciousness, are capable of being subjects of experience. As subjects of experience we are not concerned with the qualitative nature of the experiences themselves but only with the fact that experiences occur at all. Given this, we may say that sentient beings are really just streams of consciousness. In so much as sentient beings are just streams of consciousness, each sentient being is associated with one and only one p-location and all of these p-locations occur within p-space. Sentient beings may be distinguished within p-space then, by the fact that each occupies its' own particular p-location. Thus, p-locations serve to

[...]

I defer until later a discussion of the nature of p-space and the related notions that follow.
numerically individuate sentient beings within p-space. Further, regarding the content of these p-locations, each will be exhaustively described by the totality of experiences "had" by that particular sentient being or, equivalently, by a complete description of that particular sentient being's stream of consciousness.

§2

It is now time to consider the notions of p-space and p-location. These notions will here be defended in terms of certain analogies which are claimed to hold between these notions and those of physical space and time. The underlying argument is that, together, these two concepts (i.e.: physical space and time) allow us to numerically individuate all physical events occurring within physical space. That is, presuming Leibniz' "Principle of the Indiscernibility of Identicals" to be true, we can conclude that any two physical events which are qualitatively identical must necessarily occur at different locations if they occur simultaneously, or at different times if they occur at the same location. No other option is possible. Similarly, it will be argued that, given any two qualitatively identical mental events which occur within p-space, it can be concluded that they necessarily occur at different p-locations if they are simultaneous, or at different times if they share the same p-location. The following analogies then are presumed to hold with the notions of physical space and time:

69
(1) Qualitatively identical physical events can be numerically distinguished using the indexicals of location in physical space and/or location in time. Similarly, qualitatively identical mental events can be numerically individuated using the indexicals of location in p-space and location in time.

(2) Temporal location, taken together with either of the other two, is metaphysically adequate to numerically individuate all events in a purely non-qualitative way. That is, p-space and physical space share the indexical of location in time to numerically distinguish all events.

(3) Physical space and p-space endure. That is, each location in either physical space or p-space may host a potentially infinite number of distinct events (although by (2) each individual event could be distinguished from all others by the additional use of a temporal indexical).

Obviously there will also be certain disanalogies, which should not be surprising in light of the disanalogies that hold between the notions of physical space and time. These disanalogies include;

(1) Temporal movement is only possible in one direction, while movement in physical space is possible in many directions.

(2) There are only two directions in time but many in physical space.

(3) The rate of movement in physical space is variable while the rate of movement in time is not.

(4) We have some control over the rate of movement in physical space but none over the rate of movement in time.

In the same way, the following disanalogies hold between p-space on the one hand and both physical space and time on the other;
(1) p-space is discontinuous; there is no intervening topology between individual p-locations such that any mental event occurring at one can shift or "jump" to another.

(2) All p-locations occurring within p-space are equidistant from each other; each sentient being's experiences are equidistant from the experiences of every other sentient being.

Regarding the first disanalogy above, there is one necessary qualification. The exceptional case of the fusion of two distinct streams of consciousness would violate this desideratum, since we would have a case where the content of at least one stream of consciousness was now occurring at an entirely different p-location. For, in such a case, we must suppose that either, (1) both of the original streams of consciousness now occur at a different p-location or, (2) that one of the original streams now occurs together with the other stream at the p-location originally "had" by the other. On either alternative we must suppose that at least one of the streams of consciousness has leaped the discontinuous landscape of p-space to arrive at a different p-location. However, we must recognize this as a logical possibility for consider the following case of a child born without a corpus callosum which, according to Parfit in, Reasons and Persons, has actually been known to occur. Given

\[\text{Recall from Chapter 2 that the corpus callosum is a fibrous material which connects the two hemispheres of the brain. Here we are not considering severing this connection but, rather, the case of someone born without a corpus callosum to establish the logical possibility of fusion where there had never previously been a single stream of consciousness.}\]
such a case, let us imagine some future state of neuro-surgery which enables them to connect this individual’s hemispheres resulting in a single stream of consciousness where previously there had been two. Again our alternatives are to suppose that either one of the original streams now shares the p-location originally occupied by the other alone or that both now occur together at an entirely different p-location. Either of these alternatives forces us to acknowledge that the first disanalogy above can be violated though, at least to the extent that it is a logical possibility that a stream of consciousness can shift to another p-location. However, it is only in the case of such a fusion that the desideratum that nothing which happens at one p-location can "leap" to another will be violated. Under no other circumstances will this be possible.

The reason for this, which will be elaborated in the next chapter, is that streams of consciousness are caused by brains. As such then, the stream of consciousness associated with the brain which is its' cause will always occur at the same p-location precisely because the p-location itself is a concomitant cause of the same brain. Of course this view regards causal necessity as a contingent relation since it has already been admitted that it is a logical possibility that a stream of consciousness should shift to another p-location. This amounts to the claim that it is logically possible that consciousness should be transferred to a different brain and this, in turn,
necessitates the view that the connection between any brain and its associated stream of consciousness is a contingent one.

63

We must now explore further the relationship between sentient beings and p-locations. The claim is simply that, as a sentient being my stream of consciousness occurs at a particular p-location within p-space which, minimally, has the following properties. Firstly, trivially it is distinct from all other p-locations that occur in p-space. Secondly, it is temporally coextensive with my life as a sentient being; that is, one sentient being, one p-location. And thirdly, together with temporal location, it is metaphysically adequate to individuate all of my experiences. On this theory then, a sentient being is nothing more than a bundle of experiences (i.e.: a stream of consciousness) that all share the same p-location.

Again notice that no further claims are being made about the qualitative characteristics of the actual experiences (i.e.: token or occurrent) at any particular p-location. The claim is only that, if any experiences occur "here", then they belong to "this" sentient being. In so much as we are considering sentient beings, the qualitative nature of these experiences is irrelevant; they need only be of the most rudimentary sort to qualify as the sorts of experiences appropriate to sentient beings. This observation raises the question of the relevance of sentient beings in a
discussion purportedly about "personal" survival though and it is to this that we must now turn our attention.

Firstly, imagine the case of a prisoner of war who is subjected to certain highly sophisticated brainwashing techniques so that all of his personal memories have been obliterated and replaced with a complete and fabricated life history. Further, all of his personal beliefs and character traits have been radically altered. Upon meeting him, all of his family and close friends would probably agree that he was a very different person, not in the sense of numerically different but, rather, radically changed. Any theory which maintained (like those considered in Chapter 3) that personal identity was a matter of the continuity of certain psychological characteristics would be forced to conclude that he was indeed a numerically different person. As discussed above, this is an example of the open-texturedness of the concept of person. Since the continuity of psychological characteristics is a matter of their qualitative appropriateness across time, we find ourselves in this case forced to conclude that the victim is indeed a different person. This view would occasion difficulty for his family and friends who would continue to regard him as, in some sense, the same person. And I think most of would agree that there would be something unsatisfactory and even counterintuitive about the reply of the psychological continuity theorist.
Of course we have also seen that one may attempt to explain survival as the same person in terms of the continuity of the same brain (or at least a functioning part of the same brain). On this alternative we would claim that he was indeed the same person since he still had the same (i.e.: numerically identical) brain. But this is really begging the question to the extent that it entirely fails to address the issue of his complete psychological change. Additionally however, we may imagine a case like the following. Consider President Reagan, who begins to transform over a period of weeks until he comes to resemble Hitler to the extent that they are physically indistinguishable, even though he remains unchanged psychologically. I contend that we should not know what to say in such a case regarding the "personal identity" of Reagan; we would find ourselves unsure as to whether or not he had in fact survived as Ronald Reagan. But the physical theorist would reply that the continuity of Reagan's body is irrelevant to the question of his survival as the same person. What matters is that he retain the same (i.e.: numerically identical) brain. Since this is the case, it follows that, despite physical appearances, he is still Ronald Reagan. But now imagine that this change in his physical appearance is precipitated by the replacement of all the particles of his body such that, when he actually comes to resemble Hitler, none of his original cells remain. Now what might our physical theorist say? I think that he would be forced
to say that Reagan was now a different person despite the fact of his remaining psychologically unchanged.

We may even imagine a more dramatic example yet along the lines of Parfit’s combined spectrum case. In such a case we again imagine Reagan beginning to transform slowly over a period of weeks, where the changes are again presumed to coincide with the replacement of his own cells. Finally, after the process has been completed, Reagan resembles Jimmy Carter, both physically and psychologically and there are none of his original cells remaining. In such a case both the psychological theorist and the physical theorist would conclude that he was a different person.

The problem underscored by these examples, which was discussed in the last chapter, is that, in some contexts, the concept of person is open-textured; that is, the rules of our language do not always permit us to unambiguously decide whether or not someone has remained the same person. However, it is unambiguously decidable that he is the same sentient being. This is established by showing that one remains the same sentient being throughout one’s life. Again, this is taken to follow from the claim that retention of the numerically same brain is causally necessary and sufficient for the continued existence of

Combined spectrum cases will be discussed further in Part 3 when dealing with Parfit’s reductionist theory.
the same p-location and, concomitantly, of the same stream of
consciousness.

At this point one may object that, in the last two examples
where Reagan's body was gradually replaced by entirely new
material, we no longer have the same brain. But recall that, on
this theory, retention of the same brain is regarded as only
contingently necessary. What is actually significant is retention
of the cause of one's stream of consciousness and, given the
contingency of the association between cause and effect, it need
not be the numerically same cause. Indeed, this was the
significance of the observation above that the transfer of
consciousness is a logical possibility. Now since, in the cases
considered here, there existed continuously a cause of Reagan's
stream of consciousness, we can be sure that his stream of
consciousness and its associated p-location never ceased to
exist. Hence, this is really a case that is analogous to the
transfer of consciousness considered above.

So, we conclude that one's life as a sentient being subtends
the period inclusive of the occasion of one's first experience
(however rudimentary) and the occasion of one's last experience
(i.e.: one's death). Although I may cease to be a person at some
point in my life, for reasons already noted above, I will always
be a sentient being. And even though my character may change so
radically that it may no longer be clear whether or not I am the
same person, I will always be the same sentient being. In this way, the concept of sentient being remains distinct in many contexts where the concept of person is open-textured. These considerations will justify our concern with sentient beings in a discussion about personal survival, for if we cannot decide whether or not someone has remained the same person, then it follows that we cannot decide whether or not he has survived as the same person. But we may have good enough grounds for supposing that someone has remained the same sentient being and so this concept must be the fundamental notion of any discussion about personal survival.
CHAPTER 5

Constraints for a Theory of Personal Survival

The purpose for having a list of constraints for any metaphysical theory is to rule out, a priori, particular theories as implausible. To this end, the constraints are supposed to act like a sieve which "filter out" these implausible theories. It is essential then that our list of constraints should reflect, as accurately and as completely as possible, our best intuitions on the subject. In this chapter I will explicate a list of constraints which, I claim, must be satisfied by any theory that is to be commensurate with our intuitions regarding personal identity and personal survival. In §3 I present a discussion which is intended to convince the reader that, indeed, these constraints are the ones which are necessary for any theory that is to satisfy our intuitions on these matters. This discussion will be particularly crucial in Chapter 9 when we compare the sentient being theory developed in Chapter 7 with Derik Parfit's reductionist theory. There it will be argued that Parfit's theory leads to certain highly implausible results and that this can be explained in terms of that theory's inability to satisfy the constraints developed here. In §4 we will reconsider the theories discussed in Chapters 2 and 3 in terms of these constraints.
§1

There are five constraints of central importance that will be listed here in addition to several other constraints that are of somewhat less importance. These latter might be called provisional constraints or, perhaps, observations. The order in which all of these are listed is of no particular significance.

THE CONTINGENT BRAIN CONSTRAINT:

Retention of the same brain, or at least a functioning part of it, is a causally necessary and sufficient condition for survival as the same sentient being. Equivalently, we may say that retention of the same brain is a contingently necessary and sufficient condition for survival as the same sentient being. This follows from the supposition mentioned in the last chapter that the causal relation is conceived of as contingent because, as will be argued later, it is a logical possibility that consciousness could be transferred to a different brain. Given this, I make the following claims. Firstly, retention of the same brain is both a necessary and a sufficient condition for survival as the same sentient being because the stream of consciousness which is the sentient being and the p-location at which the stream occurs are both concomitant effects of the brain. Yet, to be at all plausible, it must be maintained that

Whenever reference is made to retention of the same brain in the context of a sentient being theory in this thesis, it is to be taken as elliptical for the phrase, "retention of the same brain or a functioning part of it in a state capable of supporting consciousness".

80
this requirement is contingent. To see the force of this, recall from Chapter 2 that Descartes utilized an ingenious argument to show that it is possible that consciousness could exist independently of physical substance. He thought that he had established his radical dualism by showing that, since it was possible to conceive of the two separated in this way without contradiction, it therefore followed that they are in fact distinct. As mentioned above (i.e.: Chapter 2), I have found no evidence to establish that Descartes distinguished between logical and physical possibility and I conjecture that, therefore, Descartes' mistake was to think that his conceivability claim established, as a matter of fact, that minds exist independently of bodies in the actual world. That is, he mistakenly took his "proof" of logical possibility to establish the truth of radical dualism as a fact about the actual world. In fact, all it shows is that it is a category mistake to propose that the two are inseparable in this way. In this world it is a matter of empirical fact that consciousness is always associated with brains. But this connection which obtains in the actual world as a matter of empirical fact is a contingent one, because there is a possible world in which there are streams of consciousness that are not associated with brains.

There might still be some confusion regarding the nature of the relationship that is being implied here between brains on the one hand and streams of consciousness and p-locations on the
other. The claim being made here is that, although retention of the same brain is causally necessary and sufficient for survival as the same sentient being, it is in no way constitutive of what it is to be a sentient being. The requirement for retention of the same brain is a fact of the world because of certain basic causal laws which hold. To say that it is not constitutive of what it is to be a sentient being is simply to claim that, qua sentient being, we are only concerned with the fact that there is a stream of consciousness and not with what might cause that stream of consciousness.

Let us now clarify exactly what is meant here by the "same" brain. The claim here is that, at each time \( T_i \) in a sentient being's life, the brain must be "numerically continuous" with the brain at any time \( T_0 \) previous to \( T_i \), where the points \( T_0 \) and \( T_i \) may be joined by a series of any number of points of time \( T_n \), each of which is joined to the one next to it by the same relation of numerical continuity. The sense of continuity intended here is that it should be the same brain or at least a functioning part of it in a state capable of supporting consciousness. Notice also that only numerical continuity is required here and not numerical identity. For consider the case where the brain is severed in two (as in the case of split-brain treatments considered in Chapter 2). In such a case, it is

\[ \text{\textsuperscript{61}} \] A similar notion of "numerical continuity" was developed by Sikora in the manuscript previously cited.
obvious that numerical identity is not preserved for it is impossible that the two resultant half-brains should be numerically identical with the original single brain. That is, the brain at each time $T_n$ need not be the self-same identical brain, but only numerically continuous with the original brain. The significance of numerical continuity will become evident when discussing Parfit’s interpretation of the significance of split-brain results in Chapter 9. Parfit claims that since the results of split-brain research show clearly that numerical identity is irrelevant to personal survival, it follows that being survived by an identical replica would be about as good as ordinary survival. I will argue that, in the case of a replica, not even numerical continuity is preserved for, as a cause of the stream of consciousness associated with it, the replica’s brain is certainly not numerically continuous with my own in the sense required.

Notice further, though, that this requirement is claimed to hold only for survival as the same sentient being and not for survival as the same person. Were the claim to be made for remaining the same person there could be a conflict as a result of certain other considerations that are partly constitutive of our lives as persons (i.e.: the continuity of qualitatively appropriate psychological characteristics). For example, if at time $T_n$, we are confronted with two persons such that one has a brain which is numerically continuous with a person at a previous
time $T_1$ but who shares no psychological characteristics in common with that previous person, while the second person shares many psychological characteristics but does not have a brain that is numerically continuous with the previous person, then we should not know what to say. This is another case which demonstrates the open-texture of the concept of person which was discussed in Chapter 3. Clearly, in such a case we can see that retention of the same brain, in the sense defined above, is not sufficient for remaining the same person, although it is necessary since to be the same person one must also be the same sentient being. In the case of the simpler notion of sentient being though, retention of the same brain, or a functioning part of it in a state capable of supporting consciousness, is necessary and sufficient for survival. Since the qualitative characteristics of one's stream of consciousness are irrelevant to the question of survival as the same sentient being, and since we know that continuity of the same brain is sufficient for retention of the same p-location, we know that retention of the same brain is also sufficient for remaining the same sentient being.

**THE TRANSFER CONSTRAINT:**

The intuition which underlies this constraint may not be immediately evident but I shall establish that the consequences of denying this constraint would indeed be counterintuitive. The constraint is that transfer of consciousness to a different brain should not be ruled out by our theory as logically impossible.
That is, although it is probably physically impossible that such a transfer should ever occur, it is a logical possibility. In the case of such a transfer, both the stream of consciousness and the associated p-location must be transferred together since both are concomitant effects of the same underlying cause. Also, if we consider the alternatives, we shall see that this is indeed the case. Firstly, we may suppose that the stream of consciousness is transferred without the associated p-location. Now recall from the discussion of William’s in Chapter 3 that, to accurately describe the scenario as a case of transfer of consciousness, we must imagine that the person who previously occupied the A-body now occupies the B-body. But for this to be the case, he must also be the same sentient being otherwise, ipso facto, he cannot be the same person. And, to be the same sentient being, we must also have the same p-location for this is what numerically individuates sentient beings (i.e.:sentient beings are distinguished by their unique location in p-space). So, in the case where we imagine the stream of consciousness is transferred without the associated p-location, we can no longer have the same sentient being. Properly speaking then, we should not say that the stream of consciousness has been transferred at all but, rather, that we have altered the consciousness of the B-body person such that it has come to resemble the consciousness of the A-body person.
Another alternative is to suppose that we have transferred the p-location without the associated stream of consciousness. On this alternative it follows by default that we have not transferred the stream of consciousness though and so the result is not what we are seeking to establish. So, it should be seen that the only accurate description of an event we should call a transfer of consciousness must involve the transfer of both the stream of consciousness itself and its' associated p-location.

Now for the promised explanation of the intuition underlying this constraint: the claim is that such a transfer is conceivable (i.e.:it can be thought of without contradiction; it is a coherent, albeit physically unrealizable, situation) and so it would be a category mistake to construe consciousness as a sort of entity which could not be so transferred. Like the argument above which showed that consciousness can, in some possible world, exist independently of brains, this argument establishes that transfer of consciousness is similarly true in some possible worlds. But we have also seen that, given such a transfer, both the stream of consciousness and the attendant p-location are necessarily transferred together. Further we can see that consciousness cannot "be" a brain, since to speak of the transfer of one brain to another would be absurd.
THE PHENOMENOLOGICAL CONSTRAINT:

All that matters in survival as the same sentient being is what is contained in one's stream of consciousness. That is, despite the fact of the causal dependence on the brain for survival as the same sentient being, this is in no way constitutive of what we mean by such survival. Rather, our lives as sentient beings are entirely constituted by the phenomenological content of our streams of consciousness. The proof of this lies in the fact, already discussed, that the causal dependence on the brain is contingent only since we can coherently imagine a possible world in which our streams of consciousness exist exactly as they do in this world but are not associated with brains (i.e.: a phenomenalist world). It is even conceivable that such a world would not be deterministic at all and, hence, no causal explanations are necessary at all or even possible.

Now someone might object that, while this is true, even in such a phenomenalist world our streams of consciousness must necessarily be associated with Cartesian egos or souls or some such thing. The claim here then, is that the phenomenological content of one's stream of consciousness is not entirely constitutive of one's life as a sentient being, but only partly constitutive of it; there is the further fact of one's soul or whatever to be accounted for. This objection fails though for the same reasons as above. That is, even granting the possibility of
some "entity" of some sort which could do the job a Cartesian theorist wants, without the attendant difficulties outlined in Chapter 2, it is still possible to coherently imagine a possible world in which there are streams of consciousness but no associated entity of any sort. Now I think it may be that this objection is grounded on the intuition that, in such a world, there would be nothing to make the stream of consciousness "mine". To this I reply that it would be mine precisely because it would occur at my p-location. Now it might be further objected that there would be no way to verify this, for it is a fact that, in this world, we are able to verify such things because we can, in most cases, identify the person's body. In a world where there were no bodies, we could not identify any stream of consciousness nor could we be assured that any experience, despite its qualitative appropriateness, belonged to any particular sentient being. To this I reply that this objection is guilty of the same confusion of metaphysics with epistemology which was outlined in Chapter 1; regardless of whether or not we could verify the fact, it would still be true that, in such a world, "this" stream of consciousness would be mine because it occurs "here", where "here" refers to my p-location. And, from the fact, ex hypothesi, that one cannot coherently imagine a possible world in which there are sentient beings but no streams of consciousness, it follows that the phenomenological content of one's stream of consciousness is at least partly constitutive of one's life as a sentient being. Further, for reasons already discussed, it is
claimed here that this is entirely constitutive of one's life as a sentient being.

THE DIVISION CONSTRAINT:

An account of survival as the same sentient being should be compatible with the fact that there may be division of consciousness and that this would be about as good as normal survival. It will be recalled that the empirical fact of such division of consciousness was discussed in Chapter 2 and it is further claimed here that this would be about as good as ordinary survival since both resultant streams of consciousness would be numerically continuous with the original stream.

THE TEMPORAL CONSTRAINT:

The qualitative nature of one's experiences often provide evidence for the claim that one is the same sentient being as one was a year before or ten years before or whatever, but these qualitative considerations can do nothing more than provide such evidence. Specifically, they are in no way constitutive of our lives as sentient beings for recall from the previous chapter that, in so much as we are concerned with sentient beings, we need only establish the fact that there is a stream of consciousness; we are only concerned with the brute fact of the existence of the stream itself. The qualitative considerations are irrelevant to the extent that even the most rudimentary sorts of experiences will be appropriate for sentient beings. The force
of this point will be best exemplified by considering the following. Imagine the case of a man coming out of a coma just before dying. During the last few moments of his life, he is not at all lucid and, in fact, he babbles away in gibberish and displays none of the psychological characteristics that we typically regard as qualitatively appropriate for persons. Now on most criteria of "personhood" we would say that, although the man was conscious (in the sense opposed to anaesthesia), he was not a person and, ipso facto, he was not the same person he was previous to his decline into the coma. In this way we can see how the qualitative nature of our experiences is at least partly determinative of the ownership, qua person, of those experiences.

Now consider further the same man, but this time his period of gibberish behaviour is followed by a brief period of lucidity before his death. In fact the man is able to recognize his family and the doctor and displays all the characteristic personality traits his family expect of him. Now, what are we to say about this in light of the claim above that the qualitative nature of our experiences is in no way determinative of our lives as sentient beings? For clearly it seems that, in such a case, the occurrence of those lucid experiences just before the man's death determined the ownership of the gibberish experiences. Specifically, they made those gibberish experiences belong to the same person that existed before the decline into unconsciousness, for we have already agreed that, in the absence of these lucid
experiences, the gibberish experiences would not belong to any person at all and, hence, could not possibly belong to the same person.

The conclusion to emerge from this is that, although the qualitative nature of our experiences is at least partly determinative of our lives as persons it is in no way similarly determinative of our lives as sentient beings. In terms of temporality, this means that the occurrence of experiences is all that is significant in determining the ownership, qua sentient being, of any experiences that have occurred previous to them. Although qualitative considerations may provide evidence for the claim that this is the same person and, hence, the same sentient being, they do no more than that. Alternatively, we may claim that the man in the example above is the same sentient being as he was previous to his decline in virtue of something intrinsic to the first experience following recovery of consciousness and not because of any extrinsic, qualitative considerations. To claim otherwise would force us to hold that, the experiences of the man coming out of a coma only to reach a state of bare sentience would constitute a tiny life of their own. Conversely, if these experiences were followed by a state of full

---

This constraint presupposes that one's life as a sentient being (and also as a person) is discontinuous. Specifically, there are gaps in one's life as a sentient being that correspond to the periods of unconsciousness in our lives. This topic of gaps in the lives of sentient beings will be discussed at length in Chapters 6 and 7.
consciousness, then we would say that the experiences "had" during this period of bare sentience did belong to the same person (and, ipso facto, the same sentient being) as before the decline into the coma.

In addition to the constraints listed above there are the following "observations" as I shall call them which, although not as important as the actual constraints, are still significant enough to be mentioned.

Firstly, survival as the same sentient being is, necessarily, an all-or-nothing matter. That is, identity as the same sentient being is a matter of "strict identity" and not of "loose identity" as in the case of persons. Because survival as the same person is at least partly determined by the continuity of certain psychological characteristics, where this continuity is established by the qualitative appropriateness of the later experiences, it is a loose kind of identity. That is, this kind of continuity can be shown to hold to varying degrees. This is precisely the insight which motivated Derek Parfit to formulate his reductionist theory of personal identity (this will be discussed at length in Part 3). This is also, as already discussed, the reason why the notion of personhood is open-textured. However, in the case of sentient beings, one either is the same sentient being or one is not; the continuity implied
here is a matter of strict numerical identity. This observation is a consequence of the Phenomenological Constraint and the Temporal Constraint. That is, to the extent that we are concerned to establish that someone has survived as the same sentient being, we need only establish that the same stream of consciousness has survived (i.e., that experiences are occurring); we are not concerned with the qualitative characteristics of the occurrent experiences, but only with the brute fact of their occurrence at a particular p-location.

Secondly, experience ownership claims should have the same meaning, qua sentient being, with respect to their tense. That is, they should translate from present to past or from future to present, etc., without any change in ownership, at least in so much as the ownership claim is with respect to some particular sentient being. To see this, consider that the claim, "this" experience belongs to "this" sentient being is equivalent to the claim that, "this" experience is a component of "this" stream of consciousness. Given this, it follows that tense translations will not affect ownership because the experience will still be a component of the same (i.e., numerically identical) stream of consciousness. So, the above claim, "This experience belongs to this sentient being" translates as, "That experience belongs to that sentient being", where the reference of the noun is the same. Similarly, the claim that, "That experience will belong to this sentient being", would translate as, "This experience
belongs to this sentient being", where, once again, the reference of the noun is the same. This observation is a consequence of the "Temporal Constraint".

§3

I now propose to discuss why only a theory which can meet these constraints will be commensurate with our intuitions regarding what is involved in survival as the same sentient being. I hope to achieve this goal by reductio ad absurdum; by demonstrating the consequences of denying any of the constraints. Firstly, regarding the Contingent Brain Constraint, enough has probably been said already. Suffice it to say that, to deny the contingency of this relationship would be to commit a category mistake, while to deny the causal necessity of brains for the existence of minds would contradict abundant empirical evidence to the effect that brains cause minds. Similarly, regarding the Transfer Constraint, the denial of this as a logical possibility would constitute a category mistake (i.e.: this denial would categorize streams of consciousness as the sorts of things which cannot conceivably be so transferred).

Secondly, if one denied the Phenomenological Constraint, (i.e.: that all that matters to one's survival as the same sentient being is the fact of the occurrence of one's stream of consciousness) then one would be forced to admit that there must be some further factor to be taken account of when describing the
nature of our lives as sentient beings. But what could this "further fact" be? We have already seen that it could not be anything like a Cartesian ego or a soul, for it is perfectly consistent to imagine possible worlds where there are streams of consciousness just like our own but which are not at all associated with anything like such an entity. Also, since retention of the same brain is only causally (i.e.:contingently) necessary, it cannot be constitutive (i.e.:it is not essential) of our lives as sentient beings. However, we have seen that sentient beings are essentially streams of consciousness; in fact, that is all there is to being a sentient being.

Thirdly, regarding the Division Constraint, the denial of this would contradict abundant empirical evidence and is therefore simply false. And lastly, the denial of the Temporal constraint would force one to the conclusion that experiences coming after a period of unconsciousness could determine the ownership of experiences occurring earlier. But this should seem intuitively absurd after all. For recall that, if this were the case, then the experiences of the man coming out of the coma only to reach a state of bare sentience would constitute a tiny life of their own. Conversely, if these experiences were followed by a state of full consciousness wherein the man fully resembled his former self, then we would say that the experiences "had" during the period of bare sentience did, after all, belong to the same person (and, ipso facto, the same sentient being) as before the
decline into a coma. In this way the experiences during his state of lucidity would determine, retroactively, the ownership of experiences which had occurred previous to them, which is absurd. This further underscores the observation that survival as the same sentient being is an all-or-nothing matter.

Let us now consider the theories covered in Chapters 2 and 3 in light of these constraints. Firstly, there is the Cartesian theory. Such a theory can meet the Contingent Brain Constraint, the Transfer Constraint and the Temporal Constraint. It cannot satisfy the Phenomenological Constraint, since the further fact of this "ego" which "has" the experiences must also be accounted for in the exhaustive description of what it is that constitutes a sentient being. And of course, as already covered in some detail, it cannot satisfy the Division Constraint.

Secondly there are psychological continuity theories. Such theories can satisfy the Contingent Brain Constraint, the Transfer Constraint, the Phenomenological Constraint and the Division Constraint. They cannot, however, satisfy the Temporal Constraint since ownership is determined by the qualitative appropriateness of one's experiences, character traits, etc. So, the occurrence of qualitatively appropriate experiences following a period of bare sentience would force us to the conclusion that the experiences during this period of bare sentience did, after
all, belong to the same person. Conversely, in the absence of such qualitatively appropriate experiences, the experiences during the period of bare sentience would constitute a tiny life of their own.

Lastly, there are the physical continuity theories. Such a theory would satisfy the Temporal Constraint and the Division Constraint only. It obviously could not satisfy the Contingent Brain Constraint, since retention of the self-same brain is essential. For similar reasons it also could not satisfy the Transfer Constraint. And since an account of survival must make essential reference to the retention of the same brain, it cannot satisfy the Phenomenological Constraint. It is not necessary to consider separately the combined theory.
CHAPTER 6
Sentient Being Theories: Two Alternatives

In this chapter I will discuss the details of p-space and of p-locations with the intention of showing how our construal of these notions affects the intuitive plausibility of sentient being theories generally. The chapter begins with a brief reiteration of the general details of the notions of p-space and p-locations from Chapter 4 and then moves on to a discussion of the distinction between qualitative and non-qualitative psychological relations. This leads naturally to a discussion of two alternatives of this theory that have been defended by Robert Elliot and Richard Sikora respectively. After discussing the inadequacies of both of these alternatives, the stage will have been set for the introduction, in Chapter 7, of the version of the sentient being theory to be defended here. I call this version the "minimal neural activity" theory.

§1

Let us recall then, from Chapter 4, that p-locations function as indices which pick out individual streams of consciousness in p-space by being the unique place within p-space where all of the experiences of any particular sentient being occur. It is a fact, ex hypothesi, that all of the experiences at any particular p-location will all belong to the same sentient being, for we have defined a sentient being as
essentially a stream of consciousness "at" a particular p-location. More importantly, the relation between the stream of consciousness and the p-location at which it occurs is non-qualitative, (this will be dealt with at length in §2). In this way we are able to claim that the relation between the two adds nothing to the content of the stream of consciousness itself. That is, because the relation between the two is non-qualitative and because the p-location has no ontological status independently of the occurrence of experiences, the p-location itself is not a further fact to be accounted for when giving an exhaustive description of the content of the stream of consciousness of any particular sentient being; it does not contribute any qualitative features to the stream itself. The p-location functions purely as an indexical which is metaphysically adequate to numerically individuate sentient beings within p-space, since each sentient being will be uniquely associated with one and only one p-location. But what are we to make of the claim that the relationship between a stream of consciousness and the p-location at which it occurs is non-qualitative? To answer this we must investigate the distinction between qualitative and non-qualitative psychological relations.

§2

A relation between experiences which is a function of the quality of those experiences is a qualitative relation. Some examples will make this clearer. If, while looking through a
magazine, I see a picture of Venice and am subsequently reminded of my vacation there last year, because the photograph arouses certain associative memories for example, then there is a qualitative relation between my present experience of the picture in the magazine and my memory of being in Venice last year. Qualitative relations permeate our mental goings-on in a myriad of ways. For example, any behaviour which I display that is appropriate given my character, will be qualitatively related to dispositions I have to behave in certain ways and so forth. An example cited by Dr. Sikora is this, "A has always been an impulsive person. Consequently, it would be qualitatively appropriate if A's future experiences also displayed impulsiveness."

Sentient being theories rely heavily on the notion of non-qualitative psychological relations between experiences. The reasons for this have already been explored. Recall from Chapter 3 that any theory which purports to explain the unity of a person's life by reference to the qualitative continuity of certain psychological characteristics (e.g.: character, beliefs, intentions and so forth) will, on certain occasions, be forced to admit that a particular case is undecidable. That is, in some cases we will not be able to decide, given the empirical evidence at our disposal, whether or not X is the same person as Y. We noted that this was a consequence of the open-texture of the

---

Sikora 6.
concept of person in certain contexts, when the concept is defined within an empirical theory. I further claimed, in Chapter 4, that the concept of sentient being remains distinct in many contexts where the concept of person is open-textured. This is the case specifically because, in so much as we are concerned with sentient beings, we are not concerned with the qualitative nature of the experiences that constitute one's stream of consciousness but only with the brute fact of their occurrence.

Obviously, this can only be the case if the unity of one's life as a sentient being can be cashed out in terms of psychological relations between experiences that involve only non-qualitative relations. In fact, it will be maintained here that there are two types of non-qualitative psychological relations: synchronic unifying relations and diachronic unifying relations. Synchronic unifying relations account for the unity of the simultaneous components of our experiences at a particular instant. For example, like Descartes we could speak of feeling the warmth of the fire, the smell of the burning logs, the tactile sensation of the melted wax in our hand while at the same time looking out the window we could see men walking by in the street. Now all of these sensations occur as the simultaneous components of experience and, it will be maintained here, they are related to each other non-qualitatively by the synchronic unifying relation. The relation itself consists in the fact that these components of experience are related in this way simply in
virtue of their occurring at the same p-location simultaneously. That is, the p-location "bundles" the simultaneous components of experience just by being the "place" where they all occur. Similarly, all experiences, considered as synchronically unified wholes, will be non-qualitatively related by the diachronic unifying relation. That is, the diachronic unifying relation consists in the fact that any experiences that occur within the same continuous stretch of consciousness will be diachronically related in a non-qualitative way simply in virtue of their occurring at the same p-location. Again, the p-location "bundles" the experiences that occur within any continuous stretch of consciousness just by being the "place" where they all occur. Further, it should be obvious that the bundling does not contribute anything qualitative to the content of the stream of consciousness, since the bundling is not a component of the experiences themselves; the bundling is just a matter of sharing the same p-location in much the same way that two people may share the same address.

In any case where we are trying to determine whether experiences which have occurred at different times belong to the same sentient being these non-qualitative relations will take precedence over qualitative relations. In fact, the qualitative relations will only be relevant when we are talking about persons. When attempting to determine if X is the same person as Y, we must rely on observable characteristics. Now take an
example where you have just spotted someone you think may be an old friend, call him X, but, since you haven’t seen him in years, you can’t be sure from appearance alone. Rather than embarrass yourself by approaching X, you decide to observe him from a distance as unobtrusively as possible. You will watch him to see if his behaviour, mannerisms and so forth are appropriate, given what you remember about your old friend. That is, your judgement will depend on whether or not you think that the displayed character traits are qualitatively appropriate given what you know about the person. But when we are concerned with whether or not X is the same sentient being as Y, from a metaphysical viewpoint we need only be concerned with the non-qualitative relations since these are all that determine whether or not an experience belongs to the same p-location.

Non-qualitative relations, particularly the diachronic relation, will be dealt with later in section 2 of §3 when discussing an alternative version of the sentient being theory held by Sikora. But now, with the details of the last two chapters and the discussion of qualitative and non-qualitative relations at our disposal, we can move on to an examination of two alternative versions of the sentient being theory. We will commence with the theory held by Professor Robert Elliot.\footnote{Professor Elliot is at the Educational College of Brisbane in Australia. I am presenting the details of his theory here from recollection of a discussion with him. He has not, to the best of my knowledge, published anything on the topic and I hope that the presentation here is accurate.}
1. **p-locations as self-subsistent entities:**

The distinctive feature of Elliot's theory is that p-space and p-locations are attributed an independent ontological status akin to physical space. All other details remain the same; sentient beings are still defined as essentially streams of consciousness "at" a particular p-location. The difference then, consists in the fact that, where streams of consciousness and p-space were previously described as concomitant effects of the same cause (i.e.: the brain), on Elliot's theory this is not so; p-space is a self-subsisting entity in its' own right. On Elliot's theory then, to speak metaphorically, a sentient being just "moves into" a p-location. Further, the association between the two is claimed to be contingent in the sense that the first experience of any sentient being just "falls into" some p-location or other purely gratuitously. Subsequently, it is claimed that all other experiences of the sentient being will continue to occur at the same p-location. As before then, it is claimed that every sentient being is uniquely associated with one and only one p-location. Given this, p-locations continue to serve their purpose as indexicals which allow us to numerically individuate sentient beings. Also, as before, the p-location unifies the experiences of the stream of consciousness, both

---

Unless one believes that physical space just consists in the relations between objects.
synchronically and diachronically. On this theory however, the p-location bundles the experiences and the simultaneous components of experience rather in the way that a container bundles a load of hardware; that is, the bundling is no longer simply a result of the experiences and the simultaneous components of experience just sharing the same p-location (like two people sharing the same address) but is the result of being contained in the same space.

Now there are several objections that may be brought against this theory but it may be instructive to first inquire why Elliot should have been motivated to construct such a theory. The reason is to be able to explain how it is that sentient beings survive gaps in consciousness. For consider that, as described above, the diachronic unifying relation can only unify experiences that occur within a continuous stretch of consciousness. Most of us would agree that we are all unconscious at various times in our lives during dreamless sleep for example. It is obvious then that our streams of consciousness occur discontinuously; they have a "gappy" existence. But it was previously maintained that the stream of consciousness and its' associated p-location are temporally coextensive so, it follows that p-locations occur discontinuously as well. Now if this is so, then what is to explain the reoccurrence of the same p-location across such gaps? Well, for Elliot this problem is eliminated for p-locations do not exist discontinuously and so, we can be assured that the
sentient being survives periods of unconsciousness as the same sentient being (i.e.: it continues to occupy the same p-location throughout its’ life). As we shall see in section 2, Richard Sikora defends a different theory which postulates a type of diachronic relation that is responsible for unifying the stream of consciousness across such gaps.

But let us now consider some objections to Elliot’s theory. Firstly, one might object that the very fact that we have added to our ontology makes the theory less desirable than one which does not. Elliot’s response would be that this theory gives him greater explanatory power, with respect to the problem of gaps in consciousness, and this alone is sufficient to justify the ontological commitment. Furthermore, he would reply that the notion of an independently existing p-space is not so strange if one is willing to grant property dualism, for surely these mental events must exist somewhere. He might even claim that it is no stranger than the notion of an independently existing physical space which many take for granted. Let us grant him enough then to see if this claim is justified.

Secondly though, it is a logical possibility on Elliot’s theory that there could be more p-locations than there are sentient beings. In fact, unless the claim is that an available p-location just “pops” into existence when needed, I cannot see how there could not be more p-locations than sentient beings.
And, if there were more p-locations than there are sentient beings, then there would necessarily have to be some empty p-locations. On this view, there could even be an infinite number of p-locations in p-space and some of these would also contain sentient beings; clearly, those which are not sentient beings would be merely empty p-locations. Hence, what distinguishes a sentient being from a merely empty p-location on this view is the occurrence of experiences at any p-location peculiar to some particular sentient being. The problem is that this seems to threaten their very function as indices, for consider the case where a sentient being occupies a p-location, then dies, and another sentient being subsequently moves into the same p-location. What distinguishes the two sentient beings on such a theory? It seems that Elliot has a choice between two alternatives but both involve ad hoc restrictions. On the one hand, he may claim that p-locations and streams of consciousness are temporally coextensive as before and, in this way, the above problem of identification is obviated. But then Elliot must offer us an explanation of how p-locations conveniently "pop" into and out of existence. Recall that on the previous formulation (described in Chapter 4), this is not a problem, for the p-location will necessarily be temporally coextensive with the associated stream of consciousness because both p-space and p-locations are concomitant effects of the brain. On the other hand, he may wish to deny that p-locations just conveniently "pop" into and out of existence as needed but, in this case, he
would have to face the problem above: how to individuate two or more sentient beings that may occupy the same p-location at different times.

Secondly, regardless of which of the above alternatives he should choose, he owes us an explanation of why, after the first experience gratuitously falls into just some p-location or other, all the subsequent experiences continue to occur at the same p-location. The only possible reply, it seems, would be that this follows ex hypothesis. He might go on to say something like the following. In so much as consciousness is coextensive with brains, the neuro-physiological correlates of experience occur in physical space and likewise the experiences themselves occur in p-space. Similarly, just as the particular location in physical space where the neuro-physiological correlates of experience occur is contingent, so likewise for the experiences themselves occurring in p-space. However, all the experiences of any particular sentient being will always occur at the same p-location because each p-location is equidistant from all others (recall that this was one of the distinguishing characteristics of p-space); there is no intervening topology to permit transfer of consciousness to another p-location. Hence, although the first experience of any sentient being "falls into" some empty p-location or other gratuitously, all subsequent experiences of that sentient being will also occur there as a matter of course.

---

Again assuming property dualism, which Elliot does.
Except for the case of fusion already noted, the streams of consciousness that constitute sentient beings cannot traverse the discontinuous landscape of p-space to arrive at any other p-location. However, as noted by Professor Hanson, it is not obvious that a stream of consciousness should need a landscape to traverse the gap between different p-locations. Why is it that it can't simply jump? Of course this is not a problem during any continuous stretch of consciousness, for the p-location and the associated stream of consciousness will be associated vis a vis the diachronic unifying relation. Therefore, we can be sure that, during any continuous stretch of consciousness, all the experiences of any sentient being will be diachronically united at the same p-location. But Elliot cannot rely on this as an explanation of why, after a period of unconsciousness, the first experience does not occur "at" a different p-location.

Thirdly, given the ontological status of p-locations on Elliot's theory, it would constitute a further fact to be taken account of in our description of a sentient being. Recall (from Chapter 4) that p-locations are temporally coextensive with their associated streams of consciousness and that the association between the two is non-qualitative; the diachronic unification of the experiences in any continuous stretch of consciousness is just a consequence of their all sharing the same p-location, rather like two people sharing the same address. But on Elliot's theory the p-location is conceived of as something like a
container rather than merely as an indexical. On this conception the relation between the experiences and the p-location as a container must be accounted for since it is a fact that the experiences are now "contained"; they do not simply occur "at" a particular p-location. So, any complete description of what it is to be a sentient being on Elliot's theory, must make essential reference to the p-location, qua container, as an explanation, for example, of why the experiences are located where they are, how they are contained, and so on. For example, if I am sitting in a chair, then a complete description of myself at that time must make reference to the chair; as an explanation of why I am situated as I am and so forth. The p-location is now a relata to be accounted for.

Now Elliot's reply is that the association between the self-subsistent p-location and the experiences is a contingent one. That is, the first experience of a sentient being just "falls into" some p-location or other gratuitously and subsequent to that all experiences of "that" sentient being will continue to occur at the same p-location. Given this contingent relation, he goes on to claim that the relation is also non-qualitative and therefore the p-location need not be accounted for as a relational item. If I understand him correctly then, the claim is that since the association between the stream of consciousness and the p-location is contingent it follows that it is also non-qualitative. And since it is non-qualitative, it follows that the
p-location adds nothing to the qualitative content of the stream of consciousness. Since this is the case, the p-location can be shown to serve as an indexical only, despite its independent ontological status; it need not be considered as a relational item that must be accounted for.

Now I think this response fails for the following reason. It trades on a confusion regarding the meaning of the terms "non-qualitative" and "contingent". To say that the relationship between two items is contingent does not imply that the relationship is also non-qualitative, yet this is precisely what Elliot seems to be claiming. Further, as claimed above, it does not follow from the fact that the relation between the two might be non-qualitative that, therefore, the p-location is not a relata to be accounted for. Elliot seems to be assuming that, since the relation between the two is both non-qualitative and contingent, the fact of the p-location's ontological status is irrelevant. Yet I claim that this is not so; it is now a further fact that must be accounted for in any description of sentient beings. As noted above, it acts rather like a container and, therefore, it must be referred to as an explanation, at the very least, of why my stream of consciousness is oriented this way rather than that for example.

Now Elliot has also emphasized the claim that, on his theory, p-space is a metaphysical primitive and, so, perhaps he
is to be construed as follows. Given the fact that p-space is a metaphysical primitive, its' ontological status is irrelevant when describing a stream of consciousness, in the same way that one might claim that the ontological status of physical space is irrelevant when describing a physical event. In the causal explanation of any event occurring in physical space, no reference is made to space itself as one of the relata of the event. Rather, we arbitrarily select some frame of reference, using the coordinates of space and time, to numerically individuate that event. Analogously, we lay a set of coordinates over p-space as a frame of reference to numerically distinguish mental events. Now since we might have laid out the coordinates in any way, it follows that the relationship between any experiences and the p-location at which they occur is contingent. And since I think he would also wish to say that relations to frames of reference are non-qualitative, it follows that the relation between the p-location and the stream of consciousness is both contingent and non-qualitative. I similarly remain unconvinced by this line of argument for the simple reason that it fails to establish the claim that the relation between experiences and p-locations is non-qualitative. To say that one relata is merely a frame of reference does not establish the claim that other items which are related to it are necessarily related non-qualitatively. Further, it is not clear, from the fact that he would regard the p-location as an unanalyzable primitive, that, therefore, its' ontological status is irrelevant
to its' function as an indexical. We are then obliged to ask what 
the frame of reference is laid over. The fact remains that he has 
made the claim that p-locations subsist, even when there are no 
experiences occurring at them, and this certainly seems to commit 
him to an ontological status for p-space and this needs 
explaining.

However, if we assume that his theory is compatible with the 
possibility of empty p-locations, then he would be forced to 
refine the original definition of what it is to be a sentient 
being. On the original definition, empty p-locations would now 
turn out to be sentient beings and it would follow that there are 
many sentient beings who have no experiences and never will. So 
it is necessary to make some appeal to experiential content in 
our definition;

\[
X \text{ is a sentient being iff,}
\]
\[
(1) \text{ X is essentially a stream of}
\]
\[
\text{consciousness "at" a p-location distinct}
\]
\[
\text{from all other p-locations}
\]
\[
(2) \text{there is experiential content (of at least a rudimentary sort) occurring at}
\]
\[
X'\text{'s p-location}
\]

Regarding clause (2), it is not required that these experiences 
occur continuously. In fact, there seems to be considerable 
evidence to indicate that there are frequent periods in the lives 
of all sentient beings in which there will be no occurrent 
experiences; for example, during periods of unconsciousness or 
sleep. Now perhaps it is possible to make a theory like Elliot's
work if it can be shown that there is some way of construing his claims in a way that can adequately meet the objections. However, there is an alternative theory, favoured by Dr. Richard Sikora, which relies on a non-qualitative, gap-crossing relation to explain how sentient beings survive gaps in consciousness.

2. Gap-crossing relational theory:

The distinctive feature of Sikora's theory is that he explains the diachronic unity of streams of consciousness across gaps in terms of what he calls a "non-qualitative, gap-crossing relation" which will be explicated below. On this theory, streams of consciousness and their associated p-locations are conceived of as temporally coextensive. Consequently there are never any empty p-locations. The problematic issue of the ontological status of p-locations is obviated by denying that they can occur independently of sentient beings; that is, p-space is explained as a causal effect of brains and there just is a particular p-location associated with each stream of consciousness. Since both streams of consciousness and p-space are conceived of as concomitant effects of brains, p-space does not have any ontological status independently of sentient beings, as is the case with Elliot's theory. On Sikora's theory then, the association of an experience with the p-location at which it occurs is non-qualitative; the p-location functions purely as an indexical to numerically individuate the sentient being "having" the experience. Given this, and the fact that p-space has no
independent ontological status, the p-location is not a further fact to be accounted for in a complete description of a sentient being; p-locations function only to individuate sentient beings and they need not (in fact cannot) be accounted for as a relational item in any description of a sentient being. Consequently, the description of any particular sentient being will involve no more than a description of the experiential content of that sentient being's stream of consciousness at a particular p-location. The fact that the stream of consciousness occurs at a particular p-location is a metaphysical fact that is constitutive of sentient beings but, since the relation between the two is non-qualitative, the p-location itself does not add to the qualitative content of the stream of consciousness.

But now let us consider the nature of this non-qualitative, gap-crossing relation. Prima facie, there will be a problem for any sentient being theory that does not rely on the strategy employed by Elliot in dealing with the periods of unconsciousness that are presumed to occur in the lives of all sentient beings. Since, on Sikora's theory, there are no empty p-locations and since no experiences occur during periods of genuine unconsciousness (as distinct from, say, dreaming periods of sleep) it follows that the p-location must cease to exist during such periods. So, we are faced with a variation of the problem that perplexed John Locke when he attempted to explain how it is that we are able to survive regular periods of unconsciousness as

115
the same person. Somehow, it must be possible for these p-locations to "reoccur" (as the same p-location) if this theory is to be successful. The solution to this dilemma, on Sikora's theory, involves the postulation of a non-qualitative, gap-crossing relation. The relation itself consists in the fact the experiences occurring before a gap in consciousness become linked to those following the gap simply in virtue of occurring at the same p-location. This relation then, can only link experiences within the same p-location and never across p-locations. And since it is non-qualitative, it adds nothing to the actual content of the stream of consciousness, qua sentient being. We should also pause to note that this is a type of diachronic, non-qualitative, unifying relation.

Sikora has at least three arguments to show that this gap-crossing relation only links experiences across gaps to the same p-location. In essence the claim is that this is true in virtue of something intrinsic to the first experience of any period of consciousness itself. He first argues from the continuity of consciousness during any continuous span as follows,

[All] of the experiences of any continuous span of consciousness not only share the same consciousness; they are also all associated with the same brain. The obvious explanation for this correlation is that it is because they are all associated with the same brain that they all have the same p-location. But if this explanation is correct, experiences associated with the same brain will have the same p-location whether or not they are separated by periods of unconsciousness so that there really is a non-qualitative
psychological relation for bridging periods of unconsciousness. \(^7\)

Now Sikora readily admits that this argument, in itself, can only establish that brains are a necessary condition for retaining the same p-location; this argument does not also establish that retention of the same brain is sufficient for retention of the same p-location.

But secondly, he argues by reductio absurdum that anyone who denies that the same p-location is preserved across gaps in consciousness is forced to a contradiction. Anyone holding such a view must hold some version of what he calls the "Cantilever Theory". In outline, the Cantilever Theory holds that;

(1) Since a different p-location occurs across gaps in consciousness, and

(2) Since we ordinarily suppose that all the experiences of any continuous stretch of consciousness occur at the same p-location,

(3) We must further suppose that the first experience of any continuous stretch gets its p-location in some way different than the rest and the subsequent experiences are "glued" to it in some way.

The Achille’s heel of this theory is that it must admit that the initial experience in a continuous series gets its p-location in a different way than the rest. It cannot get it from being glued to other experiences as the rest do: there isn’t an experience immediately before it for it to be glued to, and though it could be glued to the following experience, it can’t get its p-location from that because it will already have a p-location before the later experience occurs. Instead it would seem that an initial experience must get its p-location from some

\(^7\) Sikora 24.
sort of association with the brain while those that follow it get theirs from being fastened directly or indirectly to an initial experience.\textsuperscript{67}

He then proceeds to explicate all possible versions that such a theory might take and concludes that all involve contradictions.

Lastly, there is the argument from initial experiences,

1. p-locations remain the same for any continuous span of consciousness.
2. There are times when we are completely unconscious.
3. The p-locations of initial experiences are caused by the brain with which they are associated.
4. Either (a) the same brain always gives the same p-location to initial experiences or; (b) it does not.
5. It follows from (1), (2), (3) and (4a) that all experiences associated with the same brain have the same p-location even if they are separated by a break in consciousness which means that the p-location theory is true. Therefore if '4b' is false the p-location theory is true.
6. If (4b) is true changes in the brain cause it to give different p-locations to experiences occurring at different times.
7. It follows from (1), (2), and (6) that if (4b) is true, either; (a) Changes in the brain capable of altering the p-location of one's experiences only occur while we are unconscious or (b) they also occur while we are conscious but the brain is "shifted into neutral" so they have no effect.
8. Both (7a) and (7b) are false (he has argued for this in a previous argument on the grounds that both are implausible, ad hoc assumptions).
9. Therefore (4b) is false.
10. Therefore (by 5) the p-location theory is true.\textsuperscript{68}

\textsuperscript{67} Sikora 25 - 26.
\textsuperscript{68} Sikora 28 - 29.
The burden of plausibility for this account lies with the notion of a non-qualitative, gap-crossing relation. To recapitulate then, we seem faced with the following choice: we may regard p-locations as self-subsisting and consequently capable of persisting while empty. In this way we avoid the need for a non-qualitative, gap-crossing relation but find ourselves saddled with the problem of the controversial nature of the ontological status of such p-locations. Alternatively, we could regard p-locations as coextensive with sentient beings, thereby disallowing the possibility of empty p-locations. In this way, the problem of their ontological status is obviated but the gap-crossing relation becomes necessary to explain the unity of sentient beings across gaps in consciousness. In the following chapter we will examine a further alternative which will, hopefully, square with our intuitions while avoiding the dilemmas faced by these other theories.
In the previous chapter we examined two theories of personal survival that rely on the notion of sentient beings as the fundamental concept. We saw that those theories face the following dilemma: We could regard p-locations as self-subsistent entities in their own right, thereby avoiding the dilemma of how to explain the reoccurrence of the same p-location across breaks in consciousness. We must then provide an account of the ontological status of these p-locations that will not rely on any sort of qualitative relations between the p-location itself and the experiences that occur there. This, we discovered, is not an easy thing to do convincingly. Alternatively, we could regard p-locations as temporally coextensive with sentient beings and claim that they have no ontological status independently of their function as indices which serve to individuate the various streams of consciousness occurring within p-space. In this way, we may avoid the problems alluded to above but now we find ourselves forced to accept the strange (and some would say, counterintuitive) notion of a non-qualitative, gap-crossing relation.

The theory to be expounded here, like those discussed in the previous chapter, is a sentient being theory. In substance, it
most closely resembles the latter in that p-locations are here regarded as temporally coextensive with sentient beings. It differs from that theory though in that it dispenses with the problematical gap-crossing relations.

§1

The motivation to develop this theory arose from feeling that something very much like the sentient being theories already discussed must be right. It accords well with my intuition that, were an individual to decline to the point where he were no longer regarded as a person, he would still be a sentient being and, hence, capable of feeling pain at the very least. Furthermore, since he would be uncontroversially the same sentient being, he would be justified, previous to his decline, for any concerns he might have for the future welfare of that sentient being. Additionally, such a theory neatly avoids the problems encountered by those theories considered in Chapters 2 and 3 while yet admitting the intuitively obvious point that the qualitative nature of certain psychological characteristics is essential to our notion of personhood. And yet we are able to do this without sacrificing the intuition that our survival as the same sentient being is an all-or-nothing matter. This result, to anticipate a conclusion to be developed later, allows us to avoid the intuitively repugnant conclusion sanctioned by Derek Parfit's reductionist theory, to the effect that being survived by replica would be about as good as ordinary survival.
Clearly, the problems encountered by the sentient being theories considered above are considerable though and it is from the effort to circumvent those difficulties that this theory arose. If this theory is to retain it's initial intuitive plausibility, it is crucial that it be formulated in such a way that p-locations are not granted a kind of Platonic existence and then blithely claimed to be related in a non-qualitative way to the experiences that occur "at" them. Similarly, it would seem prudent to avoid, if possible, the need for anything as controversial as a non-qualitative gap-crossing relation. The "minimal neural activity" theory (hereafter: MNA theory) succeeds in its goal of presenting a sentient being theory of personal survival that obviates these difficulties. Let us proceed then, to an examination of this theory.

On this alternative account, the following apparently mutually inconsistent propositions will be true:

(1) As a person I am also a sentient being.

(2) As a sentient being I am essentially a stream of consciousness "at" a particular p-location (or equivalently, a "haver-of-experiences" or a subject of experiences "at" a particular p-location).

(3) p-locations do not occur independently of sentient beings, but are temporally coextensive with them.

(4) In the life of any sentient being there will be only one p-location associated with
it; all experiences of "this" sentient being occur at "this" p-location.

(5) No non-qualitative, gap-crossing relations are necessary to explain the unity of a sentient being across gaps in consciousness.

While I frankly doubt that we are ever totally unconscious during sleep (even during dreamless sleep), I take the case of someone in a coma to be paradigmatic of unconsciousness. Yet there seem to be unobjectionable grounds for claiming that someone coming out of a coma is the same sentient being as he was before the coma (and probably also the same person). It does not follow that any gap-crossing relations are necessary to account for this unity across such uncontroversial periods of unconsciousness though. To illustrate, consider this analogy. Certain programmable calculators that do not use software to store programs are capable of retaining programs in the memory, even while shut off, for up to 1 year. They do this by maintaining a minimal current-flow in their circuitry which is both necessary and sufficient to keep the memory "active"; i.e., it is at least able to retain the program entered, although it cannot run the program or accept input. If we consider the case of the calculator while shut off as analogous to a sentient being while unconscious, then the solution to our dilemma is clear. Even in periods of unconsciousness, the brain maintains a certain minimal amount of neural activity which is both necessary and sufficient to keep the p-location "active". This guarantees the continued existence of the p-location during these gaps in
consciousness and, since survival as the same sentient being depends on retention of the same p-location, it follows that there are no gaps in the lives of sentient beings during such periods of unconsciousness.

Furthermore, there is empirical evidence to support this view. The medical profession can easily distinguish between someone who is in a coma, or unconscious, and someone who is brain-dead. The results of encephalographic brain scans indicate that, while unconscious, there still occurs a minimal amount of neural activity. When even this is absent, the only activity of the central nervous system is that of the autonomic, life-support functions and such persons are said to be brain dead.

It may be objected that the empirical fact of this minimal amount of neural activity establishes nothing in this context unless it can be shown that it has some connection to sentience. Otherwise, without the connection to sentience, one could argue that any other characteristic you choose could be regarded as the sign that the p-location is present, for example, a certain level of blood-sugar, or whatever. Alternatively, one might construe this to be a materialistic theory of mind; the minimal neural activity of the brain is necessary and sufficient for sentience just because it "is" sentience (where the "is" indicates identity). But then it may be countered that, in so much as the brain is spatio-temporally located, there is no need for p-space.
or p-locations to numerically individuate sentient beings. We would then have a substantially different theory from those considered in the last chapter though, for consider that a theory such as the one outlined here would violate both the Contingent Brain Constraint and the Transfer Constraint of Chapter 5.

I will accept neither of these alternatives though. Rather, I will argue that the minimal neural activity is an indication that the sentient being is likewise minimally conscious. Specifically, I shall treat the concept of consciousness as a scaling concept, in accordance with its treatment by the medical profession. Hence, to say that one is unconscious is actually a misnomer if it is taken to mean "not conscious"; rather, someone who is unconscious is only minimally conscious. This underscores our earlier observation that, qua sentient being, one's level of consciousness need only be of the most minimal level to be sufficient for one to be "having" experiences. This is precisely why the concept of sentience was observed to be of much wider scope than that of person; even a reptile is sentient and, as I claim, so is person in a coma. These observations also underscore the observation made above to the effect that the medical profession readily distinguishes between someone in a coma and someone who is brain-dead. The answer should now be

---

*I contend that someone in a coma who is experiencing only a dull throb of pain is still "having" experiences although I admit that this is not the ordinary sense of "experience".*
clear that, in fact, it is the brain-dead victim who is actually unconscious in the sense of "not conscious".

Again the justification for treating consciousness as a scaling concept is empirical. The medical profession treat it as such and have an elaborate set of standard tests which measure the level of awareness of the patient in terms of his or her sensitivity to stimulation. The claim here is that these levels of awareness correspond to levels of consciousness and that, when

The medical scale of the levels of consciousness is called the "Glasgow Coma Scale". The tests measure the patient’s response along three parameters of sensitivity to stimulation, corresponding to the response of the eyes, verbal response and motor response. The lower the score, the less conscious is the patient;

(A) Eyes:
   (4) Eyes open spontaneously
   (3) Eyes open to speech
   (2) Eyes open to pain
   (1) No response
(B) Verbal:
   (5) Oriented - time, place, person
   (4) Confused
   (3) Incomprehensible sentences and/or words
   (2) Incomprehensible sounds
   (1) No response
(C) Motor:
   (6) Obey instructions
   (5) Localizes pain
   (4) Withdrawal response due to pain
   (3) Flexion response - decerebrate rigidity
   (2) Extension response - decorticate rigidity
   (1) No response

These tests are only intended to function as a diagnostic device that indicate the awareness of the patient but the claim here is that these levels of awareness correspond to measurable levels of neural activity which in turn indicate distinct levels of consciousness.

126
unconscious in the sense of being in a coma, one is actually minimally conscious. Similarly, during dreamless sleep one is less conscious than someone who is awake but yet more conscious than someone in a coma. It is here claimed then, that sentient beings do not lead "gappy" lives at all and so there are no gaps to be explained away. Further, this theory also supports the claim made earlier that, as a sentient being, the most minimal level of consciousness is sufficient to qualify as sentience.

Now it might be objected to this that I am referring to a different sort of consciousness from that talked about by Elliot and Sikora. From this it would follow that the theory offered here cannot constitute a solution to their problem. The argument would be that, when referring to consciousness, even the sort appropriate to sentient beings, they meant conscious awareness of the type referred to by Locke; that is, conscious awareness of one's thoughts as belonging to oneself and the ability to recall these as memories of thoughts had by oneself. Quoting Locke,

For, since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things: in this alone consists personal identity, i.e., the sameness of a rational being; and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.\(^7^1\)

\(^7^1\) Locke 449.
I resist this conclusion and argue that, in so much as they are sentient being theorists, they are similarly committed to the claim that consciousness, qua sentience, can be a matter of degree. For recall that sentience has been defined (in Chapter 4) as the minimal sort of awareness that would be sufficient to be a subject of experiences. In so much as they agree with the claim that even animals low on the evolutionary scale qualify as sentient beings, they cannot be construed to be talking of sentience in the same sense as Locke. This sort of consciousness is self-consciousness as opposed to consciousness simpliciter (i.e.: the opposite of unconsciousness).

It might be further objected that, despite the results of encephalographic brain scans, it has not been established that someone in a coma is conscious in any sense appropriate to the notion of sentience intended here; i.e.: such an individual may exhibit different brain patterns from someone brain-dead, but this does not establish that he is "having" experiences. To this I respond that, firstly, such individuals are not always unresponsive along all three parameters of measurable sensitivity indicated in footnote 1. They often respond to stimuli by grasping, moaning and so forth. This implies at least minimal awareness of their surroundings which in turn implies that they are "having" experiences. But secondly, even when unresponsive, I claim that the measurable brain activity is an indication that the individual is at least sentient. Consider the example of a
man in an accident who is badly injured and declines to a state of coma. If we could follow his subjective experiences as he declines we might notice that his awareness of his environment declined by degrees such that he first lost his visual component, then his auditory component, then his olfactory component and so on. Eventually, we find our victim with no awareness of his environment at all and he is completely unresponsive to stimulation yet there remains a dull and continuous sensation of pain. I argue that even this dull sensation of pain is sufficient for our victim to qualify as sentient.

§3

From these considerations it follows that gap-crossing relations are not necessary to explain the unity of sentient beings across gaps in consciousness for there really are no such gaps in the lives of sentient beings. The gap referred to occurs in the life of the person; that is, an individual who is unconscious is no longer a person but is still sentient. Furthermore, since p-locations are temporally coextensive with streams of consciousness, they do not occur independently of them. As such, they function purely as indexicals effectively individuating sentient beings by their unique location in p-space. As argued previously, they add nothing qualitative to the content of one’s stream of consciousness; they simply bundle the experiences of a sentient being in the sense of being the place where all the experiences occur. Such a theory then, offers us
the best results of the theories considered above. Like the first, it explains the unity of sentient beings without relying on any kind of gap-crossing relation. And, like the second, it regards p-locations as temporally coextensive with sentient beings, thereby avoiding the difficulties associated with postulating independently existing p-locations.

Such a theory, like the two considered previously, is, according to our classification system, a "further fact" theory. That is, the notion of a sentient being is a further fact to be accounted for in the analysis of what it is to be a person. It should be noted however, that both this theory and the gap-crossing theory differ dramatically from the sort of further fact implied by Cartesian ego theorists. Like the Cartesian ego theory, these theories regard the crucial factor in survival as non-qualitative but, unlike the Cartesian theory, this non-qualitative factor is not regarded as some sort of immaterial substance. That is, both the Cartesian theory and the sentient being theory explain the unity of person's lives in terms of something ultimate but, in the case of the Cartesian theory, this ultimate entity is conceived of as a sort of substance different in type from the physical stuff of which we ordinarily suppose persons to be constituted. Elliot's theory, on the other hand, seems to also be a "further fact" theory of what it is to be a sentient being. In so much as he regards p-space as ontologically independent of sentient beings, it constitutes a further fact in
the analysis of what it is to be a sentient being. And, like the cartesian theory it regards this further fact which is constitutive of our lives as sentient beings as something substantial yet immaterial and, in principal, beyond experience. These two then are "substance" theories and, for many, this would be sufficient reason for rejecting the self-subsistent theory of p-locations regardless of those reasons already considered in Chapter 6. We shall see, in the next section, that there are also other grounds for rejecting such a theory.

There is a further point of similarity between Sikora's gap-crossing theory and the minimal neural activity theory. These theories hold that survival is contingently dependent upon retention of the same (i.e.: numerically identical) brain. The claim here is not that retention of the same brain is in any way constitutive of what it is to be the same sentient being but, rather, that there is a causal dependence which holds in virtue of certain basic natural laws. We will consider these points of difference and similarity further in the next section when we compare the three sentient being theories in terms of the constraints outlined in Chapter 5.

For simplicity we will designate the three as follows;

1) the self-subsisting p-location theory will be referred to as "E" for Elliot's theory
2) the non-qualitative, gap-crossing relation theory will be referred to as "S" for Sikora's theory.

3) the minimal neural activity theory will be referred to as "H" for Henry's theory.

Regarding the Contingent Brain Constraint, all three theories are compatible with it but only "S" and "H" require it. It is a requirement for these theories in so much as both regard p-space and streams of consciousness as concomitant effects of brains. Again it must be stressed that this requirement is causal and not necessary (on the interpretation of the causal relation as contingent); it in no way implies that retention of the same brain is constitutive of our lives as sentient beings. It is conceivable that both p-space and streams of consciousness could occur without being associated with a brain, or even anything resembling a brain, (i.e.: there is a possible world in which the causal laws of our own do not hold and all that exists are streams of consciousness and p-space). Consequently, to construe the connection between brains on the one hand and p-locations and experiences on the other, as necessary in any logical sense would be to commit a category mistake. The claim here is simply that in this world, the actual world, as a result of certain natural causal laws, retention of the same brain is a contingently necessary and sufficient condition for survival as the same sentient being.
On the other hand, "E" is, on at least one construal, only weakly compatible with this constraint in the sense that brains would be causally necessary but not causally sufficient. They aren’t causally sufficient because they don’t determine the p-location. And, unless "E" included the stipulation that brains are causally necessary for the occurrence of experiences, brains would not even be necessary for survival as the same sentient being. Even with this stipulation though, "E" is only weakly compatible to the extent that brains are construed as necessary for the occurrence of experiences "at" a particular p-location, but are not necessary for the p-location itself; the brain, qua cause, is necessary only for the occurrence of experiences since this is all that distinguishes empty p-locations from sentient beings (i.e.:p-locations where experiences occur). That is, on this construal, brains are causally necessary but not causally sufficient.

All three theories are compatible with the Transfer Constraint. Recall from Chapter 5 that, for the event to be accurately described as a transfer of consciousness, we must suppose that both the stream of consciousness and the associated p-location are transferred together. We saw that this followed from a consideration of the alternatives for, if we suppose that the stream of consciousness is transferred without the associated p-location, then we haven’t actually transferred the stream of consciousness to some other brain at all. Rather, we have altered
the conscious at some p-location associated with a different brain such that it has come to resemble the consciousness at our original p-location; that is, to call it a transfer of consciousness at all, we must be able to say that it is now the "same" sentient being (which means the same p-location) at a different brain. On the other hand, if we suppose that we have transferred the p-location without the associated stream of consciousness, then, by default, we have not transferred the consciousness and so, once again, the event cannot be accurately described as a transfer of consciousness. So, in any case where both the stream of consciousness and the associated p-location are transferred together, we have a case of transfer of consciousness to a different brain and all three theories are compatible with this.

Clearly "E" violates the Phenomenological Constraint since there is something over and above what is contained in one's stream of consciousness, viz the p-location itself (which on his view is a separate entity), that is essential to survival as the same sentient being. As mentioned above, such a theory resembles Cartesianism, at least to the extent that it posits a independently existing substance (i.e., p-space) which is partly constitutive of one's life as a sentient being. As such it is not only a further fact theory of what is involved in being a person but it is also a further fact theory regarding what is involved
in being a sentient being. Both "S" and "H" are compatible with this constraint.

Regarding both the Temporal Constraint and the Division Constraint, all three theories should be compatible. On this point I see no difference between them. They should all agree on the two provisional constraints as well (i.e.: the all-or-nothing constraint and the experience ownership constraint).

Summarizing our results then, we see that "E" strongly violates at least one constraint (i.e.: the Phenomenological Constraint) and is only weakly compatible with the Contingent Brain Constraint. Theories "S" and "H" on the other hand are both strongly compatible with all our constraints. Given the argument in Chapter 5 which established why a theory which can meet these constraints is more consonant with our intuitions than one which cannot, we are left to conclude that theory "E" is less plausible than either "S" or "H". The decision between "S" and "H" will depend on whether one is willing to accept the notion of a non-qualitative, gap-crossing relation rather than the notion that the minimal neural activity which does, as a matter of empirical fact, occur even during unconsciousness is evidence of minimal consciousness in the sense required by such a theory. I argue that the latter choice is preferable for the following reasons. Firstly, we have already established that the experiences which characterize sentient beings at the most general level need only

135
be of the most rudimentary sort. In fact, they are so rudimentary that the particulars of the qualitative features of these experiences are irrelevant when giving a complete account of the sentient being's stream of consciousness. Secondly, the MNA theory allows us to explain what is involved in our survival in a way that is compatible with certain well established empirical results. It seems to require a substantially smaller leap of faith to believe that such neural activity could be characterized as experiences of the sort required than to accept the notion of such a gap-crossing relation.
PART 3

REDUCTIONIST THEORIES
CHAPTER 8

Reductionist Theories

We will commence this chapter with a review of Hume's position, showing why he was wrong to argue that one could not have unity in the case of something constituted of a succession of closely resemblant parts and why he was wrong to insist that to survive as the same person, one must remain unchanged. In §2, I will explicate the connection between Hume and Parfit's reductionist theory, showing how the latter is the logical descendant of Hume's theory. In §3 I will show how Parfit establishes the truth of reductionism and, finally, in §4, the theory itself will be explicated.

§1

It will be recalled from Chapter 3 that Hume claimed our notion of personal identity is ill-founded and cannot be justified. He argued that, as persons, we are constituted of bundles of impressions which are linked by the relations of causation and resemblance. Therefore, a person is, properly speaking, constituted of a succession of distinct objects (i.e.:"impressions") and is not a single object. Further, there is no particular impression which remains unchanged throughout the life of any person to establish the kind of continuity we presume is necessary to survive as the same person. Consequently, there is neither unity nor numerical identity in the lives of
persons and so, when we say that someone is the same person, we are simply wrong.

Now in the first place, Hume is here implying that it would be wrong to say that a succession of objects is, properly speaking, a single thing. Consequently, it is impossible for someone to remain the same person because, in so much as we are a succession of impressions, we are actually a collection of closely related persons and not a single person at all. Now consider a set, $\mathcal{O}$, of distinct but closely related objects, $O_1$ through $O_n$. In this series we have $n$ distinct objects but only one set. There is no contradiction in saying that we have $n$ objects and one $\emptyset$. Just as we can say that a succession of words constitute a sentence, so we can say that a succession of impressions constitute a person. Persons just happen to be the sorts of things that are constituted of parts in this way.

Quoting Penelhum,

So, in spite of Hume, there is no contradiction in saying that certain kinds of things are composed of a succession of parts, and yet are each only one thing. Whether a thing can have many parts or not depends entirely on what sort of thing it is. Most things (including people) do.\(^\text{\textsuperscript{22}}\)


\(^\text{\textsuperscript{22}}\) Penelhum 226.
In the second place, Hume was wrong to suppose that for something to remain numerically identical, it must remain unchanged. An acorn which grows to become an oak is still the same plant in the numerical sense but it is now a tree, whereas before it was an acorn. In fact, as pointed out by Penelhum, one cannot be said to have changed unless one is, numerically, the same person. The only sort of thing which cannot change without becoming a numerically distinct entity is something which is, by definition, an unchanging thing. An example cited by Penelhum is a musical note. As soon as the tone is raised or lowered, it becomes a different note. The plethora of objects in the world admit of a wide latitude of permissible changes, from musical notes which are, by definition, unchanging, to entities like persons which remain numerically identical despite dramatic change throughout a lifetime,

What kinds of changes can occur without our having to say that the thing has ceased to exist and given place to something else depends on what kind of thing we are talking about. To know what such changes are is part of what it is to know the meaning of the class-term for that sort of object. A house, or a person, is something which admits of many changes before we could say it has ceased to exist. To know what these changes are is to know, in part at least, what the words "house" and "person" mean.\footnote{Penelhum 227.}

To return to the example of the acorn, we can show how it remains numerically identical despite the changes that occur by...
temporally indexing its' properties. In this way we may say that the oak, which "is now" large, is the acorn which "was" small. Similarly, we may say that this person, who "is now" old, is the same person who "was" young, or this person, who "was once" impulsive, is the same person who "is now" cautious. Conversely, this person who "is now" impulsive is the same person who "will be" cautious.

Ironically, Hume himself distinguished two senses of the word "Identical". He refers to these as the numerical sense and the specific sense,

[Though we commonly be able to distinguish pretty exactly betwixt numerical and specific identity, yet it sometimes happens that we confound them, and in our thinking and reasoning employ the one for the other. Thus a man who hears a noise that is frequently interrupted and renewed, says it is still the same noise, though 'tis evident the sounds have only a specific identity or resemblance, and there is nothing numerically the same but the cause which produced them.]

From this it is clear that two things may be specifically identical if they are qualitatively indistinguishable, yet they will be numerically distinct. Hence, to remain unchanged is to remain identical in both the specific and the numerical sense, but the converse need not hold. Unfortunately, although Hume recognized the distinction, he seems to have missed its significance in this respect. Thus, he mistakenly assumes that

---

Hume, Treatise 257 - 258.
for a person to remain the self-same person (i.e.: numerically identical), he must retain specific identity. But, as the analogy of the acorn shows, persons can remain numerically identical without specific identity. In this respect, persons resemble acorns rather than musical notes.

In the next section we will examine Parfit's theory which takes Hume's position to its logical conclusion. Following Hume, he maintains a theory of psychological continuity in terms of the links of resemblance and causation. Given this, he concludes that personal survival is a matter of degree. Our lives as persons consist of a series of closely related episodes that belong to various persons, such that the resemblance between individuals diminishes with increasing time.

52

Essentially, the reductionist claims that it is possible to give a complete description of reality without any reference to persons. This follows from the fact that it is possible to provide a completely impersonal description of persons. To quote Parfit, "A person's existence just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events". A consequence of this, to be discussed at length below, is that it will sometimes be the case that the question of one's identity will be indeterminate. That

\textsuperscript{76} Parfit 210.
is, it will, in some circumstances, be an empty question to use Parfit's phrasing.

The connection with Hume is this. Parfit has arrived at his position by arguing that personal identity is a matter of the continuity of psychological characteristics, which he cashes out in terms of a theory of causation and resemblance along the lines of Hume. Further, he argues that the connectedness on which this continuity supervenes can hold to any degree. In cases where it holds strongly, we can say that X is the same person as Y and where it does not hold at all, we can say that X is not the same person as Y. But there are many cases, in the middle ranges, where continuity holds only to a diminished degree and, in these cases, the question of whether X is the same person as Y is an empty question. Notice that Parfit is not simply making the epistemological claim that we don't or can't know what the answer to the question is. Rather, he is making the metaphysical claim that, properly speaking, there is no answer in such cases. We can know all the facts regarding the case and still not know whether or not X has survived as the same person.

Parfit agrees with Hume to this extent then, that our standard notion of personal identity is false. He further agrees that the unity of our lives is to be explained in terms of the continuity of psychological characteristics that is broader than Locke's notion of memory and that the characteristics themselves
are linked by resemblance and causation. Now where Hume would claim that, properly, there is no identity in the lives of persons at all and that his theory of the links between our "impressions" is what we have instead of such identity, Parfit argues that we do have identity but that it does not hold throughout one's life. Rather, one's life is actually constituted of a series of successive persons who resemble each other to varying degrees. The degree of resemblance diminishes with time such that those individuals furthest removed may only resemble to a limited degree or perhaps (in some rare cases, usually only found in thought experiments) not even at all.

It is arguable that Parfit has taken Hume's position to its logical conclusion then, in so much as he begins from Hume's premise that there is no substance or entity, either material or immaterial, which explains the unity of our lives as persons. Empirically, all we can be assured of is the resemblance over time, both physically and psychologically, of the people we know and it is an obvious fact that this resemblance holds to varying degrees. Further, if there is no separately existing entity which is the "subject" of experience, then it follows that persons can be described in a purely impersonal way. Parfit is aware of Hume's flaws to the extent that he realizes that to have identity, one need not remain unchanged and nor is it impossible for a succession of resemblant parts taken together to constitute one thing. All that is required is qualitative similarity. Let us
turn our attention then to Parfit's arguments for establishing the truth of reductionism to understand how he has arrived at these conclusions.

§3

In outline, Parfit argues as follows. First he shows that, if we believe that our identity must always be determinate, then we are also committed to the belief that we are, as persons, some sort of separately existing entity apart from our bodies. That is, we exist as some sort of pure ego. He then shows that we cannot be such an entity. Given this, it follows that our identity is sometimes indeterminate. Further, he argues that, if reductionism were false, then our identity would always be determinate but, since this has been shown not to be the case, it follows that reductionism is true.

We are first invited to consider the "Psychological Spectrum" case in which a man, trapped by a diabolical scientist, is told that he will have his character altered in a series of steps by throwing switches until he comes to psychologically resemble Napoleon. Each stage will closely resemble its neighbors and the entire spectrum is intended to cover all possible degrees of psychological connectedness. Now in the cases at the near end, it is undeniable that he will be the same person. But what of the cases at the far end? Well, consider Williams' example from Chapter 3. If the man were told that, after a complete
transformation he would be tortured, we can ask, "Should he be concerned for his welfare?". Williams considered this argument to be a refutation of psychological continuity as a criterion of personal identity, on the grounds that it would still be the same man who would feel the pain of torture. He reasoned that, since we can be sure that it would still be the same man who would feel the pain in the cases at the near end, how could it be the case that a series of slight changes could eventually result in a different person? We would be forced to conclude that there must be some critical limit where the accumulated changes precipitate the transition to a new person. But this seems highly implausible for there is no empirical evidence of what such a critical limit might be. It seems that we could only arbitrarily stipulate such a limit and this contradicts our belief that our personal identity must always be always determinate. Now Williams took this to be an argument for taking physical continuity as the criterion of personal identity. In response to this Parfit invites us to consider the "Physical Spectrum" case where the same degree of small changes alter a man's physical appearance until he completely resembles Greta Garbo at age 30. Also, the changes involve the replacement of his cells with new ones so that, at the far end, there would be none of his original cells left. In this extreme case either we believe that it is still the same person, in which case we must believe that it is psychological continuity which is the mark of the person, or we believe that he is a different person. But if we believe the
latter, then we are again forced to conclude that there is some critical percentage of cells which, when replaced, result in a different person. Again, such a view is hard to believe, particularly since there are no empirical means of discovering what this critical percentage might be. If we believe that the man is the same person, we are also forced to conclude that psychological continuity provides personal identity even in the absence of the normal cause (i.e.: the continued existence of the same brain).

But now Parfit presents the "Combined Spectrum" case which involves all of the possible variations in the degrees of both physical and psychological connectedness. In the case at the far end, the man would resemble Greta Garbo at age 30, both psychologically and physically, and there would be none of his own cells remaining. If we believe that this man is still the same person, then we must conclude that his existence as the same person must consist in the persistence of some separately existing entity along the lines of a Cartesian ego. There is clearly no alternative since it cannot, in such a case, consist in the continuity of either the same brain or of the man's psychological characteristics. Alternatively, we could conclude that he was no longer the same person and this would imply that, in the middle cases, the question of the victim's identity is sometimes indeterminate or that there is some, as yet undetermined, level of change such that the victim becomes a
numerically different person. The last hypothesis is counterintuitive and, further, there are no empirical means for determining what such a limit could be. Consequently, our choices are either, (1) we are Cartesian egos and our identity is always determinate or, (2) our identity is sometimes indeterminate and reductionism is true.

Parfit next proceeds to demonstrate the implausibility of the Cartesian view. First he notes, we have no knowledge that such an entity exists, either from observation of others or from introspection. Secondly, even if we did have such evidence, we could not know that it persisted continuously. We could never know that it was not the case that it was destroyed every instant and subsequently replaced with a new one. Thirdly, its existence does not follow as a deductive consequence of our experiences as Descartes seems to have supposed. Descartes concluded that since he thought there must be an "I" which thinks. But, in fact, he was only entitled to conclude that, "This is a thought, therefore there is at least one thought being thought". Of course Descartes' mistake was a natural one given the way we talk, but his conclusion was not justified. Lastly and most damaging though, are the results of split-brain research discussed in Chapter 2. This evidence contradicts the unity of consciousness that is necessary for the Cartesian argument to go through. Any other alternative would have to deny the identity of the ego with our self-consciousness so that the ego itself would remain
unitary and undivided, despite the destruction of the unity of one's consciousness. But then this ego would be an empty hypothesis; it could not explain what it has typically been invoked for, viz, the unity of our self-conscious awareness and, ipso facto, the unity of our lives as persons. Quoting Parfit, "When the belief in Cartesian egos is cut loose from any connections with either publically observable or privately introspectible facts, the charge that it is unintelligible becomes more plausible." And this is not merely verificationism, as Parfit points out, but, rather, the reasonable assumption that we should reject any entity which we have no good reasons to believe in.

Since Parfit has conclusively proven that we cannot exist as any such separately existing entity and since there are good empirical reasons to believe that the brain is the cause of our psychological continuity, we have good reasons to believe that reductionism must be true. Returning to the Combined Spectrum case, we can see that the connectedness of our psychological and physical characteristics can hold to any degree and this implies that our personal identity is sometimes indeterminate, which in turn implies the truth of reductionism. I will now spell out the details of the theory.

---

\*\* Parfit 228.
Parfit begins with a distinction between psychological connectedness and psychological continuity,

Psychological connectedness is the holding of particular direct psychological connections.

Psychological continuity is the holding of overlapping chains of strong connectedness. ⑦⑧

Some examples will illustrate. A direct psychological connection between myself yesterday and myself today would be the action I perform today as a result of an intention I formed yesterday or, again, a memory of an event from yesterday. Now the distinguishing characteristic of connectedness is that it can hold to any degree. There may be potentially thousands of such direct psychological connections that could hold between myself yesterday and myself today, but certainly not all of them will hold in fact. For example, I cannot remember what I had for breakfast yesterday. This makes it difficult to say exactly what should count as enough direct connections to establish connectedness between myself yesterday and today. Let us say then, that, "there is enough connectedness if the number of connections, over any day, is at least half the number of direct connections that hold, over every day, in the lives of nearly every actual person". ⑦⑨ When there are enough such direct connections, then there will be strong connectedness. Psychological continuity then, will consist in the holding of

⑦⑧ Parfit 206.

⑦⑨ Parfit 206.
overlapping chains of such connectedness so that, even if there are no direct connections between myself now and myself 20 years ago, there would still be continuity.

Note that personal identity is a transitive relation however. If the senile old general is the same person as the brave young officer and the brave young officer is the same person as the boy who was flogged, then the boy who was flogged is the same person as the senile old general. But connectedness is not a transitive relation, for I can be strongly connected to myself yesterday without being connected at all to myself 20 years ago. So, strong connectedness cannot be the criterion of identity. Since continuity is a transitive relation, it must be the criterion of personal identity. Here then is Parfit's formulation of the criterion of psychological continuity:

(1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, 
(3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also continuous with Y. (5) Personal Identity over time just consists in the holding of facts like (2) to (4).\textsuperscript{80}

Regarding the qualification above, viz, "the right kind of cause", Parfit distinguishes three kinds of cause;

\textsuperscript{80} Parfit 207.
(1) The normal cause (i.e.: the continued existence of the same brain).

(2) Any reliable cause.

(3) Any cause at all.

These three causes correspond to three versions of the continuity criterion just presented, which are referred to as, the "Narrow", the "Wide" and the "Widest" version respectively. In the case of the Narrow version, the psychological criterion coincides with the physical criterion.

Parfit contends that both the physical and the psychological criterion are reductionist because both are committed to the claim that, the fact of a person's identity over time just consists in the holding of certain more particular facts and is in no way dependent on the existence of any separately existing entity. This in turn implies the claim that these facts can be described in a completely impersonal way. Using "event" to refer to all the events peculiar to persons, both physical and mental, Parfit states the reductionist position as follows,

A person's existence just consists in the existence of a brain and body, and the occurrence of a series of interrelated physical and mental events.\(^{61}\)

From which he concludes that a person just is a particular brain and body, and such a series of interrelated events.

\(^{61}\) Parfit 210.
Parfit concurs with the Widest version of the psychological criterion noted above though. To establish this he has us consider two cases of teletransportation. In the first, called simple teletransportation, I step into a booth, press a button, and then seem to momentarily lose consciousness. The machine reads the complete state description of my body as it disassembles it. This information is then beamed to a receiving booth on Mars where an exact duplicate of myself is created from new matter. On the narrow theory, I have committed suicide only to be replaced by an exact replica. On the wide version, it is unclear whether I have survived or not, since it is not clear whether or not we should call this a "reliable cause". And on the widest version, which is compatible with any cause, I have survived.

How shall we decide between the alternatives? Let us consider the second case of teletransportation, called the branch-line case. In this example, I step into the booth and push the button, whereupon I seem to lose consciousness for a second, all just as before. But this time, when I regain consciousness I am not on Mars, but still in the booth on Earth. Stepping out of the booth I am told that a new method has been devised of teleporting people such that their state descriptions can be read without disassembly and, hence, I can remain on Earth while an exact replica will carry out my business for me on Mars. It is hoped that I will enjoy the convenience of this new service.
Then, as I prepare to leave I am taken aside by the manager of the facility who tells me that there is a problem with the new system and, although there were no problems with the creation of my replica, I can expect to die in no more than several weeks from heart damage. Later, I am able to speak with my replica on Mars who tries to console me by assuring me that he will care for my wife and children, finish the book I am writing and so forth.

In the case of simple teletransportation, where I do not coexist with my replica, it is easier to believe that it is a simple method of travelling but, in the branch-line case, where we overlap, there seems to be good reason to believe that I will die, regardless of what happens to my replica. Parfit denies this claim however, and argues that being survived by an exact replica who is completely psychologically continuous with me is "about as good as ordinary survival". He argues for this on the grounds that, although we are not numerically identical, we are qualitatively identical and that this qualitative identity contains all that matters to our survival if we examine the facts dispassionately. Firstly, we have already agreed that psychological continuity is all that our personal identity consists in, so the only matter of contention here is the nature of the cause. Now secondly, Parfit argues, the only reason we are inclined to think that the continued existence of the same brain is necessary (i.e.: the normal cause) is because, in the

Parfit 201.
overwhelming majority of cases, psychological continuity coincides with numerical identity in this way. And lastly, as the case of split minds shows, we can have something which anyone would agree is about as good as ordinary survival and yet we do not, in this case, have numerical identity. For recall that we now have two people where previously there was one. Two persons cannot be numerically identical with one, so there is not numerical identity here. There is simply no other basis for this prejudice claims Parfit. So, all that matters to our personal survival is relation "R" which is defined as, "psychological connectedness and/or continuity with the right kind of cause.", where, "The right kind of cause could be any cause".\footnote{Parfit 215.}

But what if one should balk at this, claiming, "alright Parfit, I will agree with the wide version and accept that relation R is all that matters, as long as there is a reliable cause, but I cannot go so far as to accept that any cause will do". To this Parfit replies that we could imagine some unreliable treatment for a fatal disease such that, in most cases nothing is achieved but in a few cases there is complete recovery. The effect is just as good, even though its cause was unreliable. Parfit argues that it would be foolish for anyone to refuse the treatment since there is nothing to lose and perhaps much to be gained. The above case of teletransportation is analogous, "In
our concern about our own future, what fundamentally matters is relation R, with any cause".  

There is one final detail of Parfit's theory to be covered. It might be argued that, since his theory still relies on psychological continuity, it remains vulnerable to Bishop Butler's claim that it presupposes personal identity (recall that discussion of this point was deferred in Chapter 3). Parfit has a very slick response to this that involves the distinction between normal and quasi-memory. Quasi-memories include normal memories such that, "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,

and

3. my apparent memory is causally dependent, in the right kind of way, on that past experience."  

On this definition, ordinary memories are a subclass of quasi-memories. Specifically, they are quasi-memories of our own experiences. Now Butler's objection goes through because our normal concept of memory implies that we can only remember our own memories. But this is not the case with the broader notion of quasi-memory. It is logically possible that I should quasi-remember other people's experiences. For example, it might become

---

Parfit 215.

Parfit 220.
possible for neuro surgeons to implant other people's memories in anyone else's brain. Or, from a previous example we have considered, the replica of myself on Mars will quasi-remember my experiences. So, in his definition of relation R, Parfit's intended meaning is the continuity of quasi-memories, quasi-intentions and so forth.

Let us now summarize the substantive conclusions of reductionism,

(1) We are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events. Our existence just involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events. Our identity over time just involves (a) Relation R - psychological connectedness and/or psychological continuity, either with the normal cause or with any cause, provided (b) that there is no different person who is R-related to me.

(2) It is not true that our identity is always determinate. I can always ask, "Am I about to die?" But it is not true that, in every case, this question must have an answer, which must be either Yes or No. In some cases this would be an empty question.

(3) There are two unities to be explained: the unity of consciousness at any time, and the unity of a whole life. These two unities cannot be explained by claiming that different experiences are had by the same person. These unities must be explained by describing the relations between these many experiences, and their relations to this person's brain. And we can refer to these experiences, and fully describe the relations between them, without claiming that these experiences are had by a person.
Personal identity is not what matters. What fundamentally matters is relation $R$, with any cause. This relation is what matters even when, as in a case where one person is $R$-related to two other people, relation $R$ does not provide personal identity. Two other relations may have some slight importance: physical continuity, and physical similarity.

In the next chapter, we will discuss some problems with Parfit’s analysis and then move on to a comparison of reductionism with the M.N.A. theory considered in Chapter 7.

---

Parfit 216 - 217.
CHAPTER 9

Reductionism and the M.N.A. Theory: A Comparison

This chapter commences with a discussion of some general problems that plague reductionist theories and then moves on, in §2, to a comparison of reductionism and the M.N.A. theory of Chapter 7.

§1

Firstly, there is a problem with the conclusion which Parfit draws from his analysis of the results of split-brain research. He argues that, in such a case we have clear evidence that the unity of consciousness has been destroyed and yet we have something that is about as good as ordinary survival. It is evident that, in such a case, numerical identity is not preserved for it is impossible that the two resultant streams of consciousness should be numerically identical with the single stream which is their progenitor. From this Parfit concludes that numerical identity is not what matters in personal survival. Rather, all that matters is relation R, which amounts to the continuity of our psychological characteristics, with any cause. He then goes on to conclude that, therefore, we should regard being replaced by an identical replica as also about as good as ordinary survival for the two cases are analogous.
The problem here is that Parfit has failed to realize the significance of numerical continuity. Recall from Chapter 5, when discussing the Contingent Brain Constraint, that "numerical continuity" is a term intended to capture the notion that both streams of consciousness which result from split-brains are continuous with the single stream which is their progenitor specifically in the sense that they all share a common cause (i.e.: the same brain or at least a functioning portion of it capable of supporting consciousness). That is, each resultant half brain is numerically identical with the respective half of the original undivided brain. Hence, the cases are not analogous at all, for in the case of the replica, this numerical continuity does not hold. While the conclusion that numerical identity is not what matters in personal survival is correct, Parfit goes too far in concluding that being survived by an identical replica would be about as good as ordinary survival, for this ignores the significance of numerical continuity. The two cases are relevantly dissimilar to this extent, that in the case of the split-brain patient, the two streams of consciousness which are numerically continuous with the original single stream continue to "have" the experiences, but in the case of the replica, it is an entirely different (i.e.: numerically distinct and discontinuous) stream of consciousness "having" the experiences. A more forceful way of putting this is that, in the case of my replica, it is not me "having" the experiences of completing my book, loving my family and so on. Rather, it is someone else who
just happens to resemble me, both physically and psychologically. This is not about as good as survival but is death, absolutely. This will be dealt with at greater length below.

Lastly, there is a conflict on Parfit's theory with something that we should regard as at least logically possible, viz, the possibility that there could be two or more streams of consciousness that were qualitatively identical. On Parfit's theory this is logically impossible, since there are no means to distinguish them. This follows from the fact that Parfit relies on the qualitative nature of our experiences to determine ownership and, given this, it would be logically impossible to have two streams of consciousness that were qualitatively identical. Unless he can provide some other criterion for distinguishing various streams of consciousness, qualitative identity implies numerical identity. Further, it can be shown that Parfit could not, as a matter of principle, ever come up with any such criterion for this would violate the central claim of reductionism: that it is possible to give a complete description of reality without making any reference to persons because, "A person's existence just consists in the existence of a brain and body and a series of interrelated physical and mental events". Any additional criterion would mean that there would be a further fact to be taken account of when describing a person's existence; it could no longer consist solely in the

---

Parfit 210.
existence of a particular brain and body and a series of interrelated physical and mental events. This is not a problem for sentient being theories since what distinguishes a stream of consciousness on such a theory is its' unique location in p-space. We shall deal with this at greater length below when we compare Parfit's theory with the M.N.A. theory.

§2

In this section, Parfit's theory will be compared with the M.N.A. theory with respect to the constraints outlined in Chapter 5. Firstly, regarding the Contingent Brain Constraint, Parfit's theory clearly runs afoul of this. Recall that this constraint claims that retention of the same brain is causally necessary (where causal necessity is conceived of as a contingent relation) and sufficient for survival as the same sentient being, specifically because it is the normal cause of the stream of consciousness which is the sentient being. Parfit's theory rejects this in favour of the claim that all that matters to our personal identity is relation R with any cause. Further, he has arrived at this mistaken view because, as noted above, he has incorrectly appraised the results of split-brain research. Specifically, he has moved from the valid conclusion that numerical identity is not constitutive of personal survival to the mistaken conclusion that these results also prove that numerical continuity is irrelevant. But this has been shown to be false.
Parfit’s theory is able to satisfy the Transfer Constraint though. This, it will be recalled, is the constraint that it should be at least logically possible that consciousness could be transferred to another brain although it is almost certainly physically impossible due to the holding of certain natural causal laws (i.e.: retention of the same brain is causally necessary and sufficient for the occurrence of one’s stream of consciousness).

Parfit’s theory is also able to satisfy the Phenomenological Constraint, which maintains that, all that matters in survival is what is contained in one’s stream of consciousness. Similarly, both theories will agree with the Division Constraint. Clearly, the results of these experiments are crucial to Parfit’s position and he has certainly taken them into account.

Such agreement is not forthcoming with respect to the Temporal Constraint though, for at least two reasons. Firstly, for Parfit, experiences are not "owned" or "had" by any person or subject at all since, properly speaking, there are not persons but only particular bodies which have associated with them sets of interrelated physical and mental events. Secondly, even if he did admit to ownership of experiences in any sense at all, this would be determined by the extrinsic qualitative nature of the experiences and not by something intrinsic to the experiences.
themselves. This, as has been shown, is specifically the result of Parfit’s insistence on psychological continuity as the criterion of personal identity, for psychological continuity is cashed out in terms of the qualitative appropriateness of the experiences. Consequently, the problem for Parfit here is that for someone recovering consciousness, if the quality of his experiences are too rudimentary to link them to his earlier self, then they wouldn’t be his experiences unless they were followed by other experiences that were qualitatively appropriate in a way that would link them to his earlier self. Consider the following example. A man is asleep and is suddenly covered in boiling oil. It is easy to imagine that the only experience he would have the instant between recovering consciousness and dying would be sheer agony. On Parfit’s theory, these experiences of agony would not belong to the same person who went to sleep unless he survived long enough to also have experiences that were qualitatively appropriate to his former self.

Briefly summarizing our results so far then, we see that Parfit’s reductionist theory can only meet two of the five constraints which, as it was argued in Chapter 5, are necessary if we are to have a theory of personal identity that can square with our intuitions regarding what is involved in matters of personal survival. Further, it is incompatible with both of our additional constraints which, though less important, are still significant. These are the constraint that survival as the same
sentient being is an all-or-nothing matter and that experience ownership claims should have the same meaning with respect to tense. Regarding the first, Parfit's theory contradicts by default since he has not even recognized the distinction between persons and sentient beings. Still, the fact that survival is, in some sense, an all-or-nothing matter is an intuition most people have and he is not able to account for it at all. In fact, he blatantly contradicts this intuition claiming that, in some cases, the question of one's survival cannot even be answered; it is an empty question. Regarding the second, we may say that the contradiction is again by default since the essence of Parfit's theory is that there is no subject at all which "has" the experiences. It is worth mentioning though that, even if Parfit did claim that there was a subject of experiences, his theory would still contradict this constraint. This follows from the fact that he regards persons as constituted of a series of successive individuals who more or less closely resemble each other. Hence, it is perfectly compatible on Parfit's theory to claim that ownership need not be attributed to the same person at widely disparate times. As a result of these considerations then, it is here claimed that Parfit's theory is considerably less plausible than the M.N.A. theory expounded in Chapter 7. This claim will be further established in my concluding remarks which follow in the next chapter.
Let us return to the dilemma posed in Chapter 3 then. There it was claimed that, generally speaking, there are only two strategies one may take to explain what our personal identity consists in. Firstly, one might follow Descartes' lead and attempt to construct a theory of some sort of independently existing entity, construed as an ultimate particular or pure ego. This ultimate entity is conceived of as the true self, the self-reflective subject of experiences which remains unchanged throughout one's life. Parfit calls all such attempts "further fact" theories since this ego, qua ultimate entity, is a further fact in addition to the existence of one's brain and body and the interrelated series of mental and physical events that constitute all that is observable regarding persons. This strategy will not work as Parfit, Nagel and others have pointed out, because the identification of this ego with one's self-reflective awareness, which is necessary for the theory to explain the unity of our lives as persons, is incompatible with the results of split-brains and divided consciousness.

Secondly, one might take Locke's lead and attempt to frame the account in terms of the continuity of observable characteristics. Locke, convinced that the term "person" was a forensic term that identified moral agents, framed his account in
terms of the continuity of reflective consciousness which is usually established by memory. Initially, this seems intuitively plausible since our notion of person, which is closely tied to the necessary and sufficient conditions for moral agency, is usually cashed out in terms of these characteristics. Others have argued that since the brain is the cause of the characteristics that we take to be definitive of persons, it is the continuity of the body, or at least the brain, that is the basis of personal survival. Again there is intuitive plausibility to this conception for we usually identify persons by identifying their bodies. That is, in the overwhelming number of cases, we justify the claim that X at time $T_1$ is the same person as Y at a previous time $T_0$, by reidentifying the distinctive physical characteristics of Y on X; it is the physical continuity of persons that we normally employ to reidentify them. But we have also seen that no such alternative based on the continuity of observable characteristics can work either, for all such theories mistake the criteria for justifying claims we make about persons for what personal identity consists in. The criteria that justify our claim that someone has survived are epistemological criteria and cannot tell us what the metaphysics of our personal survival consists in.

The only other alternative is to follow Hume and acknowledge that our ordinary notion of personal identity is ill-founded and cannot be justified. It was Parfit's insight to take Hume's
theory to its logical conclusion and argue that, instead, personal survival is a matter of degree by showing that the kind of connectedness on which our psychological continuity supervenes is a matter of degree. Our analysis has revealed both what is right and what is wrong with Parfit’s reductionism. He is at least partly correct to assert that personal identity consists in the continuity of psychological characteristics. And, like Hume, he recognized the necessity of construing the relevant characteristics in the broadest sense, rather than limiting these to memory alone. Also, like Hume, he is correct to insist that the notion of causation must be brought into the account of our mental episodes as an explanation of what links these episodes together. And, as we have seen, Parfit is sensitive to the flaws in Hume’s analysis. He has properly recognized what eluded Hume, that a person, although constituted of a series of mental episodes linked together, can still be regarded as a single entity and that this entity need not remain unchanged to retain numerical identity. That is, numerical identity does not require qualitative identity since the properties of persons can be temporally indexed. So we are able to say of this person who "is now" large, that he is the same person who "was once" small.

However, in taking Hume’s analysis to its logical conclusion, he has remained within the empirical paradigm and, hence, his position remains vulnerable to the flaws plaguing all such theories. These have been discussed at length above and need
not be repeated here, but it is worthwhile to note that Parfit has committed some that are peculiar to his theory. Firstly, he has incorrectly assessed the results of split-brain research to conclude that since numerical identity is not relevant to personal survival, it follows that all that matters is relation $R$ (i.e.: psychological continuity defined in terms of his notion of connectedness) with any cause. From this he goes on to conclude that being survived by an identical replica would be about as good as ordinary survival. The latter conclusion, as we have seen, is mistaken precisely because this replica would not be numerically continuous with me. The normal cause of my stream of consciousness (i.e.: my brain) would have been destroyed and this is death. And lastly, we have seen that his position cannot admit the logical possibility of there being two qualitatively identical streams of consciousness. This is because he has provided us with no means to individuate persons except by the qualitative nature of their streams of consciousness. Given this, two qualitatively identical streams of consciousness would be indistinguishable, (this would not be a problem for any sentient being theory though since sentient beings are distinguished by their unique location in p-space). More importantly, he denies that, as persons, we are also subjects of experience. That is, he denies that we the sorts of things which "have" experiences. And, he denies that our survival is an all-or-nothing matter, a claim most people would probably find absurd.
In this thesis, I have argued that our personal identity consists in something ultimate: the further fact that we remain the numerically same sentient being. This is a further fact of what it is to be a person, but is not the same sort of further fact as a Cartesian ego. It is not conceived of as an ultimate particular, or an independently existing entity and this notion of sentient being is perfectly compatible with the results of split-brains and divided consciousness. Further, this theory allows us to salvage the intuitions that we are entities which "have" experiences and that our survival is an all-or-nothing matter.

There is also a Humean flavour to this theory though, for I have argued that as sentient beings we are essentially streams of consciousness or, as Hume would say, a "bundle" or succession of impressions. But we are now able to explain the dilemma which so sorely vexed Hume; what is it that actually "bundles" the impressions together? The answer, on the theory presented here, is that all the experiences of any person are bundled together simply in virtue of their all sharing the same p-location. The claim is not that the experiences are bundled by being contained within this p-location (as seems to be the case on Elliot's theory), but rather that they just occur "at" the p-location, in the same way that several people may share the same address. The argument, which assumes dualism (i.e.:that our experiences are
mental events distinct from the neuro-physical events which cause them), is that our experiences must occur somewhere. They are not located in physical space, so, ipso facto, they occur in p-space. Now since experiences are caused by brains, it is argued that p-space is similarly a concomitant effect of brains; it is just a consequence of there being experiences. Now we can be sure that all the experiences "at" any particular p-location will all be of the same sentient being since we have defined the term "sentient being" as, a stream of consciousness "at" a particular p-location. Further, given the causal dependence of experiences and p-space on brains, the same p-location will always be associated with the same brain and, by extension, with the same stream of consciousness. It is further argued that although both the stream and the p-location are causally related to the brain, they are only contingently related to each other. So it is a contingent fact that one's stream of consciousness occurs at a "particular" p-location; it is not contained there, it just occurs there.

Let us now return to Parfit and ask in what sense was he correct to insist that psychological continuity is at least partly constitutive of our lives as persons? The answer, as Locke realized, is that our lives as persons are intimately tied to the notion of moral agency. We need to be able to identify persons so that we can know whom to reward and whom to punish for actions that they have done in the past. So we need to be able to establish the attributes of persons that are crucial to being
moral agents and also what it is that connects "this" moral agent today with "that" one who performed such and such an action yesterday. But we also need to establish the characteristics of persons that remain relatively stable so that we can know what to expect from others around us. Now the very characteristics which we usually take to establish personhood (i.e.: moral agency) and which we usually look to to determine what "sort" of person someone is, are these psychological traits. And it is the continuity of these traits that allow us to predict how persons are likely to behave and, together with the physical characteristics, help us to identify persons across time.

But, as noted by Amélie Rorty, the criteria that we take to be definitive of personhood (i.e.: the necessary and sufficient conditions of moral agency) are at least partly a matter of social convention.

A society that focuses primarily on the sorts of actions that are thought to follow from rational choice will criteria for identification in the continuity of psychological traits believed to assure rational choice. If these capacities are sharply changed, the individual may no longer be thought to be a person, even though she is still regarded as the same human individual. ...

In any case, controversies about conditions for personal identity reflect differences about the conditions that establish an individual as a responsible agent.

---

Rorty 4-5.
It is for this reason that the criteria that establish personhood cannot be a description of what metaphysically constitutes our personal identity, since this metaphysical issue is a matter of ontology and not one of social convention. As Rorty points out, any analysis of persons in terms of necessary and sufficient conditions, "becomes legislative and normative, no longer straightforwardly analytically descriptive".87

---

87 Rorty 2.
Works Cited:


Sikora, Richard. Personal Identity, Psychological Location and Survival. unpublished manuscript.


INDEX

Actual world (2)
Ambiguity
  problem of (57)
Anaesthesia (67)
Ancestral relation (44)

Bodily exchange (46)
Brain processes (25)
Brain-dead (124)
Branch-line case
  the (153)
Brave Officer Paradox (43)
Bundles
  of experiences (73), (102)
  of impressions (61), (138)
Butler
  Bishop (45)

Cantilever Theory
  the (117)
Causal link (48)
Causal necessity (72)
Cause (25)
Cerebral hemispheres (27)
Coma (124)
Combined continuity theories (55)
Combined spectrum (76), (147)
Conceivability (19)
Concomitant effects (114)
Configuration (38)
Connectedness (143), (150)
Consciousness (67)
Constituent parts (40)
Constitutive (18)
Contingent (72), (81)
Continuity
  criterion of (42)
  physical (1)
    physical, psychological (32), (143)
    psychological (1), (151)
Corporeality (20)
Corpus callosum (27), (71)

Descartes
  René (11)
Discontinuous
  p-space (71)
Diversity (60)
Division of consciousness (89)
Dualism (12)

Ego (11)
Cartesian, pure (147)
Elliot
Robert (98)
Empirical
tHEME OF PERSONAL IDENTITY (1), (32)
Entity (11), (61)
Episodes (44)
Epistemological question
 OF PERSONAL IDENTITY (1)
Epistemology (59)
Essence (12), (18)
Essential nature (18)
Events
mental, physical (69)
Experiences
occurrence (73)

Flew
Anthony (45)
Further fact
theorem (12), (130)
Fusion (71)

Gaps
in consciousness, in p-locations (5), (105)
Grice
H.P. (23)

Haver-of-experiences (67)
Hume
David (32), (60), (138)

Identical
numerical sense, specific sense (141)
Impression (60)
Indestructible (12)
Indexicals (98)
Indivisible (11)
Initial experiences
argument from (118)
Interaction
mind/body (21)
Leibniz
  G.W. (69)
Locke
  John (1), (33)
Lockean memory theory (33)
Logical possibility (22), (80)
Loose identity (92)

Margalit (4)
  Avishai
Memory (42)
Metaphysical question
  of personal identity (1)
Metaphysics
  of personal identity (59)
Mind (17)
Modular
  theory of mind (29)

Nagel
  Thomas (28)
Non-qualitative, gap-crossing relation (5), (114)
Numerical continuity (82), (160)
Numerical identity (16), (61), (83), (138)
Numerically individuate (69)

Ontology (24)
Open-texture (4)
Open-texturedness
  problem of (57)

P-location (68)
Parfit
  Derek (7), (32), (142)
Penelhum
  Terrence (139)
Person (3), (41)
Personal identity (12)
Personal survival (12)
Personhood (13)
Physical continuity
  theory of personal identity (46)
Physical space (69)
Physical spectrum (146)
Possibility
  logical, physical (22)
Possible world (22), (88)
Principium Individuationis (37)
Psychological relations
  qualitative, non-qualitative (4)
Psychological space
  p-space (68)
Psychological spectrum (145)
Pure ego (23)

Qualitative nature (68)
Qualitative psychological continuity (43)
Quasi-memory (156)
Quinton
  Anthony (44)

Reductionism (7)
Reductionist theory (142)
Reid
  Thomas (43)
Relation "R" (155)
Relations
  qualitative, non-qualitative (99)
Replica (83), (159)
Resemblance (44), (144)
Rorty
  Amélie (172)

Scaling concept (8)
  of consciousness (125)
  of psychological connectedness (151)
Self-consciousness (23), (41)
Self-reflective awareness (23)
Self-same (16)
Senile general paradox (44)
Sentience (124)
Sentient being (3), (66)
Sikora
  Richard (5), (98)
Simple teletransportation (153)
Simultaneous components of experience (101)
Soul (24)
Sperry
  R.W. (27)
Split-brain (27), (148)
Stream of consciousness (68)
Strict identity (92)
Subject of experiences (67)
Substance
  immaterial, mental, unextended (12)
  material, physical, extended (12)

180
Substratum (17)
Succession
  of related objects (60)
  of related impressions (138)
Swinburne
  R.G. (1)

Temporal indexing (140)
Temporally coextensive (73), (114)
THE CONTINGENT BRAIN CONSTRAINT (80)
THE DIVISION CONSTRAINT (89)
The Minimal Neural Activity Theory (120), (159)
THE PHENOMENOLOGICAL CONSTRAINT (87)
THE TEMPORAL CONSTRAINT (89)
THE TRANSFER CONSTRAINT (84)
Total temporary state
  t.t.s. (44)
Transitive relation (43), (151)

Ultimate particular (23)
Unifying relations
  diachronic, synchronic (101)
Unity (24), (60), (138)

Verificationism (24)

Williams
  Bernard (46)