THE SURVIVAL OF SENTIENT BEINGS

by

JOHN MALYON INGLIS

B.Sc., The University of British Columbia, 1973
M.Sc., The University of British Columbia, 1974

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF PHILOSOPHY

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2000

* John Malyon Inglis, 2000
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
Vancouver, Canada

Date August 30, 2000
Abstract

This thesis is concerned with the metaphysical question of in what our survival consists. To survive in the sense to be explored herein is for there to be future experiences about which one should be egoistically (as opposed to altruistically) concerned. I focus on one's survival as a sentient being, a being capable of having experiences but possibly lacking some of the attributes of a person.

There are two categories into which metaphysical theories of survival fall: reductionist theories (theories according to which survival can be analyzed in terms of physical continuity and/or qualitative relations) and nonreductionist theories. I argue that theories of the former type conflict with our deeply-held and well-justified beliefs about our survival. Reductionist theories cannot explain why one should be specially concerned about certain future experiences.

If survival cannot be analyzed in reductionist terms, then we cannot fully refute skepticism about our survival over time. However, as nonreductionist alternatives to skepticism, I briefly consider the traditional substance view (according to which one's survival consists in the persistence of one's ego) and a view according to which one's survival consists in the continuation of one's stream of consciousness. I conclude by briefly enumerating the difficulties that this latter view would have to meet.
### TABLE OF CONTENTS

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

Acknowledgements ............................................................................................... vi

**INTRODUCTION** ................................................................................................. 1

Chapter 1: Survival and Special Concern ............................................................. 2
  1.1 Egoistic Concern and Doubts About It ......................................................... 2
  1.2 Why Special Concern Should Be Taken Seriously ...................................... 8
  1.3 Survival, Death, and Teletransportation ...................................................... 20
  1.4 Survival and Identity .................................................................................. 25
  1.5 Survival as a Sentient Being ...................................................................... 28

Chapter 2: Metaphysical Problem of Survival ..................................................... 32
  2.1 Epistemological and Metaphysical Problems of Survival ......................... 32
  2.2 Reductionism ............................................................................................. 35
  2.3 Reductionism and Survival ........................................................................ 40
  2.4 The Qualitative Approach to Survival ..................................................... 54
  2.5 The Physicalist Approach to Survival ...................................................... 57
  2.6 Methodology: Deep Beliefs and Thought Experiments. ......................... 60

PART I: Qualitative Theories ................................................................................ 66

Chapter 3: The Qualitative View ......................................................................... 67
  3.1 Introduction .................................................................................................. 67
  3.2 Locke’s Memory Theory ............................................................................. 68
  3.3 Problems for the Lockean View .................................................................. 73
  3.4 Transitivity .................................................................................................. 75
  3.5 Circularity .................................................................................................... 77
  3.6 Objections to Parfit’s Response to the Circularity Problem ....................... 79
  3.7 Identity Is Not What Matters ...................................................................... 83
  3.8 Objections to the View That Identity Is Not What Matters ....................... 87
  3.9 Causal Requirement .................................................................................. 89
Acknowledgements

I am able to acknowledge herein only a few of the numerous contributors to whom I am greatly indebted for the completion of this thesis. I would like to thank Kwantlen University College for supporting my research and generally being an excellent employer. I also owe a great deal of appreciation to the members of my supervisory and examining committees: Professors Paul Bartha, Michael Chandler, Owen Flanagan, Phil Hanson, Alan Richardson, and Carlo Testa. Among the friends that have patiently endured my ramblings, I would like to thank especially Ron Johnson, Fred Jones, and Graham Rankin for their interest in my project and their valuable insights. I am unable to adequately express my gratitude to my research supervisors, Professors Richard Sikora and Gary Wedeking, without whose guidance, inspiration and tireless efforts, this project never would have been completed. Finally, completion of this thesis would also have been impossible without the support of my daughter Sarah, my son Nicholas, and especially my wife Elizabeth. I greatly appreciate their patience.
1. Egoistic Concern and Doubts About It

Most of us feel a great and special concern about our anticipated future experiences. If you learn that someone is to experience great pain or anguish, then you may regret that this person is to suffer so greatly. However, when you learn that the person who will suffer this pain is yourself, your reaction will be quite different and perhaps more intense. Though it may be reasonable to say that one feels such special concern about the subject of the experiences in question (i.e. oneself), what one is truly specially concerned about is what the experiences of this subject will be like. That is, it is the future experiences of this subject that are the object of this special concern. No matter how deeply concerned one may be about someone else's future experiences, when the experiences in question are to be one's own, one's concern seems to be of very different nature. Thus, there seems to be a clear dichotomy between egoistic (special) concern (for one's own experiences) and altruistic concern (for the experiences of others).

I believe that we are inclined to suppose that we should be, in the sense that we have good reason to be or that it is somehow rational for us to be, egoistically concerned about our future experiences. That is, we are inclined to suppose that there is a rational basis for such concern (rather than that we just happen to
feel such concern). Indeed, self-concern has often been taken as the very paradigm of true rationality. As Sidgwick wrote, "...my feelings a year hence should be just as important to me as my feelings the next minute, if only I could make an equally sure forecast of them. Indeed, this equal and impartial concern for all parts of one's conscious life [i.e. all of one's future experiences] is perhaps the most prominent element in the common notion of the rational." (Sidgwick, 1907, p.124; quoted in Parfit, 1984, p.313) (I explore this rationality of egoistic concern in more detail later in this chapter.) For convenience, I will sometimes speak of this concern as justified, but I intend "justified" as shorthand for "has such a rational basis" and not to imply that this concern is justified in the way that, for instance, a belief may be justified. Whether or not it is demonstrable that such concern is ever justified in this sense, I think that most of us (it is the philosophically sophisticated, I suspect, who are more likely have doubts) are strongly inclined to see it as such. We naturally suppose that from one's present point of view, there are two sorts of future experiences, ones for which a special (egoistic) concern that one may have (that accompanies one's present experience) is justified (i.e. one's own future experiences), and ones for which this concern is not justified (i.e. the future experiences of others), ones for which at most an altruistic concern is justified.

We tend to direct these feelings of special concern towards future experiences to be associated with our body. Indeed, we tend
to believe that these feelings are justified for the experiences to be associated with the same body as our present experience. However, many of us stop short of the unequivocal belief that the experiences to be associated with our body are exactly the ones about which we should be specially concerned. Depending on one's religious beliefs, one may also think that there may be experiences following the death of one's body about which one should be specially concerned. Also, when asked to imagine future experiences that are not associated with one's body but contain apparent memories of one's present experience, one may think that one's feelings of special concern for such experiences are justified.

However, on reflection, the whole notion of egoistic concern (and that it may be justified) may seem highly problematic. It seems doubtful that one's feelings of supposed egoistic concern are really so different in kind or degree from feelings of concern for the experiences of others. Feelings of concern for future experiences involve a range of emotions, including fear, indifference, and hope. As any parent can attest, one may have this same range of emotions for the future experiences of one's child as for one's own future experiences. Indeed, the feelings of concern (in terms of the emotional and other cognitive states involved) for the experiences of one's child may be stronger than for one's own experiences. Hence, the distinction between egoistic and altruistic concern cannot be accounted for in terms of degree or intensity of concern. In the case of supposed egoistic concern, the person who will have the experiences is just someone who (like one's child)
one may care about deeply and closely. Perhaps the only difference is that feelings of egoistic concern for future experiences are accompanied by the belief that these experiences will be one’s own.

There is an obvious evolutionary explanation of why we feel so concerned about the future experiences to be associated with our body. In order to survive, we need to attend to not just the immediate needs but also the future needs of our body. For us to survive, it is not enough for us to seek food when we feel hungry and to seek shelter when we feel cold. Even when we do not feel these immediate needs, we must concern ourselves about future times when we may. If we do not plan ahead, it is unlikely that we will be able to meet our future needs. Hence, it is clearly a selective advantage to be concerned about the future experiences associated with our body. If we did not feel greatly concerned about what these experiences will be like, then it is hard to understand how our species could have survived.

There is also an obvious explanation of how we acquire the concept of egoistic concern for future experiences and how we come to believe that we should care specially about future experiences to be associated with our body. Right now I remember having past experiences and feeling specially concerned about my experiences at what is now the present. My present ostensible memories have an aspect (what I will call the indexical aspect of ostensible memories) that represents these experiences and feelings as my own. It is not that this aspect depicts the experiences as involving one’s body (even though it may well to do so) but that it depicts
oneself as the subject of these experiences, just as one is the subject of one's present experiences. Thus, the subject of the earlier experiences seems to have had good reason to be specially concerned about these present experiences. We naturally think that feelings of special concern that accompanied experiences (at a given past time) that we seem to remember were justified for our present experiences. Because we tend to remember these past experiences as involving the same body as our present experience, rightly or wrongly, we also tend to feel such special concern about experiences to be associated with our body in the future. That memory has a role in our acquiring our notion of egoistic concern explains why some find it attractive to suppose that we should care specially about experiences that contain apparent memories of our present experience even if these experiences are not associated with the same body.

However, even though our memories of having past experiences may help to explain how we acquire the concept of egoistic concern, they certainly do not provide conclusive evidence that this concern is justified. Because one's ostensible memories are fallible (i.e. they may not faithfully depict past experiences or even represent past experiences at all), one cannot be certain that feelings of special concern that accompanied an earlier experience for one's present experience were justified just because one seems to remember the experience. Even if one's ostensible memories do accurately represent earlier experiences, the indexical aspect thereof may be false. Even if these memories correctly depict the
earlier experiences as involving one's body, it may be a mistake to think of the subject of the earlier experiences as oneself (rather than just the subject of the experiences then associated with one's body) and hence as someone who had reason to be specially concerned about one's present experiences. Perhaps it is evolutionarily programmed that our memories should have this indexical aspect. That our memories have this aspect certainly does not establish that egoistic concern is ever justified. Perhaps, then, there is no reason to be specially concerned about future experiences.

If the distinction between egoistic and altruistic concern is problematic, if we can explain (ultimately in terms our neurophysiology as shaped by natural selection) why we have these supposed feelings of special concern, and if there is no compelling case for the commonsense belief that these feelings are ever justified (in the sense at issue), then it seems irrational to cling to the notion of egoistic concern. Perhaps it is simply a fact that we anticipate so emotionally experiences that we think will be our own. Special concern for future experiences is merely an affect that we are disposed to have, and as such it is no more reasonable to expect there to be a deep reason for this concern than for any other affect.¹ A parent may care obsessively about her child, but we don't expect there to be such a rational basis for this caring. Why, then, should we expect there to be such a basis for our feelings of special concern? Perhaps the belief that these

¹ Alan Richardson raised this point in a written communication.
future experiences will be our own (that we will be specially involved in these experiences), that distinguishes feelings of egoistic concern, is simply a delusion. Of course, there may be relations in terms of physical continuity and/or memory between our present experiences and certain future experiences. However, these are our only special relations to the future experiences, and they don’t give us a reason to care specially about these experiences.

2. Why Special Concern Should Be Taken Seriously

Because my thesis is that relations involving physical continuity and/or memory (and/or other qualitative relations) do not give us a reason (in the above sense) to care specially about future experiences, if egoistic concern is such a hopelessly flawed notion, then there is no point to my thesis. Fortunately, a plausible case can be made that we (philosophers as well as people in general) deeply believe that egoistic concern is justified and that sometimes it may actually be justified.

Suppose that oneself and someone one cares about deeply are being tortured. Certainly, one may be just as distressed about this other person’s pain as about one’s own. However, one’s access to one’s own present pain is very different from and more direct than one’s access to the pain of this other person. One feels one’s own

2 Though most people deplore physical pain as much as I do, masochists may even put a positive value on such pain. If one does not find the prospect of physical pain so awful, one can think of pain in the more general sense as an experience to which the subject would assign an extreme negative value.
pain but not the pain of the other person. Generally, one's relation to one's present experience is very different from one's relation to the present experience of anyone else. Arguably, it is because of this special relation to one's present experience that one has good reason to be specially concerned about this experience. Our philosophical views aside, I think we can easily understand why one would feel such concern. I find it hard to accept that concern for one's present experience is a merely contingent affect, that we should just happen to feel this concern.³

Now suppose that one's own body and the body of this other person will at any moment be subjected to torture. A person who knows that his body is about to be tortured anticipates a special (and clearly unwanted) relationship to the ensuing experiences. It is not that the emotions regarding the experiences (that one thinks of as one's own) soon to be associated with one's body and the experiences soon to be associated with the other person's body are essentially different. Rather, one's relation to the former experiences is essentially different from one's relation to the latter experiences. Arguably, it is such a special relation to the future experiences that justifies one's special concern for these experiences. It is because of this special relation that one's

³ Though I consider these claims about concern for one's present experience to be entirely reasonable, I realize that some may find them contentious. I remind the reader that I am not claiming that this concern is justified, only that it is reasonable to see it as such.
concern for one's own experiences is of a different species (i.e. egoistic rather than altruistic) than one's concern for anyone else's experiences. Though one's concern for others, such as one's children, may be greater than one's concern for oneself, egoistic concern seems justified in way that altruistic concern (in itself) is not.

Locke seems to have accepted that there is such a relation that justifies concern for our future experiences. He wrote:

This every intelligent Being, sensible of Happiness or Misery, must grant, that there is something that is himself, that he is concerned for, and would have happy; that this self has existed in a continued Duration more than one instant, and therefore 'tis possible may exist, as it has done, Months and Years to come, without any certain bounds to be set to its duration; and may be the same self, by the same consciousness, continued on for the future. And thus, by this consciousness, he finds himself to be the same self which did such or such an Action some Years since, by which he comes to be happy or miserable now. (Locke, 1975, II, xxvii, 25)

That is, we cannot intelligibly deny the existence of the self as the object of our sense of self-concern. Because this self may persist over time, one may have reason to be concerned about one's self in the future. One has reason to be concerned about whether this self will be "happy or miserable." As I think this talk of happiness and misery shows, he thought that what one should be really concerned with is the quality of the life of the person about whom one is concerned. Clearly, this quality of life depends on what this person's experiences will be like. Thus, one has reason to be concerned about one's future experiences (i.e. the experiences of one's self in the future). One should care about these experiences because they will be experiences of one's self.
Present actions that serve to enhance the quality of one's future experiences (that will make one happy rather than miserable) may be rational even if these actions involve some sacrifice of the quality of one's present experiences. Clearly, this concern for oneself and about the future experiences of oneself should count as egoistic (special) concern.

The belief that Locke seems to share that we should be specially concerned about our future self (and thus about experiences that are to be ours) is a component of a commonsense view that many of us on reflection hold about ourselves and our survival. Thomas Nagel articulates another aspect of this view as follows.

My nature...appears to be at least conceptually independent not only of bodily continuity but also of all other subjective mental conditions, such as memory and psychological similarity. It can seen, in this frame of mind, that whether a past or future mental state is mine or not is a fact not analyzable in terms of any relations of continuity, psychological or physical, between that state and my present state. The migration of the self from one body to another seems conceivable, even if it is not in fact possible. So does the persistence of the self over a total break in psychological continuity - as in the fantasy of reincarnation without memory. (Nagel, 1986, p.33)

Whether or not an experience will be mine appears to be a determinate matter. Moreover, "being mine is [or at least seems to be] an irreducible, unanalyzable characteristic of all my mental states..." (Nagel, 1986, p.34)

According to this commonsense view, there is a relation to one's future experiences that is conceptually independent of bodily continuity and qualitative psychological continuity (i.e. does not depend conceptually on what one will be like or on the contents of
one’s experiences) and that justifies one’s special concern for these experiences. Nagel implies that we are inclined to suppose not only that there is this conceptual independence but also that whether or not a mental state will be ours cannot be analyzed (and hence does not consist) in these continuities. That is, this relation is ontologically or metaphysically independent of these continuities. We believe that we may be thus related to experiences even after we have greatly changed, and that it is at least intelligible to suppose that we may be thus related to experiences after the death of our body. However, as Nagel realizes, it is consistent with this view that there being such future experiences may be causally dependent on the continued existence of our brain. (Nagel, 1986, p.34) According to Nagel, Derek Parfit thinks that we are in the grip of the “Simple View,” a view that accords well with this commonsense view. (Nagel, 1986, p.43) I think Nagel’s (and Parfit’s) depiction of our commonsense view of ourselves (the reflective plain man’s view) is basically correct.

Of course, because we are unable to directly apprehend this relation, this commonsense picture could be completely wrong. I certainly do not intend to defend this picture in this thesis. Though most contemporary philosophers would reject parts of this picture, few seem to reject it outright. For instance, though Parfit thinks that there is a relation to certain future experiences that justifies one’s special concern for these experiences and that this relation is independent of bodily continuity, he believes that this relation depends on the contents
13.

Of course, one may suspect that, if there is no relation of the sort that the above commonsense view posits, then there is no special relation that justifies egoistic concern. Perhaps philosophers should discard the notion of egoistic concern along with the rest of this largely discredited view. However, most commentators on personal identity and survival write as if they accept this notion. As far as I know, Eddy Zemach is the only contemporary philosopher who unequivocally denies that anything like egoistic concern is ever justified. (Zemach, 1987)

One reason why some may find the view that Nagel articulates unattractive is that on this view the nature of the self is mysterious and obscure, as is the relation of one’s present self to the future experiences about which one should be specially concerned. However, there may be things with which we are in part directly acquainted that satisfy the above requirements and make it much clearer how one’s special concern for the future experiences could be justified. I believe that because of one’s direct, privileged, and indeed intimate access to one’s present experience, one should be specially concerned about what this experience is like. I also believe that one can directly apprehend and hence

\[4\] At least, so he claims in Reasons and Persons. (Parfit, 1984, Chapter 14) However, in a later work, he accepts (rightly I think) that his view that survival consists in such a relation “undermines our concern about our own future.” (Parfit, 1998, p.315)
cannot doubt the unity of one's experience at a time. Indeed, one's feelings of concern may be unified in one consciousness with the other parts of one's experience.

Because an experience must be temporally extended (i.e. have duration), an experience has at least a brief unity over time. One can imagine how there can grow out of this unity of an extended experience a unity of temporally separate experiences. It seems reasonable to suppose that while one is conscious continuously one has a succession of overlapping experiences such that some part of a given experience is involved in the next experience. Because of this overlap, the unity of one experience extends into the next experience, and so on. Even though we can't directly apprehend this relation, there may be a transtemporal unity relation between temporally separate experiences that are linked by continuity of consciousness in this way.

There is a case that one is justified in being specially concerned about future experiences that will be transtemporally

---

5 Brentano and Chisholm clearly agree. (Brentano, 1973; Chisholm, 1987) Berkeley saw the unity of consciousness as an obvious datum for which we need to account. He did so by appeal to a mind, spirit or soul that perceives the individual perceptions or ideas that belong to this unity. (Berkeley, 1965, "Third Dialogue")

6 Russell held that successive overlapping experiences can encompass common "sense-data." (Russell, 1984, pp. 66-70)

7 Admittedly, not everyone finds it obvious that there is such a relation. According to Mark Heller, there is no unity among even successive experiences associated with the same brain. Though such experiences may be related in terms of their contents and causal relations, they are essentially no more connected than are successive experiences associated with different brains. (Heller, 1990, pp.23-25)
unified in this way with one's present experience. It seems absurd that my concern for, say, my present visual experience $V$ would be justified without my concern for, say, my present auditory experience $A$ (i.e. an experience that is unified with $V$ in belonging to the same total experience at a time) also being justified. (Suppose that I find neither $V$ nor $A$ enjoyable.) This is not to deny that if I am an accomplished and keenly visually sensitive artist who also happens to be hopelessly tone deaf, then I may be in some way much more concerned about $V$ than about $A$ or at least that I am likely to react more strongly to $V$ than to $A$. However, this is just to state the obvious: that my valuation of experiences may vary greatly. On the other hand, my relationship to both $V$ and $A$ (in terms of the direct access that I have to these experiences) is the same. If these experiences or perhaps my values were to change, then $V$ could seem comparatively inoffensive to me and $A$ could seem as cacophonous to me as heavy metal to Mozart. It is certainly hard to accept that I could have reason to be concerned about $V$ but no reason to be concerned about $A$.

It is almost as hard to accept that my present concern for $A$ would not be justified if $A$ were slightly later than $V$ but overlapped $V$. It seems reasonable that one's present concern would be justified for a future experience that overlaps one's present experience in the above way. This justification for concern should extend by transitivity to an experience that overlaps this experience, and so on. This transtemporal unity of consciousness gives one the sort of intimate but delayed access to the future
experience that justifies one's special concern for this experience. One may object that this suggestion does not explain how one's feelings of egoistic concern can be justified for future experiences separated from one's present experience by a gap in consciousness. However, here I am only trying to establish that there is a case that special concern could sometimes be justified.

Given that one can directly apprehend the unity of one's experience at a time, there is a reasonable explanation of why we have the package of beliefs (centering on the notion of egoistic concern) that Nagel articulates, in addition to that having such a package of beliefs is selectively favourable. This unity presents itself as independent of the body with which its component experiences are associated and of any qualitative psychological relations (such as similarity) between these experiences. We may reasonably assume that the transtemporal unity that grows out of this unity is likewise independent of bodily and psychological continuities. Perhaps we believe that whether or not we will have a given future experience is (and the indexical aspect of our memory presents our having had an experience as) a determinate

A whole thesis could be written on this transtemporal unity of consciousness. Obviously, the nature of this transtemporal unity is, in itself, an important issue. Here, I am offering only the briefest sketch of how we might establish this transtemporal unity and how it might justify egoistic concern. A number of the assumptions that I make about the unity of consciousness are, no doubt, highly contentious. I do not argue for these assumptions in this thesis. I consider the theory that I am suggesting here to be (like the reflective plain man's view) a mere proto-theory of how egoistic concern could be justified.
matter because there is no other way to represent our having an
type of our direct and intimate access to
our temporally extended present experience. Thus, the unity of our
experience (and our limited access to this unity) provides some
reason why we should have the above package of beliefs. Given that
these beliefs may have a rational basis (rather than just a
naturalistic explanation), these beliefs, and in particular the
notion that egoistic concern may be justified, are not so
unattractive as some suppose.

As I have shown, we can envision metaphysical proto-theories
(the reflective plain man's view and the view that there is an
extended unity of consciousness that justifies egoistic concern)
that could explain how egoistic concern can be justified. (I stress
that I am not claiming that either of these proto-theories or any
theory that I can offer adequately explains how egoistic concern
could be justified.) The debate on personal identity is also
indicative of the expectation that special concern may be
justified. Indeed, this debate has been and continues to be shaped
in large measure by our intuitive commitment to egoistic concern.
(See Unger, 1990, pp. 223-230, and Parfit, 1984, Chapter 14, on
self-concern as it may pertain to personal identity and survival.)
As we shall see in Chapter 4, the reason that the qualitative view
may seem unattractive is that it is hard to see how this view can
explain how this concern can be justified. As we shall see in
Chapters 4 and 9, there is no reason to entertain physicalism
unless we think that it is able to do so. Likewise, there is
absolutely no rationale for adding a causal requirement (discussed in Chapters 3 and 4) to a qualitative theory unless it helps to explain how such concern can be justified. (I outline the views just mentioned in the next chapter.) Though we don’t expect a deep reason to care greatly about the experiences of our children, we do seem to expect such a reason to care specially about our own experiences. Even if the conclusions of this thesis incline us to see this concern as something that we should not expect to be justified, I would be fully satisfied so long as it is recognized how dramatically rejection of this notion should change the direction of this debate.

Some philosophers may grant that egoistic concern may be justified but deny that egoistic concern is intelligible in the absence of qualitative (i.e. content-dependent) psychological continuity and/or bodily continuity. For instance, some may think that it is incoherent to care specially about future experiences that are not accompanied by memories (presumably ones that are sufficiently accurate) of one’s present experiences and concerns. Likewise, some may think that it is incoherent to care specially about future experiences that are not associated with the same body as one’s present experience. Though I reject both of these claims, I am not presently arguing against either of them. However, given the complexity of the issue and the apparent room for disagreement about what is and what is not coherent, I do think it is grossly premature to accept either of these claims. Arguably, one can be concerned for oneself and one’s future experiences (and maybe
justifiably) without appreciating the nature of the self. That we are unable to refute the skeptic who may doubt that egoistic concern is ever justified shows that it is far from obvious what supposed possibilities do and do not make sense.

For this reason, I think that it makes sense to remain agnostic for now about the circumstances in which egoistic concern is and is not justified or even intelligible. Most neutrally, whether or not egoistic concern for a future experience is justified seems to involve a relation of one's present self (or person stage) to the future experience (or perhaps a relation between one's present experience and the future experience), the exact nature of which in terms of memory, the association of the future experience with the same brain as one's present experience etc. remains to be explored. Even though it may not be clear to exactly what this special concern may be justifiably directed, we may say loosely that one (or perhaps one's present person stage, or one's present experience) is related to the future experience in this way if one has reason to be specially concerned about the future experience. That is, one is thus related to the future experience if any special concern that may accompany one's present experience for the future experience is justified.
3. Survival, Death, and Teletransportation

In a thought experiment envisioned by Derek Parfit, future interplanetary travellers can avoid the long trip to Mars in a conventional spacecraft by opting to be teletransported to the planet. (Parfit, 1984, pp.199-200) When the traveller pushes a button, his brain and body are scanned, the precise atomic configuration is recorded, the original brain and body are destroyed, and the recorded information is transmitted to Mars at the speed of light, where an exact replica is constructed particle for particle on the basis of this information. If one’s psychology (including one’s ostensible memories) is determined by the state of one’s brain, then the replica should be psychologically indistinguishable from the original. The person will emerge on Mars with an apparent memory of just having stepped into the teletransporter on Earth and will feel relieved that the procedure has worked so well. Consideration of transportation raises an obvious question: would one really survive?

Perhaps it is difficult to assess whether or not one survives teletransportation simply because it is unclear exactly what it means to survive. If one is talking about the continued existence of someone with one’s exact same physical and psychological traits (and especially one’s memories), then one does survive. However, if one is talking about the survival of one’s original brain and body, then one does not survive. However, like me, one may think that this is neither a satisfactory nor a complete answer to the question.
In order to determine whether or not one survives, it helps to compare this scenario with two other cases about which most of us have clear reactions. In a booth on the left of the teletransporter, when one pushes a button, nothing happens. This is a case of ordinary survival. In a booth on the right of the teletransporter, when one pushes a button, one's body is painlessly and instantaneously vapourized. This is a case of ordinary death. I think we have a vivid appreciation of the difference (from the point of view of the subject) between ordinary survival and ordinary death. When one pushes the button in the right booth, one’s experience abruptly and irrevocably ceases. As a matter of fact, there will be no more experiences about which one should be specially concerned. However, after one pushes the button in the left booth, as a matter of fact, there will be experiences about which one should be specially concerned. I find it very hard to accept that the belief that there is such a difference is delusional. What we want to know is whether teletransportation is like ordinary survival or ordinary death. We want to know if there will be future experiences about which the traveller should be specially concerned.

In this thesis, I am interested in a sort of survival that can be clarified in terms of egoistic concern. For one who has an earlier experience E to survive in this sense to have a later experience F is for one’s feelings of special concern for F (feelings of special concern for F that may accompany E) to be justified. For one to survive in this sense is for there to be
future experiences about which one has reason to care specially. I consider this to be an explication of our commonsense notion of survival rather than a definition of an exotic sense of survival. I believe that we tend to have a clear and fundamental appreciation of what it would mean for one to survive in this sense through a sequence of events. Though the traveller about to push the button in Parfit's teletransporter may not be sure whether or not he will survive, he should have a clear appreciation of what it means to survive this procedure. Are his feelings of special concern for the experiences to be associated with the body on Mars justified?

If egoistic concern is never justified (not even in the case of ordinary survival), then it is a mistake to wonder whether teletransportation would be like ordinary survival or ordinary death. There are no significant facts about teletransportation beyond that it destroys one's original brain and body but preserves their structure and presumably one's distinctive psychology. Then, likewise, there are no significant facts about ordinary survival beyond that it preserves both of these nor about ordinary death beyond that it preserves neither of these. Preservation of one's brain and body (physical continuity) and preservation of one's psychology (psychological continuity) do not provide a rational basis for caring specially about the ensuing experiences. In one's

---

9 I do not deny that there are other legitimate and commonsense (though perhaps less interesting) senses of survival. The continuance of someone with one's characteristics and the survival of one's original brain and/or body (perhaps in a condition to support consciousness) are two of them.
survival, physical and/or psychological continuity are "as good as it gets."\textsuperscript{10}

If such is the case, then whether or not one survives would seem to be a matter of convention, perhaps of use of language. Because we have not had to deal with cases of teletransportation, not surprisingly, just what that convention would be about teletransportation has not been decided. Perhaps, one day it will suit our purposes to treat teletransportation as falling under the term "survival" or under the term "death" or perhaps as not clearly falling under either of these terms. Whether or not one survives in the cases of ordinary survival and ordinary death is just as much a matter of convention. However, because we frequently have to deal with these cases, the convention is much better established. Though it may make sense to decide the identity of certain artifacts (and perhaps other objects) by convention, I think most of us would find it very hard to accept that our identity or survival is a matter of mere convention or decision. Indeed, it seems almost unintelligible to suppose that questions of personal identity or survival could be conventional. Whereas the identity of artifacts may sometimes be open to alternative conventions, there seems to be a fact of the matter regarding our identity and

\textsuperscript{10} I am indebted to Gary Wedeking for this expression.
survival that precludes convention.\textsuperscript{11} To render a decision about the range of cases in which one survives (in terms of, say, physical and/or psychological continuity) is to decide something, which, we are apt to suppose, instead needs to be uncovered. Could whether or not we survive to have (and have reason to be specially concerned about) future experiences really be determined by a choice about how to use language? Could a reversal by the French Academy (notorious for its linguistic dictates) regarding verb usage somehow change the facts about whether or not one has survived? Conventionalism with regard to the identity or survival of persons seems very unpalatable.

Not surprisingly, few contemporary philosophers admit unequivocally to holding that survival is a matter of convention.\textsuperscript{12} Bernard Williams finds it difficult to accept that it could be a matter of convention (and indeed anything other than a determinate all-or-nothing matter) whether or not one will have a given future experience. (Williams, 1975, pp.193-196) For Parfit, though identity may sometimes be a matter of convention, survival is

\textsuperscript{11} In the famous case of the Ship of Theseus, there are arguably two contradictory yet equally reasonable conventions regarding the identity of this artifact over time. Are we to consider the ship formed by the gradual replacement of this ship’s planks or the ship later constructed from the discarded planks to be the same as the original? (See Hirsch, 1982, pp.68-70 for a discussion of this case and Chisholm, 1998, pp.68-70, for discussion of how such a case invites convention. In its modern form, this case can traced back to Thomas Hobbes. (Hobbes, Concerning Body, chap.2, sect.7))

\textsuperscript{12} Heller holds that whether or not successive experiences are experiences of the supposedly same person is a matter of convention. (Heller, 1990)
generally not. (Parfit, 1984) (I will explain the distinction between identity and survival shortly.) If we clearly renounce the notion of egoistic concern, then the study of personal identity should begin with the acceptance that questions of survival are a matter of convention. Philosophers who think of egoistic concern as unintelligible and/or unjustifiable should purge their writing of the appearance of treating this flawed notion and accept that theirs is a conventional, pragmatic, and, I think, quite uninteresting task.

Until this notion may be generally rejected and the conventionalist consequences clearly accepted, it is worthwhile to explore whether or not theories of personal identity and survival can explain how egoistic concern may be justified. Though it is only a plausible hypothesis that egoistic concern is justified, given the present facts, being able to account for the justification of special concern is a reasonable requirement of adequacy for any theory of survival.

4. Survival and Identity

We may think that one who has an earlier experience E survives to have a later experience F because the one who has experience E is the same person as the one who has experience F. However, in extraordinary circumstances, we may question whether survival and identity coincide. Following the surgical removal of either cerebral hemisphere, a brain still supports consciousness. Though I am not suggesting that such a procedure can ever be technically
feasible, one can imagine a much more sophisticated procedure in which the healthy cerebral hemispheres are excised from the original body and transplanted into two decorticate clones. Perhaps in this case one should be specially concerned about and one thus survives to have the experiences associated with both of the resulting bodies.

Hence, it is not clear at the outset (at least not, it seems, to everyone) that a person X at an earlier time cannot survive as (can justifiably be specially concerned about the experiences of) two or more persons Y_1, ..., Y_k at a later time. However, it is clear that X cannot be identical to more than one of these. A relation of personal identity is by definition an identity relation and hence must have the formal properties of such a relation, namely reflexivity, symmetry and transitivity. It follows that an individual at one time cannot be the same person as more than one individual at a later time, because otherwise the absurd conclusion would follow that these latter individuals would be identical to each other. Hence, in such a case, identity and survival may not coincide.

If survival and identity do not always coincide, the issue of survival in my sense is the more fundamental. If one's identity as a person is to be at all meaningful, then reasonably X cannot be the same (identical) person as Y unless X survives as Y (i.e. unless X is justified in caring specially about the experiences of Y). Moreover, in a case in which X survives as more than one person Y_1, ..., Y_k, the question of which, if any, of these is the same
person as X would have to be settled (unlike the question of which persons X survives as) either arbitrarily or by convention.

One may object that if egoistic concern is understood as concern for oneself and one's own experiences, then it is incoherent to suppose that one can justifiably be specially concerned about the experiences at a given future time of more than one being (i.e. not all of whom can be oneself). However, as suggested above, to survive in the sense at issue to have an experience is to enjoy a relation to this future experience. It is this relation (rather than identity per se) that matters in identity. Though this relation may normally coincide with identity, this relation may hold even when there is not identity. If what is important is such a relation to later experiences, then, perhaps, this relation may be duplicated.

I am not suggesting that the idea that survival and identity may not coincide is even coherent. I consider this possibility only because it has been entertained in the literature on personal identity. (Parfit, 1984) (See Chapter 3 of this thesis.) Most of us find this supposed possibility to be deeply disturbing. It seems plausible that, as Williams suggests, the only way that one can feel specially concerned about future experiences is to suppose that one (the same person as one's present self) will exist and have conscious states at the future time. (Williams, 1975) That is, the only way to make sense of egoistic concern for a future experience is to think of oneself as having the future experience. However, I think it is a mistake to presume that survival and
identity always coincide. Quite possibly, investigation of survival may lead to the conclusion that cases in which one survives as more than one person (and other problem cases) are impossible. However, to make this assumption at the outset is to beg an important question. Unless and until it can be established that such is the case, it seems preferable to concentrate on the more fundamental of the two issues, on survival rather than on identity. However, though this conclusion is beyond the scope of this thesis, I believe that a strong case can be made that survival is always unique and hence, if identity is understood as unique survival, survival and identity always coincide.

5. Survival as a Sentient Being

There are, I think, a number of disadvantages to equating one's survival with one's survival as a person. Indeed, because being a person is widely considered to require the possession of a variety of higher-level cognitive functions, there is a serious danger of confusion in the inattentive use of the term "person" in discussion of survival. If one has a degenerative brain disease in the later stages of which only the most rudimentary of experiences will be associated with one's brain, then one may still feel this egoistic concern for these experiences. Certainly, one who is diagnosed with a condition that results in the loss of all these higher functions long before a final cessation of consciousness will fear the progress of the disease, perhaps more than immediate death, exactly because one does think that one will survive without
these functions. However, even if one survives the loss of these higher-level functions, one may not survive as a person.\textsuperscript{13} Also, one might think that to survive as the person one is, one must retain not just these higher-level functions but also specific individual traits. One may also fear the loss of these traits (perhaps ones of which one is especially proud) more than immediate death, again just because one thinks that one will survive their loss.\textsuperscript{14} Thus, one may not survive as a person to have all of the experiences about which one should be specially concerned. Hence unclarified use of the term "person" may obscure one's understanding of what's really at issue, i.e. survival to have experiences about which one should be specially concerned.

To avoid these problems, I focus on one's survival as a sentient being, a being capable of having experiences but possibly lacking (some of) the remaining attributes of a person.\textsuperscript{15} Though no longer a person after suffering extensive brain damage, one may still survive as a sentient being. Because sentient beings that

\textsuperscript{13} I am not suggesting that human beings with such greatly reduced capabilities and/or impoverished experiences should not be considered as persons in the sense of being extended the usual rights of a person. I avoid the issue of survival as a person only because the term "person" may suggest to some a certain sophistication of experiences and/or capabilities.

\textsuperscript{14} I am not presupposing that one may survive the loss of one's psychological characteristics and capabilities. However, I consider it a mistake to accept at the start that we are entities that cannot survive the loss of elements of their psychology other than the capacity to have bare conscious states.

\textsuperscript{15} Introduced in connection with survival by B. Thurston, the term "sentient being" has been employed extensively by R. Sikora in his work on personal identity.
are not persons may survive as sentient beings (and because of the prevalent anthropocentric understanding of "person"), it might be best to avoid the term "person" altogether and use the term "sentient being" exclusively. However, because the use of the former term may assist us, as persons, to employ the principal tools we have available, namely our intuitions regarding our own survival, I will sometimes speak of the survival of a person. However, this expression will always mean the survival of a person as a sentient being rather than as a person.

Focusing on survival as a sentient being rather than as a person has another important advantage. Recall that Sidgwick held that one should feel the same measure of special concern regarding the contents of all of one's future experiences. That is, he accepted what Parfit calls the Requirement of Equal Concern: "A rational person should be equally concerned about all parts of his future." (Parfit, 1984, p.313) It would thus be irrational to avoid torture now at the expense of greater suffering later. Clearly, how one views future experiences about which one feels specially concerned depends on what one anticipates these experiences may be like. However, I think that most of us would be inclined, on reflection, to agree with Sidgwick that whether or not egoistic concern is justified is, in any given case, a determinate all-or-nothing matter. If such justification is all-or-nothing, then whether or not one survives is also an all-or-nothing matter.

This view of survival as all-or-nothing is certainly not universally accepted in the contemporary literature. Though Parfit
accepts that "a rational person's dominant concern should be about his own future...," like other contemporary philosophers, he denies that whether or not one should care specially about future experiences is all-or-nothing. He holds that reasonably one may be "less concerned about those parts of his future to which he is now less closely connected [i.e. after which he has undergone significant psychological changes]." (Parfit, 1984, p.317) Though the claim that survival is all-or-nothing is certainly not universally accepted and is not essential to my thesis, I wish to avoid any assumptions about ourselves and our survival that conflict with this claim. Clearly, survival as a sentient being can also be an all-or-nothing matter. However, because cognitive sophistication is a matter of degree, so must be survival as a person.

Thus, there are some clear advantages to focusing on survival as a sentient being rather than as a person. Because of the intimate connection between the notion of survival in the sense in question and the notion of a sentient being (and to avoid confusion with other senses of survival), I refer henceforth to survival in this sense as survival as a sentient being.
1. Epistemological and Metaphysical Problems of Survival

As often stated, the problem of identity is to find "the criterion of identity over time." As Parfit notes, this phrase is ambiguous. It could mean "our way of telling whether some present object is identical with some past object," or it could mean "what this identity necessarily involves, or consists in." (Parfit, 1984, pp.202-203) In particular, there are two problems of personal identity: the epistemological problem of how to determine whether or not persons at different times are the same and the metaphysical problem of determining in what a person's identity over time consists. Likewise, the epistemological problem of survival is to determine how, in any given case, we can know whether or not one does survive. The metaphysical problem of survival is to determine in what survival consists.

To solve the epistemological problem of survival, we would need to be able to answer questions of the following general form. Suppose that at an earlier time there is a sentient being X and that at a later time there is a sentient being Y and that a sequence of events S occurs between these times. Does X or does X not survive as Y? (Presumably, at a given time, we can indicate a sentient being by ostension to a particular body that supports consciousness.) What evidence can there be that X does or does not survive as Y? Ideally, we would like a set of necessary and
sufficient conditions for survival, ones such that it is always clearly ascertainable whether or not they hold. Such a set would provide a complete answer to the epistemological problem of survival.

What complicates the metaphysical problem of survival is that, in cases that we may consider to be cases of survival, a number of factors that may be involved in survival tend to be conjoined. In such cases, there is not only the continued existence of one's body but also considerable psychological and physical (generally qualitative) stability. Even in pathological cases, as one ages one retains much of one's physical and psychological make-up. Normally, when one does change significantly, one does so only gradually. However, it does not follow that survival consists in bodily survival, qualitative stability, or gradualness of changes. As Parfit writes, "...what is necessarily involved in a person's continued existence is less than what is in fact involved." (Parfit, 1984, p.210) Though all of these factors may accompany survival, they may not be that in which survival consists. The metaphysical problem of survival is to determine what is necessarily involved in one's survival.

In tackling the metaphysical problem, it is important to treat the epistemological and metaphysical problems as separate. A solution to the metaphysical problem does not necessarily provide a solution to the epistemological problem, simply because it may not always be possible to determine whether or not that in which survival consists is present. For example, the conclusion that
survival consists in the continued presence of an immaterial soul would not help to determine whether or not X survives as Y because it would be impossible to observe whether or not X and Y have the same soul. Like Swinburne, I would reject verificationist arguments that the question of whether or not X survives as Y is meaningless (that there is no metaphysical fact about whether or not one survives), unless there is evidence that enables us to decide one way or the other. (Swinburne, 1973-74, p.243) There may be an objective metaphysical fact about whether or not one survives even if we cannot be certain one way or the other.

However, in attempting to answer the metaphysical question, we may also be hoping to make some progress with the epistemological question or may simply forget that this latter question is not at issue, and herein lies a danger. A metaphysical theory which is more helpful than another at tackling the epistemological question may be afforded greater plausibility than it really warrants. For example, a metaphysical theory according to which X's survival as Y consists in X and Y having the same body would make it very easy to determine whether or not one survives. Certainly, the usefulness of such a theory in answering questions of survival should not count against it, but equally this usefulness should not count in its favour. The metaphysical problem should be considered in isolation from the epistemological problem. This thesis is concerned with the metaphysical problem only.

Closely related to the metaphysical problem of survival is the metaphysical question of what we are. It is important to realize
that what we view ourselves as depends on that in which we are interested. If our interests are anatomical, then we may view ourselves as internally organized physical objects. If our interests are physiological, then we may view ourselves as organisms. If our interests are in the specific social and/or economic role that one may fill, then we may view a person as whatever may fill such a role. Also, what we are should largely determine in what our survival consists. For example, if we are bodies, then our survival should consist in bodily continuity. A focus of this thesis is how we should view ourselves given that we are beings who should be specially concerned about certain future experiences.

2. Reductionism

As we shall see, metaphysical theories of in what survival consists fall into two broad categories: reductionist theories and nonreductionist theories. Because the term "reductionist" is employed in many areas of philosophy, it is important to clarify what this term may mean.

As John Searle writes: "...the basic intuition that underlies the concept of reductionism seems to be the idea that certain things might be shown to be nothing but certain other sorts of things." (Searle, 1994, p.112) However, as he notes, people mean different things by the "nothing-but relation." Indeed, one may be a reductionist with regard to very different sorts of things: objects, properties, relations, events, persons, theories etc.. He
identifies five different senses of reduction.

1. In ontological reduction, "objects of certain types are shown to consist of nothing but objects of other types." (Searle, 1994, p.113) For example, physical objects are nothing but rigid assemblages of molecules.

2. In property ontological reduction, properties of certain types (of certain sorts of entities) are shown to be nothing but properties of other types. For example, heat is nothing but molecular motion. I take this sense of reduction to include relational ontological reduction, in which relations of certain types (between entities of certain sorts) are treated as relations of other types. In this sense of reduction, facts about entities (i.e. that they enjoy certain properties and/or relations to other entities) are reduced to other facts.

3. In logical or definitional reduction, "...words and sentences referring to one type of entity can be translated without residue into those referring to another type of entity." (Searle, 1994, p.114) For example, sentences about numbers are translated into sentences about sets.

4. Theoretical reduction involves a relation between theories: "laws of the reduced theory can (more or less) be deduced from the laws of the reducing theory." (Searle, 1994, p.114) The phenomena treated by the reduced theory are explained in terms of the reducing theory. For example, gas laws are reduced to laws of statistical thermodynamics.

5. In causal reduction, "the existence and a forteriori the causal
powers of the reduced entity [or phenomenon] are shown to be entirely explainable in terms of the causal powers of the reducing phenomena." (Searle, 1994, p.114) For example, solidity (i.e. the power of objects to resist penetration) can be thus explained in terms of (and thus causally reduced to) causal powers of molecules to interact with and thus align themselves with other molecules into rigid lattice structures.

Searle claims that his view of the relationship between the mind and the brain, that conscious states are caused by neurobiological processes, also qualifies as a causal reduction. (Searle, 1994, p.115) In making this claim, Searle's conflates two quite distinct notions: that one thing is caused by another, and that the causal powers of one thing are explained in terms of the causal powers of another. To my mind, only the second counts as reduction. That one thing results from another thing does not imply that the former is causally (or in any other way) reducible to the latter. For example, it does not follow from the fact that the assassination of Archduke Franz Ferdinand caused World War I that the Great War can in any way be reduced to this assassination. When Searle claims that consciousness can be "explained" in terms of causal interactions in the brain, he means merely that consciousness results from this interaction. (Searle, 1994, p.112) To truly explain causal power A in terms of causal power B is to account for (i.e. make understandable) the former in terms of the latter. If we are able to do so, then we have no need to postulate causal power A. That is, such an explanation of causal power A in
terms of causal power B enables us to eliminate A in favour of B (i.e. to effect a causal reduction of A to B).

If A's are nothing but B's and can be thus reduced to B's in the sense of ontological (including property/relational ontological) reduction, then there is no need to include A's along with B's in our basic ontology. That is, we can describe reality without using the terms referring to A's. (Of course, we could continue to use the term "A" as the more convenient and perhaps more compact means of expressing ourselves. For instance, despite our knowledge of molecules, "object" still seems preferable to "rigid assemblage of molecules.") Does a reduction in the other senses of A's to B's likewise justify our dispensing with the term "A"?

It seems clear that logical/definitional reductions may lead to ontological reductions. For example, if we accept that numbers can be defined as sets, then we view numbers as nothing but sets (and hence not as something that exists over and above these sets). Theoretical reductions also seem to enable us to carry out ontological reductions. For example, if the thermal phenomena of gases can be explained in terms of the laws of statistical mechanics, then heat (of a gas) can be reduced to molecular motion. There is no need to think of heat as something that exists over and above this motion. Likewise, causal reductions (as just clarified) may enable us to carry out ontological reductions and hence obviate the need to include the terms at issue in our vocabulary and their referents in our basic ontology. If we can explain how objects are
able to resist mutual intrusion in terms of the rigidity of their hypothesized lattice structure, we could perhaps dispense (any linguistic awkwardness aside) with the term "solidity" in favour of "rigidity of molecular lattice structure". If all the causal powers of one thing are reducible to those of another thing, we could perhaps dispense with the former thing in favour of the latter.

However, to establish that B causes A does not, in itself, licence ontological reduction of A to B and thus obviate any ontological commitment to A. For instance, it does not follow from the fact that experiences are causally dependent on brain states/processes that they are "nothing but" brain states/processes in the ontological sense. To show that experiences are nothing but physical states, we would, I think, need to explain how experiences have their qualitative features on the basis of the hypothesis that experiences are such states.

Though theoretical reduction may lead to ontological reduction, Searle's account of how this may happen in science is simplistic and disunified. Consider the case of the reduction of heat to molecular motion, a paradigm case of reduction in science. Perhaps, if the thermal phenomena of gases can really be understood in terms of the laws of statistical mechanics, then heat of a gas can be reduced to molecular motion. However, this property reduction of heat to molecular motion is much more complicated than he seems to suppose. It is not simply a case of a theoretical reduction leading to a property ontological reduction. Scientific laws (such as gas laws and the laws of statistical mechanics) tend
to be causal laws that express causal connections between properties of supposed entities. For example, according to a well-known gas law, increasing the temperature (measured on an absolute scale) of a gas enclosed in a fixed volume causes the pressure to increase proportionately. The theoretical reduction here involves the reduction of causal powers of the entities hypothesized by the reduced theory (i.e. the powers of confined gases to increase their pressure when heated) to the causal powers of the entities hypothesized by the reducing theory (i.e. the powers of molecules to transfer their momentum to other molecules). Not only does this theoretical reduction involve causal reduction but it also involves (or perhaps leads to) the ontological reduction of entities as well as properties. Not only is heat reduced to molecular motion but samples of gases are reduced to a collections of weakly interacting molecules. It seems that in science, ontological reduction, ontological property reduction, theoretical reduction, and causal reduction may be inseparably entwined.

The sense of reduction that is at issue in this thesis is ontological (including property/relational ontological) reduction. Reductions in the other senses are important only to the extent that these reductions may justify ontological reductions.

3. Reductionism and Survival

How does reductionism apply to the issue of survival of sentient beings? What do reductionist theories of survival claim to be reducible to what? What sorts of reduction are involved? I first
Consider reductionism with regard to personal identity. According to Thomas Reid, "Identity...[is] a relation between a thing that is known to exist at one time, and a thing which is known to have existed at another time." (Reid, 1785, p.108) For instance, to say that a ship X identified as existing at one time is the same as a ship Y identified as existing at another time is to assert that there is an identity relation between X and Y. How does this apply to the identity of persons?

According to Locke, personal identity is a matter of whether a person identified as existing at one time is the same as a person identified as existing at another time (of whether there holds an identity relation between these persons), for instance whether the person now with the body of the cobbler and the person then with the body of the prince are the same person. (Locke, 1975, xxvii, 15) He was concerned with the question of what personal identity necessarily involves. Though, as he realized, to determine in what personal identity consists we need to appreciate the nature of a person, he clearly distinguished the issue of in what identity of a person over time consists from the issue of the nature of a person. He wrote: "...to find wherein personal Identity consists, we must [first] consider what Person stands for; which, I think, is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing, in different times and places..." (Locke, 1975, II, xxvii, 9) In the same section, he clearly distinguishes this issue from the issue of what personal identity (properly understood as "the sameness of a
Rational being") over time consists in. Personal identity is a relation between person stages. The issue that has concerned personal identity theorists since the time of Locke is the nature of this relation.¹

As he is widely understood, Locke held that the identity of X at an earlier time and Y at a later time consists in Y remembering the experiences of X. (See Chapter 3.) His best-known early critics, Butler and Reid, denied that personal identity can be analyzed in this manner and apparently in any illuminating way at all. Reid held identity (and personal identity in particular) to be "too simple a notion to admit of logical definition." (Reid, 1785, p.108) Butler held that it is primitive and unanalyzable, that to attempt to define "wherein personal identity consists...would but perplex it." (Butler 1756, p.99) (Nonetheless, neither denied that

¹ Depending on one's metaphysics, one may view persons as persisting over time (existing in their entirety at a range of times) or as extending through time (or, if one is skeptically inclined, perhaps as neither). (See Lewis, 1986, p.202, for a discussion of this distinction. Note that Lewis employs somewhat different terminology.) I remain neutral on this issue. For persons considered as persisting, the issue seems to be in what a person's identity over time consists. For persons considered as temporally-extended, the issue seems to be in what the temporal unity relation (between distinct temporal parts) of a temporally-extended person consists. When I use "person stage," I do not mean to imply that such a stage is a temporal part of a temporally-extended person. (Note that both John Perry (Perry, 1975a, pp.7-20) and Sydney Shoemaker (Shoemaker and Swinburne, 1984, pp.72-91) employ this term without commitment to temporally-extended persons.) Likewise, when I use "person at (or over a small period of) time," I do not mean to imply that persons are merely persisting rather than temporally-extended. I employ interchangeably not just these terms in particular but also the overall vocabularies of these metaphysical views. Analogous remarks apply to the case of sentient beings.
personal identity over time might consist in the continued existence of a nonphysical self.) Some contemporary philosophers (such as Parfit and Perry) defend views along the lines of Locke's, and some others (such as Unger and Williams) argue that (or at least roughly that) the identity of persons consists in their having the same brain and/or body. Like Locke, these philosophers hold reductionist views, according to which identity can be reduced to more basic relations (i.e. identity consists in these relations). A small number of dissenters (most notably Swinburne) deny that personal identity consists in such relations. Like Butler and Reid, they take a nonreductionist view of personal identity.

As far as I know, Parfit is the only writer on personal identity who gives an explicit definition of a reductionist theory. According to Parfit, reductionist views of personal identity claim

1. that the fact of a person's identity over time just consists in the holding of certain more particular facts.

Such views "may also claim"

2. that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even claiming that this person exists. These facts can be described in an impersonal way. (Parfit, 1987, p.210)

On the reductionist view, "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." (Parfit, 1987, p.211)

Clearly central to Parfit's treatment of reductionism, Claim (1) shows that he too understands reductionism as the view that the relation of personal identity over time consists in more basic
relations. Though (2) may seem to be a separate and perhaps inessential claim, given his clear statement that a view is nonreductionist if it rejects "either or both of the two Reductionist claims" and his overall treatment of reduction, I am inclined to read (2) as a statement of what he (or reductionism as he thinks we should understand it) would not accept as these more particular facts.² (Parfit, 1987, p.210) Claim (2) serves to rule out supposed reductions of personal identity to facts that reasonably should not count as more particular. The need is obvious for some such restriction of these more particular facts. For example, it seems illegitimate to treat a theory of personal identity as reductionist if it avoids talking about the identity of persons only by talking about the identity of sentient beings.

If we take Claim (2) as a part of Parfit’s definition of reductionism, then we are faced with a number of ambiguities. First, (2) seems to conflate the issue of whether we can reduce persons to sub-personal elements (and hence whether we need not include persons in our ontology), and the issue of whether we can reduce the identity of persons over time. Is it the existence of a person (at a time) or the continued existence of a person over time that reductionists claim to be reducible to "the existence of a brain and body" and to such "interrelated physical and mental

² In the 1984 edition of Reasons and Persons, Parfit defines a reductionist theory as one that accepts both (1) and (2). The change in the later edition does not serve any obvious purpose and is, I think, potentially misleading.
events"? However, if (as Claim (1) shows) reduction of identity over time is at issue, then to say "this person exists" can only mean that this person continues to exist over time rather than that this person exists at a given time. Second, it is not obvious from (2) whether a theory that treats persons as a special sort of a more general species of entities G and that presupposes the identity of members of G would count as a reductionist theory. Third, it is not obvious whether or not he would accept a theory as reductionist that may claim that experiences belong to a mental transtemporal unity (such as a stream of consciousness). It is certainly not obvious from (2) just what would count as these "more particular facts."

It is clear that Parfit thinks that it is the reduction of personal identity to physical continuity (i.e. the continued existence and/or physical continuity of material entities) and/or psychological continuity that is at issue. His "more particular facts" seem to be facts regarding these continuities. He writes that, on a non-reductionist view, "...personal identity over time does not just consist in physical and/or psychological continuity..." (p.210) Ultimately, then, reductionist theories claim that personal identity does consist in such continuities. Though to thoroughly disentangle the issues in (2) would be no easy matter or a task that would be worthwhile to undertake here, I

---

3 Given his treatment of psychological continuity, we are, I think, to read "psychological continuity" as involving retention or continuity of memory, character traits, intentions etc., what I call "qualitative psychological continuity."
think that careful reading of (2), and consideration of what it
excludes, also supports this assessment of Parfit's understanding
of reductionism.

Interestingly, this characterization of reductionist theories
as involving physical and/or psychological continuity closely
resembles Swinburne's definition of an empiricist theory of
personal identity as "a theory which analyses personal identity as
a matter of bodily continuity and continuity of memory and
character." (Swinburne, 1973-74, p.233) Despite this use of
"analyzes," it is clear that Swinburne understands empiricist
theories to claim not just that our concept of personal identity
can be analyzed in terms of these continuities but that personal
identity consists in these continuities. The problem with which he
is primarily concerned is to determine "Wherein does the identity
of persons consist?" (Swinburne, 1973-74, p.240) He doesn't just
deny that our concept of identity can be analyzed in terms of these
continuities but that personal identity is "constituted" by these
continuities. (Swinburne, 1973-74, pp.240,241) Though the term
"reductionist" may not be widely employed in the literature on
personal identity, this term corresponds to an important
distinction treated in this literature. Indeed, the issue in much
of this literature is whether or not personal identity can be
reduced to such continuities and hence whether or not theories of
this type are satisfactory.

I think that it is reasonable to suppose that both Parfit and
Swinburne would consider qualitative physical continuity (i.e.
retention or continuity of physical characteristics) to be an acceptable element of a reductionist (or empiricist) theory. Indeed, Parfit accepts that physical similarity may have some importance with regard to one's identity. (Parfit, 1984, p.217) I will understand a reductionist theory to claim that one's identity over time consists in relations involving physical continuity (physical relations), and/or in terms of relations involving qualitative psychological continuity and/or qualitative physical continuity (qualitative relations). In a reduction of personal identity (a case of ontological reduction of a relation), facts regarding the holding of this identity relation between person stages are reduced to facts involving qualitative and/or physical continuity. If reductionism is true, then we can assert that X (at one time) is the same person as Y (at another time) without presupposing the identity of X and Y, and without using the language of personal identity in general. Indeed, we could "completely describe reality" without referring to the identity of persons over time. (Parfit, 1984, p.213)

The question of whether a person at a time can be reduced to sub-personal elements (i.e. whether we can account for the facts of our existence at a time in terms of these elements), and hence whether we can dispense with the notion of a person at a time, is an entirely separate issue. At a time, there are material entities, physical events, experiences, and persons. Perhaps persons at a time are reducible to physical entities, physical events, and experiences thus obviating the need to include persons at times in
our ontology, and maybe even the experiences are reducible to the physical events. Though the question of what ontology is needed to account for the facts of our existence at a time is important, in considering the identity of persons at different times this question is not at issue. An answer to the question of whether personal identity is reducible may clearly leave open the question of whether the notion of a person at a time is dispensable, and vice versa. We may be able reduce personal identity without reducing persons, and vice versa.

Adaptation of this understanding of reductionism to the consideration of the survival of sentient beings is unproblematic. Because (as Parfit recognizes) more than one person at a later time may enjoy the required physical and/or qualitative relations to a single person at a earlier time and there may thus be an element of convention involved in questions of personal identity, the definition that I have adapted from Parfit is better suited to the question of survival. (See Chapter 3 of this thesis.) Also as explained in Chapter 1, there are advantages to focusing on survival as a sentient being rather than as a person.4

---

4 There is additional advantage not mentioned therein. If reductionism is false, it may still make sense to treat the survival of a person as the survival of a person as an entity of a more general species G of sentient entities and its continuing to have the physical and mental qualifications to be a person. However, because a sentient being is the most general sort of entity capable of having mental states, we need not consider the possibility that the survival of a sentient being qua sentient being may be the survival of a sentient being qua member of a more general such species G and this being’s continuing to be sentient.
I will understand a reductionist theory of survival (as a sentient being) to claim that the survival of X (at earlier time) as Y (at a later time) consists in physical and/or qualitative relations between X and Y. There are thus two types of reductionist theories: physicalist theories (according to which survival consists in physical relations) and qualitative theories (according to which survival consists in qualitative relations). In a reduction of survival (a case of ontological reduction of a relation), facts regarding the holding of this survival relation between sentient being stages are reduced to more particular facts regarding physical and/or qualitative continuity. If reductionism is true, then we can assert that X (at an earlier time) survives as Y (at a later time) without using the language of survival (including such words as "survive," "survival," "survivor" in the sense at issue). We could completely describe reality without referring to the survival of sentient beings. Nonetheless, even if reductionism is true, it may be convenient to continue to speak of persisting or temporally-extended persons and sentient beings, and their survival. Note, once again, that the question of whether we can dispense with the notion of a sentient being at a time is not at issue. We may be able to reduce survival in this way without reducing sentient beings, and vice versa.

By a nonreductionist theory, I mean simply a theory of survival that is not reductionist. According to nonreductionism, survival does not consist in qualitative and/or physical relations. There are theories thus characterized as nonreductionist which may
seem to some to reduce survival to other facts. For instance, the stream of consciousness theory (SCT), the view that X’s survival as Y consists in the experiences of X and Y being encompassed by a stream of consciousness, might be seen as reductionist. However, whether or not a theory is reductionist depends on what we are prepared to count as "more particular facts." I have already explained the need for some restriction of these facts. As I understand these facts (i.e. as involving physical and/or qualitative continuity), the belonging of experiences to the same stream of consciousness does not count as such a fact, and hence the SCT does not qualify as reductionist. There are clear and important differences between the commitments of reductionist and nonreductionist theories (as defined) that make it reasonable to understand the more particular facts as I do.

Consider the following seemingly attractive picture of how the world at an earlier time relates to the world at a later time. At these times, there exist material entities, sentient beings, and there occur physical events and mental events. There are also physical and qualitative continuities between material entities and between sentient beings at these times, and qualitative continuities between mental events at these times. We may also suppose that there are natural laws relating what there is and what there occurs at these times and generally constraining how the world at the earlier time evolves into the world at the later time.

---

5 I am indebted to Alan Richardson for pointing this out.
Now is this an accurate and complete picture of how the world at the earlier time and the world at the later time relate? Does whether or not we survive involve more than these facts regarding the occurrence of events, the existence of things, and physical and qualitative continuities? Is there a matter regarding our survival over time not settled in terms of such facts? The difference between reductionism and nonreductionism is between rejecting and accepting the need for further facts to account for our survival over time. If reductionism is true, then there is no need to commit ourselves to such further facts. If nonreductionism is true, that is if there really is such a thing as survival and survival as a relation cannot be reduced as above, then we have to commit ourselves to survival (as a basic irreducible fact or relation) or include surviving entities or mental transtemporal entities (as encompassing mental events at these times) in our ontology. Given this significant difference between reductionist and nonreductionist theories of survival (as defined) and the focus of the relevant literature, it is perfectly reasonable to apply these terms in discussion of survival as I do.

As was the case with the reduction of heat to molecular motion, reduction of survival to these continuities may be licenced by a sort of theoretical reduction. Perhaps, reduction of our commonsense understanding of survival (including a commitment to

---

\(^6\) Again, we can remain neutral about whether some of these facts (such as the existence of a sentient being at a time) can be reduced to other such facts.
survival) to a theoretical framework involving physical and/or qualitative continuities may lead to ontological reduction of survival to these continuities. However, the case of reduction of survival and the case of reduction in science (such as the reduction of heat to molecular motion) are very different. In the case of survival, we may have reduction of the relation without reduction of the related entities, and there may be no causal reductions involved. There are also great differences in how these reduction claims might be legitimized. In the case of scientific theories, if the phenomena treated by the reduced theory can be understood in terms of the reducing theory, and if the reducing theory can accommodate further observations better than the reduced theory, then we have reason to accept this theoretical reduction and to revise our ontological commitments accordingly. Clearly, our reason for accepting reduction of survival would be very different. Rather than making any phenomena understandable, the reducing theory would need to explain in terms of qualitative and/or physical continuities how egoistic concern can be justified (i.e. why we should care specially about our future experiences). Also, whether we are prepared to accept the reducing theory may depend, in part, I think, on our deep beliefs about ourselves and our survival rather than on scientific observations.

In this thesis, I argue that survival as a sentient being does not consist in physical and/or qualitative relations. I argue that we cannot explain in terms of physical and/or qualitative relations how egoistic concern can be justified, and hence that no
reductionist theory of survival as a sentient being can be satisfactory. I readily concede that being produced by my same brain may be causally responsible for an experience being one that I survive to have. I have no immediate objection to supposed reductionist theories that offer analyses of species of survival other than the one at issue. However, I do object to reductionist theories that equivocate between different senses of survival and convey the impression that they are somehow analyzing survival in the sense at issue without accepting the need to explain how egoistic concern can be justified.

The main body of this thesis breaks naturally into two parts. In Part I (Chapters 3, 4), I discuss the qualitative view and explain why many now find this view to be unsatisfactory. In Part II (Chapters 5, 6, 7, 8, 9), what I consider the main focus of this thesis, I treat some initially plausible physicalist theories and expose some general difficulties for physicalism. In the concluding chapter (Chapter 10), I briefly explore the nonreductionist alternatives to reductionism. However, a serious treatment of nonreductionism is well beyond the scope of this thesis.
4. The Qualitative Approach to Survival

In Part I of this thesis, I treat the qualitative view. A qualitative theory analyzes X's survival as Y in terms of qualitative relations involving the psychological make-up and/or the physical state of X and Y (and/or perhaps of intermediate sentient being stages). According to such a theory, certain stages are appropriate psychologically and/or physically to follow a given stage, whereas others are not.

I focus on the major sort of qualitative theories: psychological reductionist theories (content-based theories). A pure psychological reductionist theory (a pure content-based theory) analyzes X's survival as Y in terms of psychological relations (content-dependent relations) involving the psychological make-up of X and Y (and/or perhaps of intermediate sentient being stages). That is, to say that X survives as Y is just to say that Y's psychological state is appropriate given X's psychological state, or perhaps that X's and Y's psychological states are linked by a chain of appropriate psychological states. Equivalently, to say that one who has an earlier experience E survives to have a later experience F is just to say that E and F are related in terms of the memories, desires, pleasures, pains, beliefs, attitudes, fears etc. that these experiential states do or could contain. On such a theory, certain experiences are of appropriate content or potential content to follow a given experience, whereas others are not. To survive is for there to be suitable sequels to one's present experiences. A content-based theory may add a causal
requirement, a requirement that the experiences and psychology of Y be appropriately caused given what realizes the experiences and psychology of X, that the psychological state of Y result somehow from the psychological state of X. Historically, psychological reductionist theories, such as variants of the memory theory, often attributed to Locke, have been the most widely held theories of survival.

The psychological reductionist approach to survival has obvious appeal. What seems to be important is our experiences (that we will have them and what they will be like). It seems natural to suppose that what matters to us is our future experiences rather than the fate per se of our body. For instance, suppose that I am diagnosed as having a degenerative brain condition that will result within one month in the death of my brain and permanent cessation of consciousness, but that my body will be maintained on life support for the remainder of the year. Clearly, I should view this prospect just as if I were to die in one month, and I shouldn't be any more concerned about the fate of my body for the remaining eleven months than I would be if my body were to die at the same time as my brain. Certainly, my present body is the same body as my future brain-dead body, but I am not the same sentient being as any future sentient being with that body, and indeed there is no such sentient being. My body interests me only inasmuch as it supports my experiences. This is not to say that whilst I am having experiences, I am not concerned with the fate of my body. It is natural to suppose that one will continue to have experiences only
so long as one's body (and more importantly one's brain) retains the capacity to support experiences. Moreover, the nature of the experiences that one will have will depend on the fate of one's body. Hence our concern for our bodies is derivative in the sense that bodily continuity seems necessary to provide that about which we are really concerned - the continuation of our experiences. In hoping to survive, one is hoping that there will be future experiences that will be related to one's present experiences in certain ways.

Because the psychological reductionist approach appeals to a mentalistic relation in terms of the mere contents of experiences involved, this approach has the major advantage of being compatible with both dualist and materialist views of the mind. This approach enables a dualist (who holds that experiences are not mere brain states) to explain how one who has an earlier experience can survive to have a later experience without appeal to physical considerations of persistence of the brain or body over time. On the other hand, a materialist can identify the relevant psychological contents with states of the brain. He can claim that if we had sufficient knowledge of the neural basis of psychological states, we could phrase a content-based theory in terms of physical descriptions of brain states. However, even after we may acquire this important knowledge, the present talk of psychological states may remain the more convenient and compact means of describing one's survival.

My approach to the qualitative view is to consider this view
(with a focus on the content-based view) in general and then to expose some general problems for this view. In Chapter 3, I discuss briefly the history of the content-based view and the state of the debate regarding this view. In Chapter 4, I present a general argument, inspired by Butler, that no reductionist theory can explain how egoistic concern can be justified and hence that no such theory can be satisfactory, and I examine how this argument applies to the qualitative view. (Butler, 1756) I also consider an attempt due to John Perry to explain in terms of qualitative similarity how one's concern may be justified for experiences to be associated with one's body. (Perry, 1976) I also argue that no qualitative theory can possibly explain how egoistic concern can be justified unless its causal requirement is so strong (requiring basically the normal causal basis of the ensuing experiences) that the result should be considered a physicalist theory.

5. The Physicalist Approach to Survival

In Part II of this thesis, I treat what now seems to be the dominant response to the question of what we are and the related question of in what our survival consists - physicalism. There are two basic components of this physicalist approach: a view of ourselves as material entities (bodies or brains or organisms etc.) and a related physicalist theory of survival according to which survival consists in the continued existence and function of the sort of material entity in question.

Presumably, how exactly such a theory understands what it
means to say that Y is the same material entity as X could be explicated (like the identity of any temporally-extended physical object) in terms of spatiotemporal continuity (and continued activity) of the entity as a whole and possibly of its component parts. Because there may be some physical continuity between X and Y without there being enough for these to be the same material entity in any ordinary sense, it may be the case that survival consists in this physical continuity rather than the identity of the entity over time. However, because a characterization in terms of physical continuity would be highly complex and there may be no basis to decide just how much physical continuity is needed for survival, it is convenient to talk of X and Y being the same entity.

A physicalist theory may also have a psychological requirement. The theory may require Y to have certain psychological features in addition to having mere conscious states (as would Y given that it is assumed to be a sentient being). For instance, the theory may require as well that Y’s body support a less basic or perhaps a distinctive psychology. The theory may also require that Y’s psychological state be related in an appropriate way to X’s psychological state. It may also require the continuous support of (or continuous capability to support) a certain level of psychology. One notes that there may be theories that could be characterized equally well as physicalist theories with a psychological requirement or as content-based theories with a causal requirement.
A physicalist theory presents the claim that one survives in just the cases specified (in terms of physical continuity and any psychological requirement) as an analysis of in what survival consists. That is, to say that X survives as Y is just to say Y is the numerically same material entity as X (and any psychological requirements are satisfied). Thus, a physicalist theory may equate the survival of one who has an earlier experience E to have a later experience F with the association of E and F with the same material entity.

There are two types of physicalist theories: physical and biological reductionist theories. According to a physical reductionist theory, survival claims are fully expressible in the language of physical (bodily) continuity (perhaps with a psychological requirement). According to a biological reductionist theory, survival claims are fully expressible in terms of the continued existence of organisms (again perhaps with a psychological requirement). These types are, I think, exhaustive of physicalist theories.

My approach to physicalism is to assess some of the seemingly more promising theories and then to expose some general problems for physicalism. Because Peter Unger and Peter Van Inwagen offer well-developed and initially attractive theories, I focus on these theories. (Unger, 1990; Van Inwagen, 1990)

In Chapter 5, I discuss Van Inwagen's view of ourselves as organisms and argue that this view is implausible. (Van Inwagen, 1990) In Chapter 6, I discuss Unger's physical approach to survival
as well as his objective view of persons as enormously complex physical entities (bodies and most importantly brains). (Unger, 1990) I argue that Unger's account is problematic in itself and suffers from serious methodological problems. In Chapter 7, I argue that transfer of consciousness (i.e. survival to have experiences associated with a body numerically distinct from the body with which one's present experience is associated) and disembodied survival are conceivable. Since physicalist theories imply that transfer and disembodied survival are inconceivable as well as physically impossible, if these scenarios are indeed conceivable, then no physicalist theory can analyze survival as a sentient being. In Chapter 8, I argue that recent developments in the theory of meaning and reference do not provide sufficient reason to doubt that transfer and disembodied survival are conceivable. In Chapter 9, I argue that no physicalist theory can explain how egoistic concern can be justified and hence that no such theory can be satisfactory.

6. Methodology: Deep Beliefs and Thought Experiments

Because what methodology one employs often fully determines the outcome of one's philosophical investigations, it is appropriate to end this introduction with a statement of the methodology that I accept. I believe that thought experiments are an appropriate and essential component of a reasonable methodology for investigation of personal identity and survival. Because questions of survival often seem hopelessly confusing, such a
methodology is indispensable. In some cases, we may have clear and consistent reactions about whether or not one survives. However, about other cases, we may tend to be uncertain or to react very differently. Thought experiments may serve to reveal and to clarify our responses to these questions. They may serve to bring into focus our reactions to cases and to tease out our beliefs about survival. By abstracting away the details of more concrete cases, thought experiments bring our intuitions into focus. They also assist us to disentangle factors (such as physical and psychological continuities) that tend to be conjoined in real life cases. They help not only to sharpen our intuitions but also to explore the consequences of our views. For such reasons, thought experiments have been a primary and accepted element of the methodology of personal identity ever since Locke introduced his famous imaginary case of the prince and the cobbler. (Locke, 1975, II, xxvii, 15) Many reductionists (such as Unger and Parfit) whose views I reject explicitly accept a methodology involving thought experiments and/or employ them freely.

A thought experiment employed in the investigation of survival (at least in my sense) invites us to assess whether or not the subject should feel specially concerned about certain experiences following a described sequence of events. We do so by putting ourselves in the subject's place and examining whether we would feel specially concerned about the ensuing experiences. In this way, examination of our feelings of special concern provides a important (though not always conclusive) methodological tool for
exploring our beliefs about survival. Some reductionists endorse very similar methodologies. Unger proposes the "Avoidance of Great Future Pain Test" as an accurate though not infallible test of these beliefs. (Unger, 1990, p.27) That is, if one would suffer a lesser pain now to avoid there being a greater pain later, then this preference is indicative of a deep belief that one would survive the sequence of events in question. For Williams, to assess whether or not our fears can "reach through" the sequence of events to the ensuing horrible experiences helps to determine whether or not one would survive. (Williams, 1975, p.191) According to what I call his methodology of projection, we may do so by attempting to engage in the "imaginative projection" of ourselves and our feelings of special concern into the ensuing experiences. (Williams, 1975, p.194) How exactly can methodologies that accord so well with my own lead to such different (i.e. reductionist) conclusions?

Clearly, there should be some commonsense constraints and qualifications on this methodology of thought experiments. Though thought experiments don't need to involve scenarios that are likely to occur or are technically feasible (if they did, then they would be more like real cases than thought experiments), thought experiments need to be reasonably realistic. (In my view, Parfit's teletransportation thought experiment easily meets this requirement.) It is also important to remember that there are criteria (such as explanatory power, coherence, simplicity, and even elegance) for evaluating philosophical theories other than
compatibility with our reactions to thought experiments. I consider it a mistake to discount a thought experiment just because our reaction conflicts with a deep belief or a widely-accepted philosophical view. Indeed, a methodology of thought experiments balanced with the above considerations should provide a means of scrutinizing even such well-accepted views. I believe that proponents of reductionism tend not only to misinterpret thought experiments but to ignore these constraints and qualifications. Later, I argue that certain key reductionist thought experiments (such as Unger's "spectrum of decomposition") are uncompelling if not illegitimate. (See Chapter 6.)

Of course, the critic may even reject the general methodology of using thought experiments to probe our supposed beliefs. Mark Johnston's views are characteristic of this tendency. Johnston calls the methodology of consulting our intuitions about cases (as exemplified by the practice of using our reactions to specific cases to assess purported explications of knowledge) "the method of cases." (Johnston, 1987) He identifies Williams as a prime offender in applying this supposedly questionable methodology to the investigation of personal identity. I consider it to be a mistake to impose such constraints on philosophical investigation. Ultimately, one's commitment to certain beliefs is of a psychological nature. Thought experiments serve to bring into focus the relationships between our beliefs and to sharpen our

---

7 Kathleen Wilkes argues for a methodology that relies on actual rather than imaginative cases. See Wilkes, 1988, Chapter 1.
appreciation of our relative commitments to these beliefs. A philosophical argument needs only be convincing, not to have a certain pedigree beyond this. To renounce thought experiments is to leave oneself with little reason to prefer one theory over another.

Of course, one may object that to uncover our beliefs about survival and our reactions to cases is not to uncover metaphysical facts about survival. However, the answer that we give to the metaphysical problem may depend on our beliefs about whether or not one survives in certain cases. All we can expect is the best answer given what we presently believe. There is no guarantee that our deep beliefs won’t change to accommodate a different answer. (However, I suspect strongly that our reactions will continue to be robust in light of scientific discoveries, in light of what we may learn about how the brain contributes to our mental lives.) Reductionism may seem very attractive until we grasp how deeply it conflicts with our commonsense beliefs about survival.

Clearly, the metaphysical problem of personal identity and survival and the mind-body problem are closely related. Because obviously I cannot defend in this thesis a position on the mind-body problem, I should be clear about the assumptions about experiences that I accept. I assume that Nagel, Searle, Kripke, Flanagan and numerous other experts on the philosophy of mind are right in rejecting those forms of materialism that deny that there are experiences or claim that experiences are publicly observable. I also deny the identity theory that experiences are nothing but brain states, at least in the naive sense of involving only
properties that are purely physical. (Unavoidably, I do later consider some arguments against the identity theory.) I am somewhat more sympathetic to views, such as those of Searle and Nagel, according to which physical states (understood broadly) have irreducible mental properties. (Searle, 1994; Nagel, 1986) As I think the thought experiments that I employ (together with these assumptions) show, reductionism with regard to survival is not nearly so attractive as its proponents suppose.
PART I

QUALITATIVE THEORIES
1. Introduction

Recall that a qualitative theory analyzes X's survival as Y in terms of qualitative relations involving the contents of the experiences, the psychological make-up, and the physical characteristics of X and Y (or perhaps of intermediate person stages). A pure psychological reductionist theory (a pure content-based theory) analyzes X's survival as Y in terms of psychological relations (content-dependent relations) involving the contents of the experiences and the psychological make-up of X and Y (or perhaps of intermediate person stages). Also recall that a content-based theory may add a causal requirement, a requirement that the experiences and psychology of Y be appropriately caused given what realizes the experiences and psychology of X.

It is tempting to suppose that survival consists in qualitative (and especially content-based) relations. Like Parfit, one may think that the traveller survives teletransportation. The replica on Mars will be psychologically indistinguishable from the original. He will seem to remember the traveller's life on Earth and indeed entering the teletransporter. He will have all of the traveller's plans of what to do on Mars. He will also have the special knowledge, beliefs, and character to make sense of such memories and to carry out these plans. Though physical characteristics may be of less importance for one's survival than
psychological ones, the exact physical similarity of the replica to the original may enhance the replica's ability to carry out these plans. One may think that by virtue of the psychological (and generally qualitative) resemblance to the replica on Mars, the traveller survives as this replica.

In trying to understand the qualitative view, I focus on the memory theory (the view that identity consists in memory linkages between person stages), the most discussed and criticized content-based theory in the personal identity literature. Because this theory reveals many of the difficulties of content-based theories and almost all such theories appeal to considerations of memory, this theory is both instructive and representative. In this chapter, I consider a number of objections to the memory theory and how these are handled in the literature. In the next chapter, I consider whether or not the qualitative view (and the memory theory in particular) can explain how special concern can be justified.

2. Locke's Memory Theory

To gain an understanding of the content-based view and its attractions and difficulties, it helps to consider the origins of the view of experience-memory as a criterion of personal identity and how this theory has evolved to meet its difficulties. Because Locke's memory theory and some of its twentieth century refinements are theories of personal identity (rather than survival), I start
by considering the memory theory as a theory of identity.¹

According to Locke, "...as far as...[a person's] consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person..." (Locke, 1975, II, xxvii, 9) The traditional interpretation of Locke, that Y at a later time is the same person as X who had an experience at an earlier time just in case Y remembers this experience of X, can be traced to the works Reid and Butler. (Reid, 1785; Butler, 1736) As Jonathan Bennett writes: "There is virtual unanimity among readers of Locke that what he calls...consciousness...[of an earlier experience] is just episodic memory." (Bennett, 1994, p.109) Because I am interested in the theory itself, I do not consider whether or not such a theory is correctly attributed to Locke.

In support of his view of personal identity, Locke offered his famous case of the prince and the cobbler. Locke wrote: "...should the soul of a prince, carrying with it the consciousness [i.e. memory] of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, everyone sees he would be the same person with the prince..." (Locke, 1975, II, xxvii, 15) If the person now with the cobbler's body can really remember the prince's experiences, then we may find it plausible to suppose that this person is the prince.

¹ Locke certainly considered his theory as a theory of personal identity. The modern refinements of this theory, due to Quinton and Grice, are presented as theories of identity. (Quinton, 1962; Grice, 1941) Memory here means event memory rather than factual memory or remembering how.
Certainly, it is tempting to feel that there is a close relationship between memory and personal identity. Indeed, it is easy to give a psychological explanation of the attraction of the memory theory. Because our ostensible memories of experiences (and the indexical aspect thereof) tend to represent these experiences as our own, we tend to believe that we had these experiences. If these memories are supposedly of events long ago, we tend to believe that we were present at and that our identity extends backwards to these events. For this reason, a person with apparent memories of having been a given person will certainly believe that he is that person. Locke wrote: “Had I the same consciousness that I saw the ark and Noah’s flood... as that I write now; I could no more doubt that I who... viewed the flood... was the same self... than that I who write this am the same myself now whilst I write...” (Locke, 1975, II, xxvii, 16) Clearly, we take our memories of past experiences as strong evidence of our own identity over time. Moreover, though we do not have direct access to the memories of others, we take memories that they reveal to us as strong evidence of their identity. Because considerations of memory may provide such strong evidence of our own identity and of the identity of others, it is attractive to suppose that personal identity necessarily involves memory.

As stated, Locke’s view is open to a number of obvious and seemingly easily remediable objections. Of course, one doesn't always have occurrent memories of all of one's supposed past experiences. In saying that “can be extended” rather than “is
extended," Locke seems to have recognized this fact. (Perry, 1975a, p.16) To say that Y remembers an experience means that he could have a memory of the experience (i.e. if he were prompted to remember the experience etc.). For instance, though, while concentrating on writing, I may have no occurrent memories of the video I watched last night, if I were to take a break, I could perhaps have such memories.

There are also memories of experiences that have been irretrievably forgotten. Locke's official view is that only past experiences that one can remember are one's own. For Locke, one is not responsible for a crime that one does not remember. (Locke, 1975, II, xxvii, 26) Presumably, this is so because one who does not remember the crime is not the same person who committed the crime. However, this view does not seem reasonable. Though I may not remember eating a hot dog at the time, I may clearly remember Gretzky scoring the winning goal at a hockey game. Presumably, there were not two persons inhabiting my body at that moment of the game. If it was I who witnessed the winning goal, then it was I who was eating the hot dog. (I am assuming that the identity of a person at a time is unproblematic.) So if Y can remember an experience of X, then that should be sufficient for Y to be the same person as X.

It is perhaps a more serious problem that there may be whole stretches of our lives of which we have irrevocably lost all recollection, i.e. stretches for which it is not the case that we could have a memory. For instance, I may have absolutely no
recollection of any experiences from an evening of excessive drinking or from a day during which I suffered a severe blow to the head. Locke anticipated this problem of drunkenness/amnesia, and seems to have offered a definite response. (Locke, 1975, II, xxvii, 20) According to Locke, the person who had the experiences during the stretch that I am unable to remember is simply not me. We may think that this response resorts to an extreme expedient. On arriving at work, I may have no recollection (even under hypnosis) of the drive. It seems implausible to deny that the numerous perceptions of the drive and the thoughts that then preoccupied my brain were my own. It seems obviously wrong that an experience that I do not remember cannot have been mine.

Fortunately, there is an alternative response to the problem of irretrievable forgetting - the conditional analysis - which may seem much more attractive. When we say that Y "could have a memory of an experience E," we mean that Y would, given certain conditions, have a memory of E. As suggested above, Locke seems to have accepted this way of dealing with the fact that our consciousness is not always crowded with all of our available memories. It is tempting to extend this conditional analysis to the case of experiences that have been blotted from one's memory. Perhaps, we could say that, because I would or could have memories of the experiences during the stretch if I had not consumed that bottle of tequila, knocked my head so hard, or simply been so

---

2 Perry proposes a similar analysis. (Perry, 1975b, p.136)
distracted during the drive, these experiences were my own. The conditional analysis seems to provide a plausible way of dealing with this difficulty.

Though Locke did not consider this option, it is an option that would have been reasonable for him to take. I do not claim that Locke would have accepted these modifications aimed at solving the problem of forgetting (irretrievable and otherwise), only that it would be reasonable for him to accept these and that they don't violate in any immediately obvious way the spirit of his theory. I will still refer to this view, as modified, as the Lockean view.

3. Problems for the Lockean View

Clearly, Locke's theory as stated (and modified) above suffers from some serious difficulties. I will focus on three difficulties that were noticed by his earliest critics and that have greatly concerned the contemporary debate on personal identity. These are the problems of transitivity, circularity, and duplication. I do not claim that these are exhaustive of the difficulties for this theory.

As Reid's famous example of the Brave Officer shows, the above relation that Locke held to be important cannot be a relation of identity because it is intransitive. (Reid, 1785) The brave young officer who led an heroic charge can remember being flogged as a boy for stealing apples, and the aged general can remember leading the charge, but the aged general cannot remember being flogged as a boy. Hence, the relation "can remember an experience of" is not
transitive. Because a relation of personal identity is transitive (i.e. if X is the same person as Y and Y is the same person as Z, then X is the same person as Z), this memory relation cannot be that in which personal identity consists.\(^3\)

Another of the earliest objections to Locke's memory theory was Butler's complaint that "memory presupposes, and therefore cannot constitute personal identity..." (Butler, 1736) Though expressed laconically, Butler's objection seems quite clear. As we naturally understand memory, one does not remember an experience unless one had the experience. Hence, this supposed analysis of personal identity over time in terms of memory seems to suffer from the glaring circularity that to satisfy this analysis the person at the later time must be the same person as the person at the earlier time. I call this the circularity problem.

As John Perry points out, the memory theory may also suffer from a more subtle circularity. (Perry, 1975b) Suppose that we accept the conditional solution to the problem of drunkenness/amnesia. When we say, for instance, that Y (the person who is now before us) would remember a dreadful accident if he (that very same person as at the earlier time) had not been inebriated at the time of the accident, we are appealing in a circular manner to considerations of personal identity. (Though

---

\(^3\) The case of drunkenness/amnesia also illustrates this problem. A drunk and the same man who wakes up hung-over but nonetheless sober the next day can remember the events of the previous morning. However, this sober man doesn't remember the experiences of the drunk.
implausible, Locke's response to this problem of drunkenness/amnesia does not face this difficulty.)

Reid also noticed that multiple persons in the future could have memories of the experiences of a single person. Thus, it is a consequence of Locke's theory that multiple persons at the same future time could be the same person as a single person at the present. Reid wrote that "if the same consciousness can be transferred from one intelligent being to another...then two or twenty intelligent beings may be the same person." (Reid, 1785, p.114) Clearly, this is absurd because the later persons would thus be the same as each other. I call this the duplication problem. Because Locke's original theory and some of its twentieth century refinements are theories of identity, this problem is a serious difficulty for these theories.

Though, to the best of my knowledge, none of these difficulties were resolved (or even tackled) during the time of Locke's early critics, all three of these difficulties are widely discussed in the twentieth century literature on personal identity. To illustrate the contemporary treatment of these difficulties, I focus initially on Parfit's theory. I also contrast his treatment of these difficulties with other available treatments.

4. Transitivity

To address the transitivity problem, Parfit distinguishes between direct and indirect memory connections. There is a direct memory connection between X at an earlier time and Y at a later
time if Y can remember (i.e. has or could have a memory of) an experience of X. (Parfit, 1984, p.205) There is continuity of memory between X and Y if between X and Y "there has been an overlapping [and temporally ordered] chain of direct memories." (Parfit, 1984, p.205) This relation (continuity of memory) is transitive, and hence an appropriate candidate for an identity relation. The young officer would be related to the boy and the general to the young officer, but the general would also be related to the boy. Thus, continuity of memory between person stages may be a suitable analysis of personal identity. I do not know of any serious criticism of this solution to the transitivity problem. However, as I explain in the next chapter, I do not see how continuity of memory (or any sort of qualitative continuity) can justify egoistic concern.

---

* Parfit's solution to the transitivity problem can be traced back to the works Grice and Quinton. Quinton proposed that person stages are stages of the same person if they belong to a temporally ordered succession of person stages such that each stage in the succession could have a memory of an experience of the previous stage. For Quinton, the relation in which personal identity consists is the ancestral of Locke's memory relation. (Quinton, 1962) Grice was also concerned with a further difficulty for Locke, the Senile General Case, in which the general can remember the experiences of the boy but not those of the young officer. Grice proposed that person stages (what he called "total temporary states") are stages of the same person if these stages belong to a sequence (not necessarily temporally ordered) of person stages such that each in the sequence either could have a memory of an experience of its predecessor or vice versa. (Grice, 1941)*
5. Circularity

Parfit's approach to the problem of circularity is based on his concept of quasi-memories (q-memories). On Parfit's definition of q-memories, a person Y has a q-memory of a past experience E (Y q-remembers E) if (1) Y seems to remember having experience E, (2) someone X did have experience E, and (3) Y's apparent memory is causally dependent on experience E in the right kind of way. (Parfit, 1984, pp.219-223) Parfit's use of "seems to remember" in his condition (1) is vague, and we might clarify this condition by substituting "is disposed to have [is capable of having] an apparent memory of" ("has an apparent memory of" for short) for this expression. Condition (3) requires that an appropriate causal process leads from X's having experience E to the later corresponding apparent memory (i.e. Y's having an apparent memory of E). Clearly, it is an important question just what would count as an appropriate causal process, that is what we should require as the causal basis for Y's apparent memory. Presumably, the ordinary neurological process that leads from an experience to the apparent memory thereof would satisfy the requirement for such a process. I defer until later the question of what else might count as an appropriate causal process.

Clearly, this definition of "Y q-remembers E" does not depend on the identity of a person over time. To say that Y q-remembers experience E (or would q-remember E if circumstances at the time of

---

5 Perry offers a similar strategy. (Perry, 1975b)
E had been different) does not imply that Y is the same person who had experience E. For example, to say that a person now would q-remember the accident if the person who experienced the accident had not been inebriated at the time does not imply that these persons are the same person. Q-memories are more general than memories. Presumably, an ordinary memory is a q-memory of one's own past experience. Also, depending on how we understand "appropriate," the suggested definition of "q-remembers" does not presuppose that Y's q-memory of E has the normal causal basis for a memory or even that the q-memory is associated with the numerically same brain as was E. A process of replication or a memory implant may satisfy condition (3). Thus, the person who materializes on Mars may have q-memories of experiences of the traveller whose body the teletransporter destroyed on Earth.

Say that there is a q-memory connection between a person X at an earlier time and a person Y at a later time if Y q-remembers an experience of X. For Parfit, when there are enough q-memory connections ("if, over each day, the number of direct quasi-memory connections is at least half the number in most actual lives..."), there is strong connectedness of q-memory between X and Y. (Parfit, 1984, p.222) There is this strong connectedness when there are sufficiently accurate and numerous q-memory connections between the experiences of X and the q-memories available to Y. Say that there is continuity of q-memory between X and Y if there are overlapping chains of strong connectedness of q-memory between X and Y. Then we may say that X and Y are the same person if there is continuity of
q-memory between X and Y. This refinement of the memory theory does not suffer from the circularities diagnosed above.\(^6\)

6. Objections to Parfit's Response to the Circularity Problem

There are two notable objections to Parfit's response to the circularity problem. The first of these is a problem that Antony Flew raised initially for the Lockean view. (Flew, 1951) Flew posed a dilemma for Locke. If Locke requires that, for X to be the same person as Y, Y must truly remember an experience of X, then (as we have seen) the theory is viciously circular. If, on the other hand, Locke (or more contemporarily Parfit) requires only that Y have a (perhaps merely) apparent memory of an experience of X, then (as Flew argues) it unreasonable to believe that this makes Y the same person as X. Surely, if one has an accurate apparent memory of the

\(^6\) Parfit considers the possibility that personal identity can also be analyzed in terms of intentions. That is, to say that X at earlier time is the same person as Y at a later time is just to say that Y carries out intentions of X. (Parfit, 1984, p.205) This view faces problems that are analogous to those faced by the memory theory, problems that can be resolved in similar ways. First, this intentional relation is intransitive and hence cannot serve as an analysis of personal identity. (It may be the case that Y could carry out an intention of X and that Z could carry out an intention of Y without it being the case that Z could carry out an intention of X.) Say that there are direct intentional connections between X and Y if X intends some actions or possible actions of Y. Say that there is continuity of intentions between X and Y if X and Y are linked by a chain of direct intentional connections. Continuity of intentions between person stages is transitive and may thus provide a suitable analysis of personal identity. Second, the intentional theory suffers from a circularity similar to that which afflicts the memory theory. To say that X intends an action of Y implies that X and Y are the same person. In order to avoid the problem that intentions presuppose personal identity, Parfit employs the term "quasi-intentions" in a manner that is analogous to his use of "quasi-memories."
Napoleonic experience of rallying the French troops at Austerlitz, then this doesn’t make one Napoleon. Such an apparent memory is simply delusional.

Of course, Parfit is well aware that if Y has a single q-memory of an experience of X, then it is implausible to suppose that this makes X and Y the same person. He accepts that such a q-memory would be delusional. As he writes: “...we should not claim that, if I have an accurate quasi-memory of some past experience [or even a few such q-memories], this makes me the same person who had this experience.” (Parfit, 1984, p.222) To be the same person as someone in the past, one would need to have “countless quasi-memories of earlier experiences.” (Parfit, 1984, p.222) Only if Y is linked to X by chains of myriad q-memory connections, is it reasonable to suppose that Y is the same person as X. However, we may doubt that numerous merely apparent memories are any more capable of constituting identity than is a single apparent memory.

Perhaps, what is needed to solve the false memory problem is an adequacy condition for veridical (i.e. genuine) memories that does not appeal to considerations of personal identity. According to C.B. Martin and Max Deutscher, for someone to truly remember an experience “...the causal chain between the past observation and the present representation [i.e. the memory] of it [the observation or experience] should continue without interruption within the body.

Likewise, Parfit considers it unreasonable to suppose that X is the same as Y if X merely q-intends an action or possible action of Y. However, if there are multitudinous q-intentional links between X and Y, then this may make X and Y the same person.
of the person concerned." (Martin and Deutscher, 1966, p.182) The need for such a causal chain may seem clear. If one has an experience, forms a memory as a result, irretrievably forgets this memory, and a similar representation (apparent memory) is later reimplanted (surgically, chemically, or hypnotically), then this may not count as true remembering. They argue that memory is impossible without a physical "memory trace" leading from the experience to the memory thereof, that is, the usual sort of physical trace preserved in the numerically same brain persisting over time. Without the usual causal basis, an apparent memory would be illusory. Perhaps Flew would accept such a memory trace requirement as that which distinguishes genuine from false memories. Though this may require that the representation be supported throughout by the same brain, it does not require that X and Y be the same person. (I revisit in Section 9 below the issue of what might count as a suitable causal basis.)

The second difficulty is that it is implausible that personal identity can consist in retention of apparent memories alone. Marya Schechtman argues convincingly that without accompaniment by appropriate beliefs and character traits etc., even multitudinous and highly faithful q-memories cannot secure personal identity. (Schechtman, 1990) On their own, such apparent memories could seem very inappropriate indeed. For example, without a range of appropriate beliefs, desires, special knowledge, abilities, manners of thinking etc., even the most vivid of Napoleonic memories would be obviously delusional. (She makes a similar point with regard to
q-intentions.) However, Parfit accepts that connections involving beliefs, character traits etc. (i.e. when these psychological features "continue...to be had") may also be important for identity, that what is involved may be more complex than continuity and/or connectedness of q-memory. (Parfit, 1984, pp.205,206)

Parfit distinguishes between two general psychological relations that he believes to pertain to identity: psychological connectedness and psychological continuity. "Psychological connectedness is the holding of particular direct psychological connections." (Parfit, 1984, p.206) Person stages are strongly psychologically connected if there is a high degree of psychological connectedness between them (perhaps, over a day, at least half of the normal degree of psychological connectedness). (Parfit, 1984, p.206) "Psychological continuity [between person stages] is the holding of overlapping chains of strong [psychological] connectedness." (Parfit, 1984, p.206) He considers the connectedness and continuity of psychological factors other than just apparent memory to be involved in personal identity.  

---

8 Parfit thinks that physical similarity may also have some importance but only in as much as it is needed to maintain the relevant psychological relations. A radical physical change (such as replacement of one's body with one of the opposite sex) may result in too rapid and extreme a psychological change for one to retain one's identity. (Parfit, 1984, p.285) For this reason, "It may be rational to want the body of my Replica to be like my present body." (Parfit, 1984, p.286)

9 Unlike Parfit, Perry treats only connectedness (similarity) as important. (Perry, 1976)
7. Identity Is Not What Matters

Parfit's solution to the duplication problem is to grant that (because they may not be one-to-one) relations involving memory (and/or other qualitative relations) do not constitute personal identity but to deny that, in itself, identity really matters. Though, at a given future time, there can be at most one person that is the same as one's present self, one can have more than one survivor at this future time. Perhaps, this distinction between survival and identity is Parfit's most important contribution to the debate on personal identity.

Clearly, most of us are deeply committed to the belief that our identity matters greatly. It is our identity that seems to distinguish certain experiences (those that will be ours) that will be important to us in a special way from experiences (those of others) that will not be important to us in this way. According to Parfit, identity does not have the deep importance that we tend to accord it (and that the Lockean view seems to recognize). In an example that he offers in support of this claim, memory linkages (and other qualitative relations) are duplicated. (Parfit, 1984, pp.253-261) I will consider a variant of this example.  

Suppose that of three identical triplets involved in a car crash, two (Y and Z) suffer massive brain damage but no damage to the rest of their bodies, and that the third (X) suffers a fatal injury to his body that leaves his brain fully intact. To minimize

---

10 Parfit's example and my variant thereof are similar to a case introduced by David Wiggins. (Wiggins, 1967, p.50)
the impact of this tragedy, surgeons decide to remove X's brain from his dying body, to divide the hemispheres, and to transplant these into Y's and Z's brainless skulls. (We may assume that Y's and Z's brainstems are intact, and hence that only the cerebral hemispheres of X's brain need to be transplanted.) Given the strong empirical evidence that memories are dispersed throughout the higher brain, it is reasonable to suppose that Y and Z will have the same apparent memories and that there will be complete connectedness of memory between X and both Y and Z. Generally, Y and Z will exactly resemble X psychologically (in terms of memories, plans, desires, beliefs, character traits etc.).

Parfit argues that the persons following the surgery with the bodies of Y and Z must both have what matters (or should matter) from the point of view of X before the surgery. That is, X survives as both Y and Z. As is well-known, people have survived (at least in the sense that their brain and body appear subsequently to support consciousness) the excision of an entire cerebral hemisphere, in some cases without serious intellectual deficit.  

---

11 Some researchers even note an improvement in intellectual function following hemispherectomy as a treatment for intractable epilepsy. L.T. Dunn and J.B. Miles observe "an improvement, often marked, in...cognitive functioning following the procedure." (Dunn and Miles, 1995) Eileen Vining et al also note such an improvement in pediatric epilepsy patients. (Vining et al, 1997) They find that following the surgery, many of their patients showed some "improvement in intelligence" and none of these patients were "worsened by the surgery." Indeed, the researchers feel "awed by the apparent retention of memory after removal of [either] half of the brain...and by the retention of the child's personality and sense of humor." They speculate that the intellectual improvement in these children is partly the result of a decrease in seizure activity, and a corresponding decrease in the administration of
Also, it is hard to imagine how it can matter whether the preserved hemisphere remains in its present body or is transplanted into the vacant cranial cavity of a physically identical though numerically different body. Hence, X would survive if half of his brain were transplanted into either Y's or Z's skull and the other half were destroyed. But how can it make X's situation worse if the other half of his brain is transplanted into the remaining body rather than destroyed? Parfit quite rightly wonders how a double success can be a failure.

Though both Y and Z have everything that would have mattered in the normal preservation of the preoperative patient's identity over time, Y and Z following the surgery cannot both be identical to X before the surgery, because if they were the absurd conclusion would follow that they would be identical to each other. Parfit concludes that identity is not what matters.\(^\text{12}\)

Though, in this example, there is considerable physical medications with deleterious side-effects, and perhaps partly the result of the cessation of "the continuous electrical activity of...[the] severely dysfunctional hemisphere...[that] interferes with the function of the other, more normal hemisphere." Apparently, having only half of a brain may be less impairing intellectually than having a whole brain with one side "badly misfiring."

\(^\text{12}\) As he claims, the above procedure is at worst only technically infeasible rather than deeply impossible. In fact, patients who have had the connections between their cerebral hemispheres severed as a treatment for epilepsy show evidence of having two separate centres of consciousness. (Sperry, 1966) So even if the technical problems of the above procedure are insuperable, there seem to be actual cases in which the above psychological connections are in effect duplicated. Because this example is a near actual case, it should satisfy even those with qualms about drawing conclusions on the basis of imaginative cases.
continuity involved, Parfit does not require that there be continuity of matter between what realizes one's psychology and what realizes the psychology of one's supposed survivors. All that is required is that one enjoy appropriate qualitative relations to these survivors and that these relations have "the right kind of cause." (Parfit, 1984, p.215) Thus, one may survive teletransportation. Even if Y and Z both result from X due to a teletransporter malfunction (rather than the above physical division), then X may still survive as both Y and Z.

As Parfit recognizes, survival as understood in terms of continuity and connectedness of an array of psychological features is (as is any qualitative relation) very much a matter of degree. Apart from any claims of identity, what might be appropriate to say is that the original person survives to a certain (and perhaps varying) extent as all of its survivors (or continuers). If a person Z at the later time is more strongly qualitatively related to X than is Y, then we may say that X survives as Z to a greater extent than as Y.

Of course, if one person at the later time is more strongly qualitatively related to X, is a closer continuer of X, than are the other candidates, then we may adjudge this person to be the same person as X. We may do so on the basis of a closest continuer schema, such as that of Robert Nozick. (Nozick, 1981) According to

\[13\] In fact, Parfit employs his Psychological Spectrum, Physical Spectrum (discussed in Chapter 6), and Combined Spectrum thought experiments in support of this view. (Parfit, 1984, pp.231-243)
this schema, X is the same person as Y if and only if Y is a sufficiently close continuer of X, and Y is the closest continuer of X. In the case of a tie (or near tie) between Y and Z as continuers of X, neither would be the closest continuer. According to Nozick, Y can be a close enough continuer of X by virtue of content-based relations between X and Y, even if their bodies are not physically continuous. For Parfit, which if any of the continuers (i.e. of the qualitatively qualified candidates) is thus adjudged to be the same as the original person is a largely arbitrary matter of trivial importance (or at least of importance that does not pertain to the issue of survival). Once, we know the facts about the qualitative relationships these continuers bear to the original person, we know everything about the survival of this person. Which one is adjudged the same as the original doesn’t tell us anything additional.

8. Objections to the View that Identity Is not What Matters

Both Nozick and Parfit accept that it is reasonable to identify one with one’s closest continuer. However, unlike Parfit, Nozick considers judgements concerning the identity of persons to be of great importance. He writes: "...we are not willing to think that whether something is us can be a matter of (somewhat arbitrary) decision or stipulation." Nozick treats identity as if it is what matters. Clearly, he does not just intend his schema as a means for deciding the quasi-legalistic matter of which among qualified survivors is to be considered the same as the original.
If one's survival does consist in qualitative relations, then Parfit's position seems the more reasonable. Certainly, there is a sense in which considerations of identity should matter to us greatly, even when there is more than the usual one candidate for being the same identical person. A person may enjoy certain advantages by being adjudged the same as the original. Who among the candidates is to move back in with the original person's wife and to take possession of his holdings may depend on which of the candidates is adjudged to be the same person as the original. However, given that all of the candidates are survivors of the original, such considerations of identity do not pertain to the issue of survival.

Even if one does reject (rightly, I think) Nozick's view, one may still not consider Parfit's theory to provide a satisfactory response to what's really at issue with the duplication problem. Perhaps, the problem is not the logical one that two different persons cannot be the same person but that it is counterintuitive to suppose that one may survive in any meaningful sense as more than one person at the same time. Perhaps, the prospect of surviving as a duplicate is not just unsettling but is simply incoherent. Swinburne denies that one can survive fission as both of the resulting individuals, that (in Parfit's case) X can survive as both Y and Z. Psychological similarity (such as that of X to both Y and Z) may be evidence (though apparently not entirely conclusive evidence) of survival, but it is not that in which survival consists. He claims that in the case of fission: "Maybe I
go where the left hemisphere goes; and when my right hemisphere is separated from the left hemisphere and comes to control a body by itself, either a new person is formed, or the resulting organism...is really a very complicated non-conscious machine.” (Shoemaker and Swinburne, 1984, p.19) A forteriori, one cannot survive as multiple duplicates produced by a defective teletransporter.

9. Causal Requirement

Many psychological reductionist theories add a causal requirement, a requirement that the content-based relations (psychological connectedness and/or continuity etc.) in question have an appropriate causal basis.\(^\text{14}\) That is, such a theory may require that the psychological state of one's future survivor be linked by a causal chain to one's present psychological state and experiences. Perhaps, the addition of a causal requirement may solve the problem of false memories (i.e. by filling out condition (3) above). Another rationale may be to solve the duplication problem. Perhaps, an analysis tightened in this way may rule out the unsettling prospect of survival as more than one individual at the same time. (I consider a further rationale in the next chapter.) The main point of disagreement among proponents of the content-based view is just how robust such a requirement needs to

\(^{14}\) Any qualitative theory can have a causal requirement, a requirement that the qualitative relatedness in question have an appropriate cause. For convenience, I confine discussion to the case of content-based theories.
be.

The specific causal requirements vary greatly. At one extreme, David Lewis holds that the later state should be causally dependent on the earlier state in the "normal" way (i.e. by the earlier state leading to the later state through a continuing neural process within a single brain and body). (Lewis, 1976) At the other extreme, Kolak and Martin deny the need for any causal requirement. (Kolak and Martin, 1987) Parfit considers the requirement that the later state be causally dependent in a reliable manner on the earlier state but rejects this requirement in favour of one on which the causal basis need not even be reliable. For Parfit, "The right kind of cause could be any cause." (Parfit, 1984, p.215) That is, connectedness/continuity may secure one's survival even if this "does not have its normal cause." (Parfit, 1984, pp.284,285)\(^{15}\)

It is perplexing that Parfit would draw the line at this point. As Kolak and Martin argue convincingly, it is implausible to suppose that there is a significant difference with regard to survival between the case of teletransportation and the case in which by random chance one's body dematerializes on Earth and an exact replica materializes on Mars. (Kolak and Martin, 1987) Though I won't consider their argument in any more detail, their position seems entirely reasonable. If there really is no need for the later state to be dependent on the earlier one in the normal way (or

\(^{15}\) Nozick agrees with Parfit. (Nozick, 1981) Like Kolak and Martin, Perry seems to reject the need for a causal requirement altogether. (Perry, 1976)
least a near normal way), then it is hard to understand what could be the rationale for a causal requirement at all.

As suggested, a content-based theory can block the duplication problem by including a causal requirement. However, to solve this problem, such a requirement would have to be very strong indeed. Arguably, because the process of teletransportation could conceivably be modified to produce multiple survivors of the original, this requirement would have to be sufficiently restrictive so as to rule out survival of teletransportation. Clearly, Parfit's causal requirement does not satisfy this need. It is hard to imagine how a causal requirement that is weaker than the normal causal basis can rule out survival as a duplicate. At the very least, there would have to be considerable physical continuity between the body that supports one's psychology and the body that supports the psychology of one's future survivor. (It is hard to see how even a requirement of considerable physical continuity can rule out double survival of fission, such as (in Parfit's case) X's survival as both Y and Z.) Likewise (if a veridical memory is one that is secured by the usual process), the false memory problem can be solved only by making the causal condition very strong. As we shall see, a supposed qualitative theory that can handle these difficulties might be better described as a physicalist theory.
CHAPTER 4
Qualitative Relations and Special Concern

1. Introduction

The most serious difficulty for the qualitative view (and for the content-based view in particular) is to explain why one should be specially concerned about the future experiences of someone who will be qualitatively related to one's present self. However, despite any other difficulties that this view may have, if we could overcome this difficulty (i.e. explain in terms of this view how egoistic concern can be justified), then there would be good reason to accept this view.

In this chapter, I argue that this explanatory requirement is reasonable and that there is a strong case that it cannot be met. As we shall see, Butler recognized the need to be able to explain in terms of an analysis of personal identity why we should care about our future experiences. I also examine John Perry's attempt to explain in qualitative terms how such concern may be justified and argue that this attempt fails. The problems of this attempt illustrate the difficulty in explaining in terms of the qualitative view how egoistic concern can be justified. Perhaps, a qualitative theory with a strong causal requirement may be able to account for justification of egoistic concern. However, I argue that, if the causal requirement is sufficiently robust to have any chance of solving this problem, then the resulting theory is better treated as a physicalist theory.

2. Butler's Principle
As Butler wrote, "...when it is asked wherein personal identity consists, the answer should be...that all attempts to define [it], would but perplex it." (Butler, 1756, p.99) Butler made this claim as a criticism of Locke's view, according to which personal identity can be analyzed in terms of memory. However, clearly, this claim seems to apply to any analysis of personal identity. What exactly could Butler have meant? Surely, if we could establish that personal identity consists in a qualitative relation or in physical continuity, then (far from perplexing identity) we would come a long way to understanding what, left unanalyzed, is a very perplexing notion.

I think Butler's point becomes much clearer if we remember that what we are seeking is an account of survival on which egoistic concern may be justified. Though Butler did not talk explicitly of egoistic concern, there is reason to suppose that such concern was what he had in mind. For instance, he also wrote that "personal identity has been explained so by some [presumably by Locke], as to render the inquiry concerning a future life of no consequence at all to us, the persons who are making it." (Butler, 1756, p.99) It appears that what Butler was claiming is that if Locke's analysis of personal identity were correct, then we would have no reason to be specially concerned about our future experiences. As I explain in the next section, there is good reason to suppose that he was correct.

I believe that this claim that the memory theory cannot explain how egoistic concern can be justified can be generalized to
any supposed analysis of personal identity (or survival). If one claims that personal identity (or survival) can be analyzed in terms of a certain relation, then one has to explain in other terms why this relation matters. But then one has to explain in still further terms why these terms matter, and there ensues an infinite regress. Hence, any attempt to analyze personal identity is doomed to failure. I will refer to this claim as Butler’s principle.¹

Perhaps it makes sense to think of Butler’s principle as encompassing a number of claims that together lead to this regress and hence imply the futility of attempting to analyze one’s identity and survival. (1) Unless it is self-evident that we should be specially concerned about the terms of a supposed analysis, whoever offers the analysis has the burden of explaining why the terms of the analysis should matter in this way. (2) An explanation of why we should care specially about the terms of an analysis must be given in still other terms. (3) It is never self-evident that we should be specially concerned about the terms of an analysis. (4) If (1), (2), and (3) are true, then any attempt to analyze identity and survival will lead to an infinite regress that can only “perplex” these further. That is, no analysis in reductionist terms can be given of identity and survival.

I should clarify that, as I use the term in (1) and (3), to

¹ One may read Butler’s principle to assert that personal identity cannot be analyzed so as to explain how one’s identity can matter in other ways as well. One should recognize that whether or not Butler’s principle is correct may depend on how one’s identity is understood to matter.
say that something is "self-evident" means only that (perhaps on reflection) it is obvious rather than that it is an a priori truth. For instance, it may be self-evident in this sense but not an a priori truth that hard work improves one's chances of success. It is possible that someone who does not immediately see something as self-evident in this sense may be convinced or otherwise come to see it as such. It is also important to understand that Butler's principle does not exclude the possibility that one's identity and survival could be analyzed in terms of the persistence of some unanalyzed subject of consciousness. I believe that, when the justification of egoistic concern is at issue, Butler's principle is correct, and hence that reductionism with regard to survival as a sentient being is wrong. I do not claim that Butler's principle is correct for other understandings of what matters.

Just how strong a case is there for each of the above four claims? Claims (2) and (4) should be uncontroversial. (2) seems to derive from a reasonable understanding of what would count as an explanation, and (4) is a logical truth. (1) seems to be an obvious requirement of rationality. If something is not self-evident, then the demand for an explanation is entirely reasonable. Of these claims, only (3) may not seem so obviously correct. It is not obvious that the resulting supposed regress cannot terminate in terms about which it is self-evident that we should care specially.

What should be uncontroversial is that if survival as a sentient being is to be analyzed in terms of a certain relation, then either it should be self-evident or we have to say why (in
other terms) one should care specially about whether or not this relation obtains. If, as Butler thought, personal identity is primitive and unanalyzable, then it may be a basic fact not requiring explanation that we have reason to care specially about experiences that will be ours. However, if we attempt to analyze personal identity in reductionist terms, then some reason is required why the terms of the analysis should be important in this way. According to Butler's principle, there is no hope of finding such a reason.

3. The Qualitative View and Egoistic Concern

According to the memory theory, as attributed to Locke, to say that I will suffer terrible pain in the future means that someone will suffer terrible pain in the future and that this person will remember, or be able to remember, my present thoughts and actions. To avoid circularity, the theory should require only that this person could have apparent memories (q-memories, to use Parfit's term) of these thoughts and actions. But why should I care specially about future experiences that contain or could contain apparent memories of my present experiences? It is certainly not obvious why it should matter to me in this way whether or not someone will have such memories of these present experiences. (Clearly, the response that I should care specially about the future experiences because they will be my own is viciously circular.) Rather, there are good reasons to doubt that the availability of apparent memories of our present experiences to
someone who has a later experience can justify our special concern for this later experience.

First, it seems clear that we feel a special concern for future experiences to be associated with our bodies no matter what these experiences may be like. All such future experiences normally matter to us a great deal and in a special way. A person who has been diagnosed as having a degenerative disease of the brain in the final stages of which he will have lost all his memory will dread the progress of the disease and its final agonies no less for knowing that his memory will be erased. (See Van Inwagen, 1990, pp.182-184, for similar arguments.) We seem to have a clear intuition that one can survive even without one's present memories and/or memories of one's present experiences.

Second, consider the transfer of an individual apparent memory from one person to another, for instance, the transfer of a single such memory from Adolph Hitler to Simon Wiesenthal. If apparent memories can on mass justify special concern, then transfer of even a single and insignificant apparent memory must justify some such concern for the experiences of the recipient. However, it is just as absurd that Hitler would thus have some reason to be concerned about the experiences of Wiesenthal as that Wiesenthal would thus bear some responsibility for the holocaust. Clearly, transfer of such a memory does not draw with it reason for special concern, responsibility etc. How then can retention of a collection of one's memories justify one's egoistic concern for future experiences? Butler's charge that the memory theory (as envisioned
by Locke) renders one's identity of "no [special] consequence" seems correct. (Butler, 1756, p.99) That is, if this theory is true, then there can be no basis for special concern for what will befall us.

Even Locke was equivocal about whether apparent memories of one's earlier experiences that could accompany a later experience justify one's egoistic concern for the later experience. For Locke, consciousness (i.e. memory) of an action is "a present representation of a past action..." (Locke, 1975, II, xxvii, 13) We cannot rule out the possibility that such a representation of a past action could "be transferred from one thinking substance to another." (Locke, 1975, II, xxvii, 13) For all we know, "...one intellectual substance may have...represented..., as done by itself, what it never did, and was perhaps done by some other agent..." (Locke, 1975, II, xxvii, 13) He also wrote: "And that it never is so...be best resolved into the goodness of God, who, as far as the happiness or misery of any of his sensible creatures is concerned in it, will not...transfer from one to another that consciousness which draws reward or punishment with it." (Locke 1975, II, xxvii, 13) Though one is accountable for actions that one apparently remembers, it would be an injustice (one that God would hopefully not allow), to punish (with painful experiences) one substance for an action of another substance. Clearly, the wrong substance could thus be unjustly punished only if it is the future experiences of the same substance as one's present substance (rather than of a substance that apparently remembers one's action)
about which one should care specially. Apparently, it wasn’t even so clear to Locke that memory can justify special concern.

Perhaps, even though the memory theory cannot explain why we should care specially about future experiences, a more developed qualitative theory involving a richer array of factors (memories, projects, beliefs, character traits, physical characteristics etc.) might be able to do so. However, it is certainly not self-evident that one should care specially about the future experiences of someone that is even psychologically and physically indistinguishable from oneself. Nor is it clear how to explain in qualitative terms why one should care specially about the future experiences of a qualitatively similar sentient being. Moreover, the above two objections to the memory theory are applicable to any qualitative theory. We feel specially concerned about future experiences to be associated with our body no matter how we will change (psychologically and/or physically). If such concern is justified, then qualitative relations cannot be that which justifies egoistic concern. Also, it is absurd that transfer of an

---

2 Gary Wedeking considers it hopelessly implausible (as do I) that the representation of a past action, in itself, draws with it responsibility for this action. (Wedeking, 1990, p.179) He also exposes an inconsistency in Locke’s account. As Locke’s Section 13 (II, xxvii, 13) shows clearly, Locke regarded the mental substance as agent. But how can one who may represent a past action be responsible for this action unless one is the agent thereof? (Wedeking, 1990, p.180) To resolve this inconsistency by denying that the substance is agent (and presumably accepting that it is the experiences that do or could represent one’s present experiences and/or actions about which one should care specially rather than that it is the past actions of one’s substance for which one is responsible) is to render the theory all the more implausible.
individual physical or psychological characteristic can carry with it any justification of special concern for the experiences of the recipient. How then can retention of these characteristics collectively justify such concern?

Parfit fails to explain adequately in terms of his theory (and indeed hardly addresses at all) why we should care specially about future experiences that we supposedly survive to have. Indeed, he seems very unclear about whether it is connectedness or continuity (of memories, intentions, beliefs, character traits etc.) that matters in survival. As he states, “Of these two general relations [connectedness and continuity], connectedness is the more important both in theory and in practice.” (Parfit, 1984, p.206) Indeed, when he writes that “My concern for my future may correspond to the degree of connectedness between me now and myself in the future,” he seems to be asserting that it is connectedness that justifies egoistic concern. (Parfit, 1984, p.313) However, he follows this with the claim that both of these relations “give me reasons to be specially concerned about my own future.” (Parfit, 1984, p.313)

Though he never explains why continuity should matter at all in one’s survival, he does suggest that psychological connectedness should matter because the specific elements of our psychology (our specific memories, intentions, beliefs, and character traits) may be important to us. (Parfit, 1984, pp.301,302) As seems clear, for instance, “most of us value some aspects of our character.” (Parfit, 1984, p.301) We may also value “the various relations between ourselves and others, whom and what we love, our ambitions,
achievements, commitments, emotions, memories...” (Parfit, 1984, p.284) If these specific aspects of our psychology (and the relationships with others that these aspects promote) matter to us, then they should be important in our survival. Because these aspects are preserved in replication, ordinary survival (in which there is the full measure of physical continuity) is little better than replication.  

Clearly, Parfit’s suggestion is not at all satisfactory. First, this suggestion equivocates on the word “matter.” To say that these aspects matter to us means that we value these aspects. However, to say that these aspects matter in our survival means that they have a role in securing our survival. Second, whether or not we should care specially about the retention of certain features of our psychology depends on whether we survive to enjoy these features. Hence, that we would enjoy them is no good reason to suppose that they help to secure our survival. He fails to explain adequately in terms of how the various qualitative factors interact why we should care specially about our future experiences.

4. Duplication Revisited

As suggested in the previous chapter, it is intuitively implausible to suppose that one may survive in any meaningful sense as more than one person at the same time. I believe that careful

3 For Parfit, the attachment to one’s actual body (as opposed to a perfect replica) may be “sentimental” as is one’s attachment to a particular ring over an exact replica. (Parfit, 1984, p.286)
examination of this intuition further undermines the qualitative view. In this discussion, it is helpful to make a clear distinction between a duplicate and a replica. By a replica of a sentient being, we may mean a sentient being with an exact copy of one's body. By a duplicate, I mean a replica that comes into existence before the original ceases to exist or that coexists with other replicas of the original. Henceforth, by a replica I mean a replica that is not a duplicate. We are apt to find the idea of survival as a duplicate to be unsettling, if not intuitively objectionable. We may think that it thus counts against a qualitative theory if it is a consequence thereof that one may survive as a duplicate. Why we should feel this way becomes clear if we remember that to survive to have an experience means to have reason to be specially concerned about this experience. The assumption that one survives as a duplicate has the consequence that survival is not something that merits our feelings of special concern.

Suppose that in a failed attempt to duplicate X, the original is destroyed but the perfect copy is produced as planned. Call the sentient being with this body $Y_1$. (This is essentially Parfit's teletransportation case.) The qualitative view implies that X

4 For example, the result of teletransportation is a replica of the original. The result of the modified process of teletransportation (what Parfit calls the Branch-Line Case) in which the original body is not destroyed is a duplicate of the original. Because it is reasonable to suppose that one's psychological state is completely determined by the state of one's brain, it is a reasonable assumption that a duplicate (and a replica) would be psychologically indistinguishable (in terms of memories, intentions, beliefs, character traits etc.) from the original.
survives as the replica $Y_1$ to the same extent as in ordinary survival, that is to the same extent that $X$ would have survived as the sentient being $Y$ with $X$'s body, had this not been destroyed.

Now suppose that to make up for this failure, two perfect copies are produced of $X$'s body. Call the sentient beings with these copied bodies $Y_1$ and $Y_2$. Again, the qualitative view implies that $X$ should be just as specially concerned about the anticipated experiences of $Y_1$ and $Y_2$ as he should be about the ensuing experiences in the case of ordinary survival. Assuming that one survives as a duplicate, $X$ should survive as both of these sentient beings.\footnote{If the qualitative theory has a causal requirement, I assume here that this requirement is not so strong as to rule out survival as a duplicate.}

But should $X$ really be as specially concerned about the experiences of each of $Y_1$ and $Y_2$ as he should be in the case of ordinary or unique survival? Suppose that before the procedure, $X$ learns that the experiences to be associated with $Y_1$'s body will be pleasurable and that those to be associated with $Y_2$'s body will be painful. Should $X$ look forward to or dread the future? Clearly, in the case of ordinary survival, if $X$ expects an experience of $Y$'s to be particularly pleasurable or particularly painful, then we believe that $X$ has reason to look forward to or to dread the experience. It is also clear, in the imagined case, how $X$'s survivors would react to their respective situations. $Y_1$ would consider himself fortunate to enjoy such pleasure, whereas $Y_2$ would
be distressed at his misfortune in having to endure such pain. However, because it is uncertain whether he will "find himself" as the one with the pleasurable or the one with the painful experiences (understood metaphorically and not to imply that he is truly identical with either), X should not look forward to or dread these experiences in the same way or to the same extent that we normally look forward to or dread pleasurable or painful experiences. Because it is uncertain exactly what to expect, X is not as justified in being specially concerned (or perhaps is not justified in being as specially concerned) about the future experiences of Y₁ or Y₂ as he would be about future experiences in the case of his ordinary survival. Perhaps, it may be reasonable to say that X has half the reason to care specially about the experiences of each of these survivors as about those of a unique survivor.

Now suppose that the duplicator is modified so that it not only destroys the original but produces a thousand perfect copies of X's body.⁶ Again, the qualitative view implies that X survives as all of the sentient beings, Y₁, Y₂,...,Y₁₀₀₀, with these copied bodies to the same extent that X would survive as Y. That is, X should be just as specially concerned about the anticipated experiences of all of these sentient beings as he should be about the ensuing experiences in the case of ordinary survival. Suppose

⁶ Antony Collins first raised the duplication argument (that numerous physical duplicates might share exactly the same memory links (more generally, qualitative links) with a person at an earlier time). (Perry, 1977, p.400)
also that only the experiences of \(Y_{1000}\) will be painful, the experiences of the other survivors being very pleasurable. Clearly, X has no more reason to be specially concerned about the future painful experiences than one has reason to fear getting a painful ailment that afflicts only one of every thousand people. Perhaps, X has a thousandth the reason to care about the experiences of each of these survivors as about those of a single survivor.\(^7\)

In conflict with the claims of the qualitative view to the contrary, as the number of X's survivors increases, his reason for special concern for the experiences of any one of these survivors seems to become a progressively smaller fraction of the reason for concern for the experiences of a unique survivor. Because, on a qualitative theory, the extent to which one survives as a given being is determined fully by the qualitative relations that hold, on such theories having more survivors cannot attenuate the degree of one's survival.\(^8\) Hence, the only way to resolve this conflict is to suppose that X has no reason to be specially concerned about the

---

\(^7\) I consider it a misleading to think of X as having reason to care specially about a pain that is one thousandth as intense as that which \(Y_{1000}\) suffers. After all, there is no sentient being here who ever experiences such a diminished pain. In Chapter 6, Section 9, I argue it is a mistake to view attenuated survival as survival to have experiences of diminished intensity.

\(^8\) I consider this claim to be essential to the qualitative view. Though the existence of multiple survivors may pertain to legitimacy of identity claims, I know of no proponent of the qualitative view who claims that survival can be attenuated by numerical (as well as qualitative) considerations. When Parfit wonders how a double success can be a failure, he implies clearly that one's survival as one being is not attenuated by one's survival as others. (See Chapter 3, Section 7.)
experiences of any of his survivors in any of these cases. (This follows on mathematical grounds. Two numbers can be the same but the one only a mere fraction of the other only if both numbers are zero.) If X’s relation to the experiences of a large number of survivors \( Y_1, Y_2, \ldots, Y_n \) is so tenuous, then X’s relation to a unique survivor \( Y_x \) or \( Y \) is equally tenuous. Hence, on a theory that implies that one survives as a duplicate, there can be no justification for special concern for future experiences, even in a case of supposed ordinary survival.

Indeed, if we believe that we have reason for special concern about the experiences of our survivors, then it is, I think, absurd to suppose that one could survive as a duplicate. Like Williams, I find it deeply puzzling how one could “project” one’s feeling of special concern towards the experiences at a given future time of more than one supposed survivor. (Williams, 1975) Perhaps, we find it so hard to conceive of survival as a duplicate because we have strong intuitions that one cannot survive as a sentient being without being that same identical sentient being. (I do not wish to argue here that survival and identity must always coincide but only to demonstrate that we are deeply committed to this belief.)

Consider the case of a twenty year old and the seventy year old with the same but greatly aged body. Most of us believe deeply not only that the twenty year old will survive fifty years later to have the experiences associated with and only with the seventy year old body but also that the seventy year old will be a survivor fifty years later of and only of this twenty year old. One is able
to project oneself (and one's feelings of special concern) forward into the experiences of a single survivor and one is able to project oneself (and one's feelings of regret, satisfaction, acceptance etc.) backwards into the experiences of a single sentient being (a precursor) of which one is a survivor. Such reactions only make sense if the survivor and the precursor are the same sentient being.

Thus, to suppose that one can survive as a duplicate conflicts with an unshakeable intuition about our survival. If at a future time there will be supposedly multiple survivors of oneself, then one seems justified in asking "Which one will be me?" Of course, it might be objected that this intuition may be unreliable and that to appeal to this intuition begs the question in favour of the view that there cannot be survival without identity. However, I believe that careful reflection actually strengthens the intuition that survival as a duplicate is impossible.

Hence, if the causal requirement of a qualitative theory does not rule out survival as a duplicate, then it is hard to understand how, on such a theory, there can be reason to care specially about the experiences of one's supposed survivor. Acceptance of a qualitative view, according to which survival consists in duplicable qualitative relations, leads inevitably to skepticism about survival in my sense.
5. Perry on the Importance of Identity

Clearly, in order to defend the qualitative view, a much more serious attempt to explain in terms of qualitative relations how egoistic concern may be justified is needed. In order to defend this view, one needs to specify the relative importance of the various factors, to explain how these factors should matter to the specified extent, and to explain how these factors and their interaction can justify special concern for future experiences. In this section, I consider John Perry's attempt to explain in terms of intentions and their interaction with other qualitative factors why we should care about future experiences. As I argue, though his insightful account does explain why a certain sort of concern is justified, it is not the sort of concern (i.e. egoistic concern) that is at issue.

For Perry, the question that needs to be answered is "what reasons would we have for present actions which would ensure future benefits for ourselves?" (Perry, 1976, p.74) That is, why should one strive to enhance the future experiences of a person who will be identical to oneself. In order to be consistent in my terminology and to do justice to his position, I will need to reconstruct his arguments quite liberally.

---

9 It is important to note that as Perry understands "identical," an earlier experience and a later experience are experiences of the same identical person if they are associated with the same body. Hence, his question might be better rephrased and clarified as follows: What reasons would one have for present actions which would ensure that future experiences to be associated with one's body will be positive?
In order to tackle the above question and to explore which factors pertain to survival, Perry considers the case of a person who by pushing a button today can prevent a horribly painful experience being associated with his body tomorrow. Clearly, Perry is seeking a special reason for pushing the button, a reason unique to the subject himself. He writes: "If I am told that by pushing a button I will prevent someone from being in great pain tomorrow, I will have reason to push it. But intuitively, if the person is me, I will have more reason, or perhaps special reasons for pushing it." (Perry, 1976, p.74) What exactly are these supposed "special reasons" to push the button today to prevent this great pain tomorrow? Actually, he distinguishes (quite rightly I think) two questions: (1) On account of what special reason(s) should one push the button? (2) Why does one feel that one has a special reason to push the button?

Of the two questions, the second is the easier. Clearly, one feels a deep motivation to ensure that future experiences associated with one's body will be positive. As stated in Chapter 1, it is hard to imagine how human beings could survive if they did not feel such motivation. Such future-directed concern is clearly a selective advantage. There doesn't seem to be the same direct advantage to our concern with avoiding future pain. However, it is also clearly a selective advantage to feel such concern. Pain tends to be caused by events that are injurious to the body. Hence, by avoiding pain, one is also avoiding physical injury. Hence, it is easy to understand how the feelings of special concern have evolved
that would specially motivate one to push the button. One notes that though Perry does not accord memory a direct role in explaining why one should act in one’s supposed self-interest, as I have explained in Section 1 of Chapter 1, memory is necessary for one to acquire this indispensable sense of future-directed concern.

So what exactly are the special reasons that Perry seeks why one should push the button? The answer to this first question seems quite obvious. If I do not push the button, then someone will be in pain tomorrow and that someone will be me. Clearly, because this response appeals to an unanalyzed notion of identity of a person over time, he can salvage only a part of this response. That is, my reason for pushing the button is that if I do not, then someone will be in pain tomorrow. My pushing the button will promote the positive quality of this person's experience. However, this can hardly be the special reason that Perry seeks. I would have exactly the same reason for pushing the button for someone with a different body.

Central to Perry's answer to the first question is his concept of a project. He understands one's projects quite generally as one's wishes as to which events are to occur in the future. He distinguishes between one's private and nonprivate projects. A nonprivate project is one's bare wish that a certain event will occur. A private project is one's wish that not only will the certain event occur but that one's own actions will lead to this event.

There are two main reasons to embark on a project. First, one
may enjoy striving to complete the project. Second, the project is a wish for a future state of affairs that one considers positive. It is because of a general fact about human beings that both of these reasons are sound in terms of enhancing the quality of future experiences to be associated with one's body. Usually one's values, character, and desires change only gradually. As Perry notes, we are reliable. We expect to have the same goals and preferences tomorrow. Hence, the person tomorrow with one's body is likely to derive as much enjoyment from working on the project as does the person who embarks on the project. The project itself will enhance one's life as one continues to work on it. One's enjoyment will not change because of the stability of one's projects. Later, when the project has been successfully completed, one is still likely to see the outcome of the project as positive, to like the resulting state of affairs. The memory of starting the project and seeing it fulfilled also adds to one's satisfaction.

An example may help to clarify Perry's idea. I decided to write this thesis because I enjoyed writing and felt it desirable to finish my degree. Now some time later, reliable as I am in my preferences, I still enjoy writing and hence spend my days more enjoyably than I would otherwise spend them. Should I finish this thesis and receive my degree, I will have achieved a state of affairs that I am still likely to find positive. I will remember working on it and feel a sense of accomplishment as a result. In this way, my projects may interact with my memories and my persistent preferences and characteristics to make future
experiences to be associated with my body positive. Completing my thesis seems to be mainly a private project, but I might view it also as of a partly nonprivate nature. It is important to me that people realize that all of the well-known theories of personal identity suffer from certain intractable difficulties. If I were unable to complete this thesis, I would view it as positive (though perhaps less so) if someone else should do so.

Perry suggests in terms of projects another reason for pushing the button. He writes: "If I am not in pain tomorrow, I will contribute to the success of many of my projects..." (Perry, 1976, p.74) Certainly, excruciating pain would tend to impede tomorrow's progress on my thesis. However, unless I presuppose that it is I who will do the writing, I would seem to have exactly the same reason for pushing the button for someone with a different body who is pursuing an identical project.

However, Perry argues that one's projects provide the basis of a reason for special (i.e. extra and/or different, maybe not "special" in my sense) concern about future experiences to be associated with one's own body. First, certain contributions to my projects are ones that only a person with my brain and body could make. One may be in a special position to promote one's projects due to special abilities and circumstances. For instance, it is by virtue of the information and records of experiences stored in my brain that I have formed the views that I have of personal identity and remain motivated to expound these views. No other person has precisely the same information physically stored in his brain.
Second, one may remain in a special position to complete one's present projects by virtue of the persistence of one's body no matter how the associated psychology may change. For instance, the presence of a celebrity at a fund raiser may help to attract donations even if the celebrity's desire to help raise the funds has vanished since committing to an appearance. Third, there is also an advantage in terms of the quality of one's present experiences to embark on a project that one enjoys and that may further something one values.

The final reason why one's projects provide a reason for such concern is more subtle. As Perry writes, "I am derivatively justified in having the private project...because I have the relevant nonprivate project...and beliefs that if the nonprivate project is not contributed to by me, other of my nonprivate projects will fail." (Perry, 1976, p.79) I think that his point is that I have a cluster of projects. For any given one of these projects, even if someone else is equally qualified and motivated to pursue it, there is an advantage to my brain and body continuing to survive and function and thus to be involved in completing the project. If they were not to survive, the other projects in my cluster would not be completed.

However, as Perry seems to realize, these reasons would be just as compelling if I should be destroyed and replaced by a benign imposter who is physically and psychologically indistinguishable from myself. Not only would the imposter have the same intentions as myself but he would also have the distinctive
collage of qualitative features needed to carry out these intentions. The reasons for pushing the button would be the same for the imposter as for a person with my same body. Yet, I may not feel the same motivation for pushing the button for such an imposter. Perry thinks that the only reasons for thinking that a "benign imposter" does not merit the same concern is the belief fostered by experience that one's projects won't get done without one and an innate commitment programmed through evolution to one's own projects. Because we never have to deal with such cases of benign imposters, our habit of thinking of our own bodies as somehow special in this way is very strong indeed. In fact, we have just as good reason to press the button for the imposter.

Equally, Perry thinks that there is no reason to care specially about someone with one's body who is greatly altered psychologically. For instance, Methuselah at age 100 would have no reason for special concern for Methuselah at age 900. According to Perry, the young Methuselah not wishing to give the old and greatly changed Methuselah cancer by smoking now is almost like not wanting another individual to be exposed to one's second-hand smoke. The only reason why one may care in a special way about oneself so greatly altered is habit. Indeed, even if the psychology associated with one's body will be (as sometimes happens) dramatically altered, this habit is just as strong as when there will be no such dramatic changes. Any feeling that I may have that there is an "ineffable me-ness" preserved in the case of normal aging of my brain and body that is lacking in the case of the benign imposter
is only habit. (Perry, 1976, p.84) Indeed, I have a better (or at least a more personal) reason to push the button for the imposter than for the aged and greatly changed individual whose body will be numerically the same as my present body.

Hence, Perry thinks that the importance of identity (i.e. the numerical identity of one's body over time) is derivative. Such identity is only important for the relevant qualitative relations that it normally provides. Of course, my relation to the imposter would be different from the relation to the future person with my body, but this former relation would be of the "same species" as the latter relation. Both relations would involve what matters, but they would involve different numerical relations of bodies. Though my relation to the imposter would not be the same, I would have the same reason to act in his behalf as in behalf of the future experiences to be associated with my same body. In contrast, the relationship of the younger to the older Methuselah would involve the same body but not the relation that matters.

Perry thinks that having the same brain is the best guarantee that the qualitative relations that matter in this way will hold. Certainly, the continued existence of someone with my brain provides the best chance of my intentions being carried out. Having the same human body (apart from the brain) may also pertain because it will promote retention of any special physical abilities required to complete one's projects and enable others to identify one as playing a certain role (as was the case with the celebrity above). However, Perry thinks that such physical considerations are
of less importance (at least in the case of humans) for one's projects than are psychological considerations.

I should resolve, as this point, what I think to be some potential for terminological confusion regarding Perry's account. Perry states that "...what is of importance in survival [presumably the continued capability to carry through with projects etc.] need not occur in every case of survival..." (Perry, 1976, p.86) He doesn't make it clear just how he understands survival, but perhaps he means the survival of one's body and its continued support of experience. In contrast, Parfit treats qualitative relations as constituting survival. Survival is the obtaining of relations that matter in a special way. A case in which these relations are not present is not a case of survival. I believe that any dispute here is largely terminological. Because the question for me is in what survival (as something that matters to us) consists, I seek to evaluate the case that the qualitative relations that Perry considers important do constitute survival. There are a number of obvious objections to the resulting theory.

First, one's desire to avoid pain does not have the utility in furthering one's projects that Perry supposes. If one is told that one will suffer great pain tomorrow, one fears the pain rather than that the progress on one projects will temporarily suffer. Moreover, unless the relevant project is one of survival or avoidance of still worse suffering, one is likely to be much more concerned about avoiding the brief episode of horrible pain than about completing the project as a whole. One is likely to regard a
brief delay in progress as even less significant. Also, if I were
told that I could finish my thesis at the cost of torture tomorrow,
I would not accept the offer. Yet the pain that I would experience
for only one day would have an insignificant effect on the
completion of my thesis project. Hence, it is difficult to explain
in terms of one's projects why one has a special reason to push the
button to avoid the future pain.

Second, the above theory puts an unrealistic emphasis on non-
selfish motives. Apart from the immediate satisfaction that one
derives from committing to a project and believing that one's
actions will lead to changes to the world that one now sees as
desirable, the only special reasons that the theory allows for
acting in certain ways is that they further the quality of future
experiences that are in no deeper sense even one's own and that
they may improve (as gauged in terms of one's present values) the
future lives of others. Even the seemingly most selfish of acts
would be of no real benefit to their instigator. The whole concept
of a selfish action would seem to be incoherent. Clearly, what we
mean by what matters in this special way is not the furthering of
any such unselfish goals. There may be a reason for such concern
about one's goals, but this concern is of an entirely different
nature than the special concern that we are attempting to
understand. We wouldn't feel the same concern for the future
experiences of someone else, for instance an imposter. As Eddy
Zemach admits, it wouldn't please him to think of his present
actions as enriching the life of an imposter who takes over his job
and moves in with his wife. It would make him mad. (Zemach, 1987)

Third, we should of course be suspicious of the idea that our special concern is to enrich anyone's experiences. Notice that we struggle to survive even if we will be miserable. Quality of anyone's experiences, even ones to be associated with our own body, is not just what we want. We want to survive even if, as is usually the case, our lives are awful. We do not just feel specially concerned about what certain experiences will be like. We also feel specially concerned about whether or not there will be certain experiences, those that we may survive to have. We seek a theory that provides a reason for both of these aspects of this special concern. Clearly, Perry's theory is unable to provide such a reason.

I believe that Perry's approach is the best attempt to explain in qualitative terms how concern for future experiences to be associated with one's body (i.e. ones about which most of us find it reasonable to suppose that one should care specially) can be justified (and perhaps the logical conclusion of the search for qualitative relations that justify special concern). Given that this attempt fails to justify the sense of concern in question (i.e. egoistic concern), I think the prospects for a qualitative theory doing so are very poor. Any attempt to analyze survival in terms of qualitative relations so as to explain how egoistic can be justified is, I think, doomed to failure.
6. Similarity and Convention

Though Perry’s theory is unable to explain how qualitative relations can justify egoistic concern, it does explain how the holding of qualitative relations can matter in other ways. Not only has Perry explained how, with one’s projects, one is qualified to enhance the future experiences associated with one’s same body but it is clear how one may be uniquely qualified to enhance (or at least to greatly affect) the future experiences associated with other bodies as well. One’s projects (such as providing for one’s dependents, being a supportive companion to one’s family and friends, following through on one’s commitments etc.) may be important to other people. With one’s projects, one’s memories, and one’s psychological characteristics, one may fill a unique economic, social, and human niche. Because of the stability of other people’s needs, it may matter greatly to others that this niche continue to be filled. As long as there is someone with one’s projects and the memories and traits that promote their fulfilment, this niche will be filled.

One’s continuing to have certain physical characteristics may also enhance one’s ability to fill such a niche. Clearly, one’s continuing ability to perform certain physical tasks is important. Physical appearance in itself may also have some importance. For

---

10 Hence, though it may be correct that (as Butler’s principle asserts) one’s survival cannot be analyzed so as to explain (or make clear) how egoistic concern may be justified, it is false that an analysis of survival cannot explain how one’s survival can matter in other ways. For some understandings of what matters in survival, the regress that Butler feared may end.
instance, just the presence at the fund raiser of someone who looks just like the celebrity may help to secure donations. The more physical and psychological similarity that one may have to the person who has filled such a niche, the better able one is to continue to fill this niche.

Now suppose that egoistic concern is never justified and hence that no one ever survives in my sense. (As I have conceded, there is no guarantee that such is not the case.) Then, as argued in Chapter 1, whether or not one survives would be a matter of convention. Clearly, it would be a very reasonable convention to identify one who is qualified qualitatively to fill the niche of the original as a survivor of the original. (When there is more than one candidate in these terms, we may decide which is the same as the original on the basis of a closest continuer schema.) Thus, whether or not certain qualitative relations hold may be a suitable basis for convention regarding survival.

A person who is teletransported back to Earth following a stay on Mars will have all the qualitative traits (psychological and physical) to fill the same niche that the traveller filled before teletransporting to Mars. Indeed, unless there have been great changes during the interim, the returning traveller would be just as well qualified to fill this niche as the person with the original body. Clearly, the traveller's family will be delighted about the successful return and the prospect of their lives returning to normal. For them, following the teletransportation home, life will go on as normal. In a world in which egoistic
concern is never justified, people would still employ the word "survives." Because of the perfect resemblance of the result of teletransportation to the original, it would be a reasonable convention to suppose that one survives teletransportation. It makes sense to deny that one survives teletransportation only if one survives in my sense normal aging but not teletransportation, that is only if the traveller has reason to care specially about the experiences to be associated with his own body but not about those associated with the replica on Mars. Thus, as noted in Chapter 1, if we do not think that egoistic concern is ever justified, then it is perplexing why anyone would find physicalism (i.e. a view on which physical continuity is necessary for survival) at all attractive. (I revisit this difficulty for Physicalism in Chapter 9.)

7. Conclusion

Recall that by adding a strong causal requirement, a qualitative theory can solve the duplication problem (and perhaps the problem of false memory). Perhaps, if the causal requirement is strong enough, then there may be reason to care specially about the experiences of one's supposed survivor. As I have argued in the previous chapter, to solve the duplication problem, we would have to require that the psychological state of the supposed survivor result from the psychological state of the original in the normal way or at least that there be considerable physical continuity (i.e. of tissue, cells, matter) between the bodies that support
these psychological states. Perhaps an analysis tightened in this way may be sufficiently robust to account for justification of egoistic concern and hence to provide sufficient conditions for survival. I suspect that Parfit's clinging to a causal requirement is due to an appreciation that a qualitative theory without such a requirement cannot justify egoistic concern. However, he fails to recognize that the very minimal causal requirement that he accepts cannot work.

How exactly would a qualitative theory that requires what is basically the normal causal basis for the psychology of the survivor differ from a physicalist theory (perhaps with additional qualitative requirements)? Indeed, the resulting theory may be more fairly described as physicalist than as qualitative. Thus, the duplication problem and the difficulty in explaining in qualitative terms how egoistic concern can be justified may incline one to physicalism.
PART II

PHYSICALIST THEORIES
Chapter 5
Are We Organisms?

1. Introduction

Recall that there are two sorts of physicalist theories: physical and biological reductionist theories. In this chapter, I consider the physicalist view of ourselves as biological entities (the biological reductionist view). In general, such a view has considerable initial plausibility. Unless we have strong religious beliefs to the contrary, we tend to think that our existence lasts just as long as our biological life continues. I focus on Peter Van Inwagen’s organicist theory, arguably one of the most systematic, and initially plausible versions of physicalism. According to Van Inwagen, we are organisms. (Van Inwagen, 1990) Hence, we may suppose that one (a person or animal etc.) survives so long as one’s biological life continues. Roughly speaking, for Van Inwagen, a life is a continuous process in which an organism maintains a homeostatic inner environment thus enabling this process to continue. I argue that, despite its initial plausibility, his theory has serious difficulties. I believe that any biological reductionist view faces similar difficulties.

1 Another well-known theory of this sort is Mark Johnston’s. Johnston thinks that we should hold the “specific conception of ourselves as evolved animals of a particular sort, the conception that locates us most easily within the naturalistic framework taken for granted by scientifically validated common sense.” (Johnston, 1987, p.75) According to Johnston, each of us is a locus of a biological kind, a specific human being. Unfortunately, he offers only a sketch of a treatment of the nature of ourselves and our identity. Eric T. Olson also views us as human animals. (Olson, 1997)
2. Van Inwagen’s Ontology

According to Van Inwagen, what there is in the world are physical simples and organisms. (By physical simples, he means the elementary particles of physics.) Organisms are the only entities that have parts. He offers this ontology as his answer to the Special Composition Question. That is, in what ways can the fundamental building blocks of matter (the above physical simples) be combined to yield new things that exist over and above the building blocks themselves? His “Moderate” answer to the Special Composition Question is that there is something that is composed of multiple simples if and only if the activity of these simples constitute a life. (Van Inwagen, 1990, p.115) His ontological position conflicts with the commonsense view that there are also ordinary objects (natural objects and artifacts). What exactly is his case against ordinary objects?

One may think that there are plenty of objects composed thereof of whose existence we must take account. For example, a rock (a natural object) is something on which one might stub one’s toe. A table (an artifact) is something which might support another physical object. However, Van Inwagen denies that things can combine (by “Contact, Fastening, Cohesion, Fusion” etc.) to generate new things. (p.122) According to Van Inwagen, what we take to be physical objects are just collections of cooperatively-arranged simples. That is, what we take “to be the product of the activity of a single object...[is] the product of the joint activity of many.” (p.118) For example, a rock is just a bunch of
simples cooperatively arranged (i.e. rock-wise) in a highly cohesive way that resists intrusion from other simples (and arrangements thereof). A table is just a bunch of simples that are cooperatively arranged (i.e. table-wise) in such a way that, by virtue of their interactions, they are able to support other objects. It is as a mere convenience that we employ and cling to our concepts of rocks and tables. He thinks that when the simples are arranged to form an ordinary physical object, there is nothing that we need to think of as existing over and above the simples as arranged. Artifacts and natural objects are mere "virtual objects."  

In a way, Van Inwagen denies that virtual objects exist. Indeed, he denies that virtual objects have properties. Though arguably it is rather artificial to deny out-and-out that ordinary objects exist, he seems correct that there is no need to include them in our basic ontology. By treating them as nothing but arrangements of simples, he has located them in his atomistic picture of the world. For convenience, I will continue to speak of "virtual objects" as (physical or ordinary) objects. In such claims, I think that he is absolutely correct. I have always found it strange that anyone would be interested in the individuation criteria and identity conditions of ordinary physical objects. Such criteria and conditions strike me as entirely pragmatic and

---

2 Because Van Inwagen’s case against ordinary objects does not pertain to my thesis, I do not consider this case in any more detail. He explains his reasons for rejecting artifacts in Section 13.
3. Survival as a Thinking Organism

In itself, Van Inwagen's understanding of survival (i.e. persistence of an organism) does not seem to accord well with survival as a sentient being. If a conscious being (for Van Inwagen, a conscious organism) were to lose its ability to support conscious states but still maintain its life processes, then this being would survive as an organism but not as a sentient being. Following the loss of this ability, there would be no experiences about which this being should be specially concerned simply because, following this loss, this being would have no experiences. Also, for Van Inwagen, there is no such thing as a sentient being. Rather sentience is an attribute of an organism. (He understands "think" sufficiently liberally so that having any experience (such as a pain) counts as thinking. A thought is just an experience. Thus a thinking organism is the same as a sentient organism.) How can it make sense to think of sentient beings as surviving if sentient beings do not exist?

However, it may be possible to find a place in Van Inwagen's theory for a sense of "persistence of thinking organisms." (p.182) A thinking organism is just an organism that is capable of thinking (i.e. has the property of sentience). A thinking organism X at an earlier time persists as a thinking organism Y at a later time if X and Y are the same organism and Y has conscious states. Perhaps one's survival as a sentient being is just one's persistence as a
thinking organism. That is, an organism's feelings of egoistic concern are justified for and only for future experiences of the same persisting thinking organism.

This understanding may seem to violate the spirit of Van Inwagen's theory. Though he does not argue that relative identity is incoherent, he presupposes an "absolute" conception of identity. (p.183) However, this understanding does not imply that there is a distinction between our persistence "qua organisms" and our persistence "qua thinking beings." (p.183) To say that one persists as a thinking organism is just to say that the organism that one is persists and continues to have conscious states. That is, persistence of a thinking organism may be nothing but the persistence of an organism and its continuing to have the property of sentience. If we can analyze persistence of thinking organisms in this way, then we can avoid talk of persistence of thinking organisms and equally of survival of sentient beings.

4. Van Inwagen's Theory of Organisms

Of course, if there are no ordinary objects, then a physicalist may think that all there is are physical simples. One may think that organisms are just simples arranged organically. However, Van Inwagen insists that there are also organisms. According to Van Inwagen, organisms have four main characteristics that distinguish them from ordinary objects. Arguably, these differences are sufficiently significant to provide some reason to treat in metaphysical terms organisms very differently from
ordinary objects. One should note that, as with ordinary physical objects, he thinks that parts (such as an arm, the brain, a kidney etc.) of an unmutilated organism are mere "virtual objects" to which no properties can be ascribed simply because such things do not really exist.

First, organisms are much more complex than ordinary objects. Even single-celled organisms have huge numbers of specialized molecules, specialized structures and specialized processes, the relationships between which are almost indescribably complex. The animals and plants of the macroscopic world are hugely more complex still.

Second, organisms exhibit much more uniqueness and distinctiveness than do physical objects. For instance, there may be numerous artifacts that are virtually indistinguishable and indeed interchangeable. We think nothing of exchanging a telephone for one from the same manufacturer and of the same model. Though naturally occurring physical objects even of the same sort may appear quite different on the surface, they may have indistinguishable composition and internal structure. In contrast, every organism is quite unique. Even identical twins raised in the same environment will have characteristics that distinguish them.

Third, unlike ordinary objects, organisms are highly control centred. Numerous coordinated feedback mechanisms are needed to maintain the homeostasis of even the simplest organisms. Animals with well-developed nervous and endocrine systems are the most control centred entities in the world. Arguably, the control
mechanisms of objects with the most sophisticated man-made processors are rudimentary by comparison.

Fourth, unlike with physical objects, the relationship of an organism to its environment is one of jealousy. An organism seeks to further (by consumption, predation, etc.) its welfare at the expense of the environment and the other organisms therein. An organism maintains the inner environment needed for its continued life at the expense of its outer environment. To use its resources to reproduce and preserve its genome is the most jealous act of all. The complexity and control-centeredness of organisms help to secure the organism's jealously-sought outcomes.

Certainly, organisms have some very distinctive features and characteristics. However, these characteristics in themselves do not distinguish organisms from physical objects sufficiently that we need to include organisms in our ontology. Though, perhaps, the special characteristics of organisms are some reason (perhaps a prima facie case) to suppose that the proposed answer to the Special Composition Question is not absurd, Van Inwagen acknowledges that he still needs to provide a case that this answer is correct. (p.115)

5. The Consciousness Argument

Why should we think that there are organisms as Van Inwagen understands them? His main argument that organisms exist is that there is a good case that I am an organism. (Van Inwagen, 1990, Section 12) This argument is as follows: I cannot doubt that I
exist. I am not immaterial. So I am physical, and hence I am composed of physical simples. Obviously, I am not a single physical simple. Hence, I am composed of multiple simples. So what could I be? The most reasonable explanation of how I can be a composite being is that the simples that compose me are involved in a life. That is, I am an organism. Therefore, at least one organism exists. If I am an organism and I exist, then reasonably other organisms exist as well.

I think that most of us, like Van Inwagen, would find the first premise (i.e. that I exist) to be very compelling. Van Inwagen basically accepts Descartes' well-known arguments that I exist and rejects the empiricist challenges to the contrary. (p.116) However, he suggests a particular challenge that a materialist who doubts that I exist may give to the view that having experiences demonstrates my existence.

Such a materialist may argue as follows. Suppose that the simples (perhaps those of my body or brain or a subset thereof) that I may think compose me don't compose anything. However, even if such is the case and hence I do not exist, "[These] simples...stand to each other in just the causal relations they would stand in if they composed...[me]: The same electrical currents flow, the same chemical reactions take place, and so on." (p.117) If, as a token-token materialist believes, any particular thought can be identified with some particular physical neural process, then, even if the simples involved compose nothing, so long as all these processes go on, there will be these thoughts
nonetheless. It seems that there can be an experience without this being anyone's experience, that is without anyone existing to have this experience. Perhaps, then, it is quite reasonable to doubt that an experience that I take to be my own is really my own. Perhaps it's no one's experience. Why then suppose that I exist?

The point here is that thinking may be just as much a cooperative activity as are the supposed activities of ordinary objects. Consider the case of a table supporting an object. The supporting of an object gets done even though there is no table that does the supporting. This activity gets done by cooperation of table-wise arranged simples. Van Inwagen writes: "Our initial impression is that there is a certain middle-sized object, the...[table], that does a thing called supporting weight. Later, under the influence of our theory, we decide that what we took to be an accomplishment of a single object was the accomplishment of many." (p.118) Given that the supporting of weight (and any other supposed activities of the table) are the product of the joint activity of numerous simples, there is no reason to suppose that the table exists.

Couldn't we say something similar about my supposed thinking? Van Inwagen writes: "Our initial impression is that there is a certain object, I, that does a thing called thinking; later, under the influence of our theory, we decide that what we took to be the activity of a single object [I] was the activity of many?" (p.118) That is, my supposedly having experience is really just a cooperative activity of certain simples. Perhaps just as a supposed
ordinary object may be a collection of cooperatively arranged simples, I am a collection of simples arranged sentiently, that is arranged in such a way as to be able to produce experience. (He talks about simples arranged "intellectually." However, given his liberal understanding of "thinking," I think it less confusing to speak of the simples as arranged "sentiently.") Just as there is no need to suppose that there is a table that supports the weight, there is no need to suppose that there is a thinker (i.e. the referent of "I") that has what I take to be my thought. If such is the case, then (though the simples that compose me do exist), I do not exist.

However, Van Inwagen thinks that there is a need to suppose that I exist, to suppose that there is an "individual thing that thinks." He considers it to be highly implausible "to regard thinking as a mere cooperative activity." (p.118) He thinks there is a disanalogy between supporting objects and thinking. He writes:

Things that work together to support weight are not forced, by the very nature of the task set them, to support weight by composing a single object that holds things aloft. But things cannot work together to think - or, at least, things can work together to think only in the sense that they can compose, in the strict mereological understanding of the word, an object that thinks. (p.118)

For Van Inwagen, "...pain cannot be...[an activity] that a lot of simples can perform collectively, as simples can collectively...support a weight." (p.118) There is an individual thing that thinks (i.e. has experiences). My mental states are not just the cooperative activities of simples.

Van Inwagen concedes that this is not a decisive "knock-down"
response to this challenge and that this response may seem to some to beg the question. I doubt that any argument can convince someone that having an experience demonstrates one's existence who does not see this as self-evident. Indeed, I suspect that to concede the need to supply arguments for this belief is to capitulate to the skeptic. However, Van Inwagen's response is more decisive than he seems to recognize. It is by virtue of their microproperties (electromagnetic, nuclear, gravitational) that the simples arranged table-wise are able to interact and cooperate to support objects. However, it is hard to believe that a thought is decomposable into properties of a multitude of distinct physical simples. The simples have no identifiable microproperties by virtue of which they are able to interact cooperatively to produce experience. How, then, can thinking be a cooperative activity of physical simples? It seems, then, that a thought is a property of (or perhaps is an activity or perhaps somehow belongs to) some entity, or is an entity in itself. Either way, there must be an entity involved over and above the physical simples themselves. It seems that there is good reason to suppose that I exist.

Now, given that I exist, what are the alternatives about what I am? One alternative is that I am immaterial. Van Inwagen

---

3 Though I refer to the whole argument that organisms exist exposit in this section as the "Consciousness Argument," the case that thinking cannot be a cooperative activity is clearly central.

4 I am not claiming here that it is indubitable, only that there is a strong case, that I exist. However, I do not believe that one can intelligibly doubt one's existence.
explicitly rejects the view that I am immaterial. I think that he is correct that one cannot "deduce...[one's] immateriality from the fact that...[one is] the sort of being whose existence...[can] be proved by 'Cartesian' methods." (p.119) As he presents it, the view of myself as immaterial seems to encompass two distinct views: (1) I am something immaterial over and above the experience, something that has the experience in question, and (2) I am "composed of ideas," that is composed of (perhaps a succession of) related experiences. On this view, I have the experience in question by containing it. (Even if I am composed only of the experience, I can still own the experience in this way.) He clearly dismisses both of these views. He writes: "...I do not understand what it means to say that I am composed of ideas. Whatever things compose me, they are all of them material..." (p.117) His view that we cannot be "composed of ideas" seems to derive in part from a rationalist conviction that talk of conscious states is unintelligible unless there is an entity to which these states can be ascribed. For Van Inwagen, a conscious state is an attribute, and there can be no attribute without an entity.5

5 See Van Inwagen, 1990, pp.206-207, for a discussion of the virtues of rationalism and the vices of empiricist views of the mind. Though I do not intend to develop the distinction between empiricism and rationalism, I should mention how he tentatively characterizes the views in question. On a "typically empiricist" view of the mind, "...the extended event or process that is the persistence or continued existence of a thinking being is composed entirely of mental events." (p.206) On a "typically rationalist" view of thinkers, "...a sequence of thoughts...[are] thoughts of the same thinker just in virtue of their being modifications of one, continuing substance: the understander that stands under them." (p.207)
So I am material, and hence I am composed of physical simples. Clearly, it is absurd to suppose that I am a single physical simple. Hence, I must be composed of multiple physical simples. That is, I am material and I have parts. So there is at least one composite material being. We know that it is possible that there are composite material beings because it is actual. For Van Inwagen, we need to answer the following question. "When, in general, does this happen [i.e. that physical simples compose a material being]? What allows a material being to have parts?" (pp.119-120) So what is the best hypothesis about what I am? This hypothesis must explain how I can be composite.

One may suppose that the simples that compose me are able to do so because they compose something that thinks. That is, I am a thinking object, a composite object that has (essentially) the property of thinking. However, Van Inwagen denies that the fact that I am a thinking being explains how I can be a composite being. As he writes: "...the fact that I am a thinking being shows that there is at least one composite material object. But it does not explain how it is possible for there to be a composite material object." (p.120) In particular, this fact does not explain "the fact that certain simples are capable of composing me." (p.120)

Van Inwagen is certainly correct that the fact that a state of affairs is actual (in this case, that there is a thinking being) does not explain how it is possible. To use his example, the absence of money from a vault proves it's possible to get inside the vault, but it doesn't explain how it is possible to do so.
Arguably, he is also correct that what explains how something is possible can be present even in the absence of what is to be explained. For example, a security guard could get into the vault by the same method (i.e. what explains how money could be removed) as the burglar but take no money.

Likewise, what accounts for the unity of the simples that compose me may be present in the absence of thought. I might cease to think but continue to exist. As it is, my existence as a composite being might be explained in the same way as for an object incapable of thought. There may even be composite objects that are "essentially incapable of thought." According to Van Inwagen, the fact that without thought my existence is of no value to me does not show that thinking is metaphysically essential to me, that I am a thinking object. Though he accepts that the capacity for thought may be my most important possession, he denies "that thinking has anything to do with my existence." (p.120)

However, even if "thinking is not my essence," there must be something that "binds the simples that compose me into a single being..." (p.121) There must be some "ground of my unity..." (p.121) According to Van Inwagen, "...I exist because the activity of certain simples constitutes a life..." (p.121) That is, "what binds them together is that their activities constitute...a homeodynamic storm of simples, a self-maintaining, well-individuated, jealous event." (p.121) There must be something that is the bearer of my mental states, and according to Van Inwagen this bearer is an organism. It seems that I am an organism and
hence that at least one organism exists.

If I exist because the activity of my simples constitutes a life, then it would be implausible and indeed arbitrary to deny that other organisms also exist. That is, if the activity of a collection of simples compose a life, then they compose an object. For simplicity, I avoid the issue of how to handle things such as viruses that exhibit some but not all of the usual characteristics of living things. I also do not wish to consider much further Van Inwagen's case for the converse claim that a collection of simples compose an object only if their activity constitutes a life (that there are no ordinary objects).

However, I think it worth mentioning one more reason why Van Inwagen denies the existence of ordinary objects. He writes:

> Cartesian arguments show that we are forced to grant existence to some organisms (to the ones that think, at least), owing to the fact that thinking cannot be understood as a disguised cooperative activity. But all the activities apparently carried out by...artifacts and natural bodies can be understood as disguised cooperative activities. And, therefore, we are not forced to grant existence to any artifacts and natural bodies. (p.122)

We have no such reason to suppose that there are artifacts and natural bodies. Hence, it is reasonable to deny that they exist.

There are two features, here revealed, of Van Inwagen's account that I find very attractive. First, his account is not anthropocentric. We are organisms just as are bacteria, ants, and other higher animals, and we are thinking (sentient) organisms just as are whales, dogs, bats, and octopi. Though we are very different from all of these organisms and more complex than some of them, our
persistence as an organism is no different in kind from the persistence of these other organisms, and our persistence as a thinking organism is no different in kind from the persistence as a thinking organism of these other thinking organisms. Perhaps, then, the notion of persistence as a thinking organism (in Van Inwagen's sense) may accord well with my notion of survival as a sentient being.

Second, Van Inwagen has arrived at the conclusion that organisms should be the bearers of conscious states just on the basis of metaphysical considerations. Hence, he can deny allowing the epistemology to drive the metaphysics. He explicitly repudiates verificationism as a methodology for personal identity and survival. He thinks, as do I, that it is a mistake not to distinguish between that which provides decisive evidence in questions of personal identity and that in which personal identity consists. (pp.186-187) Nonetheless, because organisms in his sense are quite readily reidentifiable, his theory does seem to provide a neat solution to the epistemological problem.

6. Brain Transplants

Van Inwagen's handling of brain transplants (and other scenarios that are presently technically infeasible but perhaps not deeply impossible) provides important insights into his account. He explicitly claims that, when one's brain is excised from one's body and transplanted into a brainless body, one is the same organism as the resulting organism. Indeed, he argues that it is
more accurate to think of a brain transplant as a brain-complement (i.e. a whole organism minus its brain) transplant. (p.173)

According to Van Inwagen, when as the first step of this procedure, one's brain-complement is cut away, one survives as a thinking organism. He argues that a thinking organism can thus become "a naked brain," a "radically maimed" human being. (p.172) (Of course, for Van Inwagen, before the brain is excised, it is a mere virtual object.) Then one's brain is fully connected with a different brain-complement. Once the connections have been fully restored, the normal flow between brain and brain-complement resumes. The flow from the brain-complement to the brain is the flow of oxygen and nutrients. In contrast, "The flow of human life from a naked brain ... into a severed brain-complement is largely a flow of electrically and biochemically encoded information." (p.174)

By means of a "political analogy," Van Inwagen tries to convince us that the flow of life is from the brain out rather than from the brain-complement in. (pp.174-175,178) He also suggests that there is a relevant asymmetry in the relation between the brain and the brain-complement that supports this conclusion.6 (pp.177-178) If the brain is provided with a suitable environment, it will maintain itself. Presumably, a high-tech system of pumps could serve to do just that. However, the brain complement will

6 Actually, Van Inwagen finds such an asymmetry in the relationship between the head and the head-complement, but clearly he also believes that there is such an asymmetry in the relationship between the brain and the brain-complement.
need to be fed "a constant supply of instructions" without which "it will not be able to coordinate its activities." (p.178) A life-support system for the brain-complement would have to be much more complicated, probably involving a very sophisticated computer. As he admits, "The contention that this difference does make a difference is a basic premise of my metaphysic." (p.178) Clearly, he realizes that he can give no compelling argument for this contention, only "paint a picture" to make it seem reasonably plausible. (p.178)

I think that most of us would find it very plausible that, if one's brain were transplanted into a different body, "one would go where one's brain goes." (Johnston, 1987, p.78) In a famous example due to Shoemaker, the brain from Brown's dying body is transplanted into Robinson's brainless body, the resulting individual being called "Brownson." (Shoemaker, 1959) One's untutored, and I think quite reasonable, reaction to this case is that Brown survives as Brownson, but of course not as Brownless (Brown's now brainless body), that is not a person or even a mere sentient being, in exactly the same way and to exactly the same extent as one survives when one's intact brain remains housed by the original body. It is certainly natural to assume that one survives multiple organ transplants provided that one's brain remains intact. A reasonable extension of this assumption is that one would survive a total body transplant (besides the brain), that is, the transplantation of one's brain into a different body. It seems strange to deny that the resulting person would be the same as the person who previously
had the brain in question along with the original body.

Given that Van Inwagen's theory accords with our commonsense reactions to brain transplants, given the two advantages identified above, and given our tendency to think of ourselves as (perhaps special kinds of) living organisms and as surviving so long as our biological lives continue, his view may seem, at least initially, very attractive. If there were to turn out to be no serious difficulties with this view, then it would be very attractive indeed.

7. Commissurotomy

As Van Inwagen clearly recognizes, commissurotomy presents a serious problem for his ontology (for his answer to the "Special Composition Question"). He writes:

Observations of some patients who have undergone commissurotomy have suggested to some observers that present in each human life there are two 'consciousnesses' or 'centers of mental life' or 'psychic arenas' or 'parallel streams of consciousness,' one of them associated with one cerebral hemisphere and the other with the other hemisphere. And this, one might argue, could only mean that two thinkers or two subjects of experience are somehow present in the normal human brain. But if two thinkers are present in each brain, then two thinkers are present in each

---

7 Johnston's own reaction to this case is equivocal. He recognizes that there is "a sense in which Brown's mental life may seem to continue on in Brownson." (Johnston, 1987, p.76) He accepts that one would survive transplantation of one's brain in some sense in which, for instance, one would not survive teletransportation. However, he rejects the view that one is one's brain. As he writes, "The characteristic extent and form of a human being is that of unmutilated human being..." (Johnston, 1987, p.79) The brain is merely the organ of primary importance. Though in some sense Brown might survive in Brownson and Brown's body may be the same as that of Brownless, Brown is not the same human being as either of these.
His alternatives for dealing with this conclusion are quite limited. Clearly, thinkers are neither physical simples nor single cells. Nor can thinkers be virtual objects, because such things do not really exist and hence cannot bear mental states. As stated above, he also denies that thinkers are immaterial. If, as seems obvious, a single human being does not encompass two multicellular organisms, it seems to follow that thinkers are not always organisms. Thus, "there are material objects that are neither simples nor organisms" and his answer to the Special Composition Question must be false. (p.190) He dismisses as a desperate expedient the view that only one hemisphere "is a seat of consciousness" and the other "a mere computational instrument." (p.190) His only alternative is to accept that though there may be two separate streams of consciousness, both of these belong to only one thinking being.

To make this strange response seem plausible, Van Inwagen imagines an amoeba-like organism called Cerberus with two brains $B_1$ and $B_2$, each of which is the seat of all the functions of the human brain (reason, judgement, sensory processing, motor control etc.) except that "neither contributes any direction to the homeodynamic processes [respiration, heartbeat, digestion, endocrine activity etc.] that hold the organism together." (p.191) "All such 'maintenance' - in so far as it is localized - is under the control
of a quite separate organ M." (p.191) He also supposes that \( B_1 \) and \( B_2 \) have comparable capabilities, that neither serves as a mere computational instrument for the other. He argues that even if (as some may say) the events occurring in \( B_1 \) and \( B_2 \) constitute separate consciousnesses, there is only one conscious being within Cerberus, Cerberus himself, that thinks the thoughts associated with both brains. (p.194)

This position may seem quite reasonable if the activities of \( B_1 \) and \( B_2 \) are synchronized so that they support thoughts of the same content. Instead of picturing both brains as thinking (for instance) that it is raining, why not suppose that Cerberus thinks that it is raining? (p.195) However, clearly, Van Inwagen also thinks that the same would be true even if \( B_1 \) and \( B_2 \) support contradictory thoughts, if for instance \( B_1 \) supports the thought that it is raining and \( B_2 \) supports the thought that it is not raining. (pp.196-197) According to Van Inwagen, there is still only one thinking being involved. After all, it should be easy for an organism with two brains to think contradictory thoughts. But surely someone thinking these thoughts would be aware that he is thinking both thoughts and would be unable to escape the obvious contradiction. However, Van Inwagen argues that Cerberus could be aware that he is thinking \( p \) and aware that he is thinking \( q \), but

---

8 It is important to remember that Van Inwagen thus uses "brain" in two quite different ways: as the sort of central nervous system that we do in fact have and (here and in the case of Neocerberus to follow) as an organ with this more limited set of responsibilities.
not aware that he is thinking p and thinking q. Perhaps, with p and q tokened in different brains this situation is not so farfetched. (p.201)

8. The Neo-Cerberus Thought Experiment

As Van Inwagen realizes, the case of Neocerberus (an organism different from Cerberus in one important respect) may seem to present a more serious difficulty for his theory. Suppose that Neocerberus, like Cerberus, has two brains \( B_1 \) and \( B_2 \) but also has two organs of maintenance \( M_1 \) and \( M_2 \), with \( B_1 \) and \( M_1 \) on one side and \( B_2 \) and \( M_2 \) on the other. Clearly, after we surgically divide Neocerberus right down the middle, there are present two organisms: say Alpha (with \( B_1 \) and \( M_1 \)) and Beta (with \( B_2 \) and \( M_2 \)). It is tempting to suppose (especially if the brains support consciousness continuously throughout the surgery), that there are two thinking beings all along and hence that his answer to the Special Composition Question must be wrong. (pp.203-204) However, he denies that there would be any continuous consciousness present throughout. According to Van Inwagen with his rationalist leanings, "To find out whether a certain situation contains a 'continuous consciousness'... we have first to find out whether the situation contains a continuously existent thinker," not the other way round. (p.206) In the case of the division of Neocerberus, there is no entity available that can be such a thinker. Thus, the surgery terminates the life of Neocerberus and brings Alpha and Beta into existence. He denies that there is any sense in which a being like
Neocerberus survives partition. He writes: "When I say that someone has survived a certain adventure, what I say entails that this person existed both before and after the adventure." (p.212) Here, there is no entity that exists both before and after partition.

Van Inwagen's treatment of a number of variants of this example may help to clarify his position. First, suppose that we remove from Neocerberus the organs of maintenance that go to Beta but do nothing else. He insists that because this procedure destroys one of the organs of maintenance, it ends Neocerberus' life. According to Van Inwagen, "The remaining organism would be Alpha, with his thoughts tokened in two brains instead of one." (p.208) Second, suppose that we destroy both brains but leave the organs of maintenance alone. According to Van Inwagen, this procedure would not destroy Neocerberus. However, he would no longer be capable of thought or sensation, and he would hence not survive as a thinking being. Third, suppose that we remove one of the brains and maintain it with a life-support system (perhaps based on the system of pumps described above). According to Van Inwagen, the result would be no one. "There'd be a lot of cells that did not compose anything (and hence did not compose anyone)." (p.209) He writes: "...no mass of cells, however arranged and connected, can be a locus of conscious experience unless those cells compose something." (p.210) There would not even be thoughts because there would be no one (and indeed no entity) there to have thoughts.
9. Difficulties for Van Inwagen’s Account

Van Inwagen’s treatment of Cerberus and Neocerberus brings into focus the difficulties of his view and especially his organicist ontology. Consider an organism that is just like Cerberus except that it has only one brain (understood as the organ of thought only). Alpha or Beta would be such an organism. In the case of Alpha, he would have to say that if we destroy $B_1$, then this organism survives but is no longer capable of thought. That is, it does not survive as a thinking organism. If instead we cut away the $B_1$-complement and maintain $B_1$ with a life support system, then the original organism also does not survive as a thinking organism (and indeed does not survive at all). With the centres of control of maintenance gone, there is no entity left to have thoughts. This description is implausible in itself and has implausible consequences.

How exactly does this procedure differ from the procedure in which a normal brain-complement is cut away leaving one as a naked brain? Recall that Van Inwagen insists that one survives in this latter case as both an organism and a thinking organism. The only difference is that in the organism now under consideration certain lower brain centres that control maintenance but supposedly do not support consciousness do not lie in the brain $B_1$ but rather in $M_1$. The functions of these centres could be very limited indeed. Though, in a normal brain, these centres may have some regulatory effect on heartbeat, digestion, endocrine activity etc., the only one of these functions that the organism could not survive without
in the short run is control of respiration. Someone could be born without these other functions and still survive. It seems then that whether or not the respiratory centres remain intact makes the difference between one surviving and not surviving. How can whether or not brain tissue with such a narrow range of function (the above list or perhaps just control of respiration) is excised make the difference between one surviving and not surviving?

Also, on Van Inwagen's theory, one survives when we cut away the normal brain-complement (and maintain the organism on life support), and Alpha survives when we destroy B₁ (but keep the centres lying in M₁ intact). It seems to follow that Alpha would survive if all that remained were M₁ maintained on life support. It is implausible that one would survive so radically truncated and retaining such a narrow potential. Even if M₁ were reconnected with a new brain and body, it is implausible that Alpha would survive the procedure.

Clearly, Van Inwagen's description of the case in which the B₁-complement is cut away conflicts with our commonsense reaction to this case. Anyone who accepts that one can survive as a naked brain is likely to accept that one can survive as a naked brain minus the lower brain centres. The only hesitation may be that these lower centres contribute to the function of the whole brain in a way that enables it to support consciousness. Yet it is clear that these centres in M₁ (as separated from B₁) do not support consciousness and make no direct contribution to B₁'s ability to do so. Most of us would find his claim that with M₁ gone there are no
experiences at all to be ludicrous. Given that the local environment of \( B_1 \) is the same as normal and that supporting consciousness is not the function of the organs cut away, to accept his claim is to invest these organs with magical powers. If (as I consider obvious) \( B_1 \) does on its own support experiences, then whose are these other than Alpha’s.

There is also an obvious argument that \( M_1 \) is dispensable. A person whose spinal cord is severed and whose body is maintained on a ventilator survives. In a person whose spinal cord is severed where it meets the brainstem, the lower brain centres have no control over maintenance at all. If these centres had no effect on the brain’s ability to support consciousness, then they could be excised without affecting survival. In the case of Alpha, if \( M_1 \) were separated from the rest of the body in an analogous way, then it is reasonable to say that the original organism survives. Why then would destroying \( M_1 \) have any effect on whether or not this organism survives. It seems clear that when the \( B_1 \)-complement is cut away and \( B_1 \) is maintained on life support, Alpha survives.

Our intuitive reactions to brain transplants and brain switches for beings such as Alpha and Beta also conflict with Van Inwagen’s theory. Suppose that \( B_1 \) is transplanted into Beta’s brainless (i.e. \( B_2 \)-less) body. Though Van Inwagen would have to say the result is Beta, I think that our natural reaction is that the result is Alpha. Clearly, our willingness to accept (as does Van Inwagen) Shoemaker’s description of the Brownson case (i.e. that Brown survives in Brownson) is due to the fact that the organ that
Brownson receives from Brown supports consciousness, not that this organ has the potential to control maintenance.

Consider also a double brain switch between Alpha and Beta. In this procedure, Alpha's brain $B_1$ is removed from the $B_1$-complement, Beta's brain $B_2$ is removed from the $B_2$-complement, and $B_1$ and $B_2$ are transplanted into the $B_2$-complement and $B_1$-complement, respectively. Given that the $B_1$-complement and $B_2$-complement contain $M$ and $M'$, respectively, Van Inwagen would have to hold that the organism now with the transplanted Beta brain is the same thinking organism as Alpha, and the organism now with the transplanted Alpha brain is the same thinking organism as Beta. Thus, if the brain is viewed as the organ of thought only, then on his theory one does not go where one's brain goes. I believe that this consequence is deeply contrary to commonsense. We believe that Alpha should care specially about the organism with the Alpha brain and that Beta should care specially about the being with the Beta brain. Clearly, to consider ourselves to be thinking organisms and our survival as sentient beings to be the persistence of thinking organisms (in Van Inwagen's way) has some highly counterintuitive consequences. Perhaps if his metaphysical case is strong that we are organisms in his sense, then these consequences may be acceptable.

However, Van Inwagen's metaphysical position seems problematic and some of his arguments in support of this position are weak. He holds that reality cannot be described fully without speaking of organisms, and he includes them in his basic ontology. Unlike nonliving objects and their activities which can be described in
terms of cooperatively arranged simples, organisms and their activities cannot be described fully in these terms. However, he holds that assertions regarding survival can be replaced with assertions regarding organisms and physical processes. It seems incongruous that we should be able to dispense with the language of survival as a sentient being but not with the language of organisms and their survival. His ontology seems to me to be extravagant in one way and yet inadequate in another. I also consider his case for this ontological position to be quite weak.

Van Inwagen has not established that organisms (at least in the general sense of including non-sentient organisms) are not just collections of simples arranged organically. Though organisms are more complicated than most other ordinary objects and artifacts, I don't really think that they are different in kind from other physical objects. With the further development of microelectronic- and nano-technologies, the artifacts we will be able to produce may be comparable in complexity and control-centeredness to organisms. Moreover, such artifacts could be just as distinctive as are organisms and even exhibit what he characterizes as jealousy. It is important to understand that the "storm" of simples that supposedly constitutes a life is just the activity of a collection of simples that interact as they do as a result of their physical properties. The behaviour of the simples involved and the apparent non-mental properties of the storm are (in principle) describable in physical terms and predictable on the basis of physical laws. Generally, his prima facie case is weak that organisms are so special (i.e.
essentially different from supposed ordinary objects) that we need to take account of them by including them in our basic ontology.

Of course, his Consciousness Argument is intended to show that organisms exist and are the only things other than simples that exist. I certainly agree with him that I exist (and that my thinking is not a cooperative activity of simples). I also consider it reasonable to suppose that there are other thinkers. However, those of us who do not subscribe to his ontology may not be convinced that our being organisms is the only reasonable hypothesis about what we are. Apart from the Cartesian and anti-cooperative reasoning involved, his Consciousness Argument is basically an argument from elimination. He has failed to refute (and indeed he says very little about) two alternative positions that may seem quite attractive and may accord better with some of our intuitions about our survival.

On one of these views, we are immaterial. Van Inwagen gives us no reason why we should reject this view. It is because of his rationalist scruples that he rejects the specific view that we are “composed of ideas.” (p.117) However, one may think that it makes perfect sense to suppose that we are so composed. The rationalist scruples to which he appeals (and indeed on which he bases much of his case) are just as much a matter of faith as is the verificationism that he rejects.

However, let’s suppose that we are material. If such is the case, then it is reasonable to suppose that we are composite. Van Inwagen has said very little against the view that we are thinking
objects. The only activity that he identifies of which composite material beings may be capable, but mere collections of simples are not, is thinking. He does not claim (nor, as explained above, do I think) that the supposed non-mental properties of organisms are not just the result of the cooperative activity of the simples involved. Recall that one of his reasons for rejecting ordinary objects and concluding that we are organisms is that Cartesian arguments do not oblige us to suppose that ordinary objects exist. However, they also do not oblige us to suppose that organisms that do not think exist. Hence, his own reasoning supports the position that we are thinking objects rather than organisms. Clearly, he has not succeeded in refuting the alternative that we are thinking objects.

To my mind, Van Inwagen makes his ontological commitment (in addition to the existence of physical simples) in the wrong place. To account for the facts of our existence and our survival, we need to think of ourselves as one of the alternatives he rejects. One who accepts that thinking cannot be a cooperative activity but rejects the view that we are immaterial should hold that we are thinking objects (thinking composite material beings) rather than biological objects (organisms). No matter how fantastically complex a supposed composite physical entity may be (and indeed organisms always are), it does not truly exist unless it supports consciousness and hence has a non-cooperative property. Perhaps it makes sense to enrich our ontology to include such thinking objects (as well as or, more plausibly, instead of organisms). I conclude
that Van Inwagen's metaphysical position, as it is, is neither sufficiently attractive nor sufficiently well-supported that we should overlook its implausible consequences.
1. Introduction

I consider Unger’s account of ourselves and our survival because, along with Van Inwagen’s account, it is one of the most systematic and initially plausible versions of physicalism and because it is representative of the physical reductionist view and the view of ourselves as thinking objects. Unger’s view has two components: an objective view of ourselves and a related physicalist account of our survival. One may find this view plausible because it seems to account for the facts of our existence and yet accords with the now widely accepted naturalism that there is just the physical world. However, I argue that Unger’s theory suffers from serious difficulties. Not only are we not so deeply committed to this objective view as he supposes but he does not provide a strong case why we should accept this view. Without this view as a basis, there is no reason to accept his physicalist theory of survival (or any other account according to which our survival consists in physical continuity of the brain and/or body).

2. Unger’s Objective View of Persons

At the heart of Unger’s account is his objective view of persons. According to Unger, we are enormously complex physical entities - living human beings. (Unger, 1990, pp.37-39) Though our
physical organization is quite different from and much more complex than that of most other living organisms and certainly inanimate objects, we are nonetheless ourselves physical objects. Certainly, we have some quite unique characteristics, certain mental capacities that are essential to our being human beings, but these capacities are realized physically in our brains. The fact that we have these characteristics does not discredit this objective view of ourselves but merely underlines the fact that our brains are our most important component. Presumably, just as Van Inwagen thinks that conscious states belong to organisms, Unger thinks that conscious states belong to bodies (or perhaps to brains). For Unger, a human body is a subject of experience.¹

This view of ourselves accords with a more general view of reality. Unger claims that we deeply believe in an objective world view according to which reality consists of a system of things and events in objective space and time, a system in which everything is related to everything else. (p.37) In this system, all thoughts and events are spatiotemporally related. Presumably, on this world view, mental phenomena and our survival (as located in this system) would be just as much the province of, and susceptible to,

¹ However, unlike Van Inwagen, Unger has no basic ontological commitment to organisms. It is not clear nor is it relevant to our discussion whether he includes physical objects in his basic ontology or whether he might view (as does Van Inwagen) such things as collections of cooperatively arranged physical simples. Though his view of ourselves as objects that are organized to have certain mental capacities and Van Inwagen's view of ourselves as organisms that have the property of sentience are not nearly exhaustive of contemporary physicalist views of ourselves, I think that, together, they are fairly representative.
scientific investigation as are any physical phenomena.\textsuperscript{2}

In trying to understand Unger's objective view, it is helpful to contrast this view with what I call the subjective view of ourselves. According to this view, we are subjects of experience whose necessary features may be quite different from the necessary features of physical objects. Though perhaps temporally extended, subjects of experience may not be spatially extended and may lack some other characteristics of physical objects. For instance, subjects of experience may not have mass.\textsuperscript{3} On such a subjective view, we may also have some necessary metaphysical properties. Perhaps the most important of these properties is what Unger calls the doctrine of absoluteness of subjects, that whether or not a subject survives is an all-or-nothing matter. (p.41) For now, I wish to leave this subjective view as general as possible. Later in this thesis, I consider a more specific subjective view of ourselves. Clearly, this general subjective view does not accord with the above objective view. He thinks that we are much more deeply committed to the objective view than to any such subjective view. Perhaps, we cling to the subjective view because we do not see how deeply it conflicts with the objective view.

\textsuperscript{2} I do not take Unger to deny that there may be a subjective (i.e. first-person) perspective on a mental event in addition to an objective (i.e third-person) perspective.

\textsuperscript{3} I should clarify that Unger uses the term "subjective view" instead for the extreme view that a person is "the bare, transcendental subject of his own experiences" that exists "neither in objective space nor even in objective time." (p.48) I have in mind a more moderate and more general view.
I should clarify my reasons for discussing Unger's objective view and contrasting it with such a general subjective view. If, as Unger thinks we deeply believe, we are certain objects (bodies or perhaps brains), then it is these objects that are the subjects of our experiences. If, as seems reasonable, we should care specially about certain future experiences because they will be ours, then it is because certain future experiences belong to these objects that we should care specially about these experiences. Hence, contrary to my anti-reductionist thesis, the survival of one who has an earlier experience to have a later experience consists in the association of the experiences with the same body or brain (i.e. in a physical relation). My intention is not to defend any particular subjective view or even to argue conclusively at this point that the objective view is wrong but only to refute the claim that we are so deeply committed to the objective view and hence to preempt this knockout of my anti-reductionist position. However, I do think that most us have deep and robust beliefs about ourselves and our survival that are (on balance) much more along the lines of a subjective view (for most of us, the specific details being intractably difficult to articulate) than along the lines of the objective view. What I consider decisive against the objective view is the complete and overall case against physicalism.
3. Unger's Theory of Survival

Not surprisingly given his objective view, Unger favours a physical approach to our survival, that is, an approach according to which to say that a person X survives as a person Y is just to say that there is sufficient physical continuity between the bodies (or perhaps the brains) of X and Y. If we are bodies or brains, then our survival consists in the persistence and continued function of our bodies or brains.

According to Unger, a person X at an earlier time survives as a person Y at a later time just in case "there is a sufficiently continuous physical realization of a core psychology between the physical realizer of [X's] core psychology and the physical realizer of [Y's] core psychology." One's core psychology consists of those psychological features common to all normal human beings. (See below for more detail.) Though sometimes he seems to accord the physical continuity of the brain complement some importance in survival, I think it makes sense to understand the physical realizer as whatever (i.e. the brain) directly supports the relevant psychology and not to include anything peripheral that serves to maintain this physical realizer.

Clearly, on this characterization, for X to survive as Y,

---

4 Actually, Unger presents his theory as a theory of identity of persons, but it can be easily understood as a theory of survival. Here, I have adapted the analysis of identity that he gives on page 109 to the purpose of discussion of survival. On page 36, he gives a rough characterization of survival that seems to accord well with this adaptation. I do not claim to expoit this theory in detail, only to sketch its major features.
there must be sufficient physical continuity of the matter between the realizers of X's and Y's psychology. Gradualness of changes of the cells and particles involved (as well as in their physical relationships) is an important element of this physical continuity. Clearly, on any physicalist account of survival, if the particles of the brain were replaced all at once, there would not be enough physical continuity for one to survive. Hence, any physicalist account of survival must have some requirement of gradualness. Unger's requirement of gradualness is very weak indeed. He writes: "the rapidity of replacement is not in itself an important matter: As long as they are relevantly even and gradual rather than uneven and abrupt, I can survive the most rapid of serial replacements." (p.123) For example, if all of the particles of my brain are replaced sequentially and gradually (i.e. perhaps each particle removed is replaced before the next particle is removed) over a second, then I survive. (p.124) However, if the exchange of the matter of the brain is uneven or abrupt, then one does not survive. For example, if the quarters of my brain are replaced successively at fifteen minute intervals, then I do not survive. (p.124)

For Unger, another important element of physical continuity is the opportunity for assimilation. (pp.147-152) If the cells of my brain are all replaced sequentially but so rapidly that the replacement cells do not have adequate time to assimilate with the other cells present, then I may not survive. However, if each cell is given time to assimilate with the other cells and into the overall function of the brain before other cells are replaced, then
I may survive. One notes that considerable opportunity for assimilation may make up for a lack of gradualness, and vice versa. For example, if each of the above quarters of my brain is given sufficient opportunity to assimilate with the rest of the brain before the next quarter is replaced, then I may survive. In the above case in which the particles of my brain are replaced rapidly, though there is no time for assimilation, the replacement may be sufficiently gradual for me to survive. Unger thinks that there are two sorts of assimilation that pertain to our survival: psychological and biological. For psychological assimilation, the new matter must have "a chance to contribute to the realization of, and to the causal support of... the mental life [i.e. thoughts, experiences, intentions etc.]" of the subject. (p.149) For biological assimilation, the new matter must have "a chance to get caught up in, and to contribute to, the biological life of the person involved." (pp.149-150) That is, there must be opportunity for the new matter to interact biochemically and physiologically with the rest of the organism.

Unger believes that psychological assimilation or biological assimilation can satisfy singly a need for assimilation. If the replacement occurs when the subject is in a coma but otherwise functioning biologically (i.e. there is much biological assimilation but no psychological assimilation), then the subject may survive. If while the subject is conscious the cells of his brain are replaced gradually with bionic units (i.e. there is much psychological assimilation but no biological assimilation), then
the subject may also survive. Generally, replacement that is sufficiently gradual and/or that provides sufficient opportunity for assimilation is favourable for survival.

Unger carefully contrasts the assimilation (as it pertains to our survival) of matter into our brains and/or bodies with the assimilation of parts into artifacts. Whether or not an artifact is adjudged to survive a number of changes depends at least partly on external relations of the result to the beings that make such judgements. (p.153) It is our attitude to the result that is important rather than just the changes themselves. For example, if a ship is renovated and then returned to its original service, we may be inclined to say that the result is the same ship as before. However, if a ship is extensively refitted for an entirely new role, then we may be inclined to say that the result is a new ship. In contrast with the survival of artifacts, "our own survival is much more heavily involved with the sorts of processes that, naturally or normally, may occur in us, that is, with internal processes." (p.154) What is important to our survival is assimilation involving such internal processes, what I will call "intrinsic assimilation." (Psychological and biological assimilation are intrinsic in this sense.) Any sense of assimilation involving external relations to other beings, what I will call "extrinsic assimilation," is not relevant to our survival.

---

5 I'm very doubtful that bionic replacement components, at least as they are presently envisioned, could have a role in supporting consciousness. However, I am prepared to give Unger the benefit of the doubt on this point.
On Unger's theory, at all times in the process in which X's brain is transformed into Y's, the core psychology must be continuously realizable. In one's core psychology, he includes the capacities to have conscious experience, to reason at a rudimentary level, and to form simple intentions, capacities that one shares with all normal human beings. In contrast, in one's distinctive psychology, he includes the aspects of one's psychology that one may share with some but not all normal human beings or that may even distinguish one from all other actual human beings. (p.68)

One may wonder what would count as a "sufficiently continuous realization of a core psychology." Because there may be times during which we are totally unconscious, it does not make sense to require that this psychology be realized at all times. What is important is that this psychology be realizable. Of course, the extent to which a psychology is realizable is a matter of degree. For instance, the psychology of a person who is in a reversible coma, of a person who is anaesthetized, and of a person who is simply in dreamless sleep are realizable with quite different degrees of directness. Clearly, Unger thinks that in all these cases the person's core psychology is sufficiently realizable for the person to clearly survive. Even in the case of a person who has been reversibly frozen, the potential is sufficiently directly realizable for the person to survive. In more extreme cases, he is not so certain. For instance, if the scattered molecules of a subject's brain and body have to be retrieved and reassembled, then the subject may not survive. While the molecules are scattered, any
potential that these molecules may have to support a core psychology may not be sufficiently directly realizable.

For Unger, the physical continuity and continued function of the brain is important only inasmuch as it provides for the continued existence of the subject’s (core) psychology. He writes: "A person’s physical parts and structures are important to her survival only insofar as they continue to support, and to realize, her basic psychological capabilities. If I am made into living meat, with no mental capabilities, I do not survive. If, a day later, a person is configured from this living meat, that person is not me." (p.71) I think that he is right to place so much importance on the continued capacity to realize psychology. Many of the implausibilities of Van Inwagen’s account are due to its failure to recognize this importance.

However, I should mention one difficulty that I can see for Unger at this point. Though, as he admits, his understanding of the term “core psychology” may be “highly, and perhaps irremediably vague,” it is clear that to require that Y’s brain support a core psychology is stronger than to require that Y’s brain support consciousness. (p.68) A brain can support experience without supporting anything resembling structured thoughts and/or intentions. However, as I have suggested in Chapters 1 and 4, it is reasonable to suppose that one may survive as a sentient being even if one’s brain is so damaged that it supports only the most rudimentary of experiences. That is, X’s special concern for Y’s experiences may be justified even if on Unger’s characterization X
does not survive as Y. Thus, in some respects, this characterization may be overly restrictive.

The source of this problem is, I believe, Unger’s failure to recognize that what is at issue is survival as a sentient being rather than survival as a person. He betrays his failure to appreciate this point when he writes: “When there is only a part of my brain left that, in addition to conscious experience, realizes only the reasoning ability of a typical dog, then I will no longer exist. There may be a subject of experience...but that will be a numerically different subject.” (p.196) If survival as a sentient being is at issue, he should require only that there be a sufficiently continuous realization of the capacity for conscious experience. This modification of his theory would not violate its physicalist spirit.

Of course, there is no proof that one survives in and just in the cases which satisfy Unger's characterization (or my suggested modification thereof) nor that one’s survival consists in the continuities that this characterization requires. Unger claims only that his characterization is the best available view of ourselves and our survival given the state our knowledge and our best general world view (a scientific world view according to which everything including ourselves is objective).
4. Unger's Methodology

What is the methodology by which Unger arrives at and defends his view of ourselves and our survival? As explained in Chapter 2, he accepts thought experiments as part of a methodology for personal identity theory. He writes: "On our methodological approach, we give great weight to our responses to examples, not only to actual cases and those hypothetical situations permitted by a correct science, but also quite a few scientifically impossible situations as well." (p.8) Recall that Unger proposes the "Avoidance of Great Future Pain Test" as an accurate though not infallible means of clarifying our responses to thought experiments. However, he insists that this methodology should be tempered in three ways.

First, though the description of a case may conflict with some of our scientific principles, the description should conform with what seems possible and with what makes sense given our overall world view. Second, the description of the case should be sufficiently and realistically detailed. Third, we may discount the significance of our reaction to a thought experiment if this reaction conflicts with a fundamental belief, a tenet of our world view, or a widely-held philosophical hypothesis. (p.8) Unger considers a methodology of thought experiments tempered in these ways a "sensibly balanced methodology" for personal identity theory. (p.8) The two following thought experiments stand out as cases that he considers to fully satisfy these requirements and to which he claims that we have definite reactions. (pp.3-4,21-23) For
convenience, I present these cases as sophisticated modes of interplanetary transport.

In the first case, the traveller is flown to Mars in a conventional spacecraft. However, on departure from Earth the traveller’s brain and body are “superfrozen,” and on arrival at Mars they are “superthawed.” That is, his brain and body are frozen and thawed in such a way that they survive and later support his original psychology. During the voyage, with the molecular activity of his body and brain suspended, the traveller is in state of suspended animation. The traveller is deeply unconscious, has no experiences at all, does not age, and on being revived is unaware of the passage of time. I will call this means of facilitating long space trips “cryotransportation.” Unger claims that our clear reaction to this case is that one survives. The second case is just the same as Parfit’s teletransportation case. He claims that (despite Parfit’s response) our clear reaction to this case is that one does not survive. He thinks that such reactions to these cases should be taken seriously.

Presumably, Unger thinks that these thought experiments show that we believe deeply that one survives in a clear case of bodily identity and that one does not survive in a clear case of bodily non-identity. That is, X survives as Y if Y’s body and brain are uncontestably numerically the same as X’s and Y retains a core psychology, but X does not survive as Y if the bodies and brains of X and Y (no matter how qualitatively similar) are uncontestably numerically different. These judgements seem to accord with and
hence to support his characterization of survival as well his
objective view of ourselves. A full treatment of the numerous
thought experiments that he offers in support of the details of his
physical characterization is beyond the scope of this thesis.

5. Metaphysical Doctrines Concerning Subjects of Experience

As Unger freely admits, the subjective view accords better
with certain beliefs (metaphysical doctrines) that many of us hold
about ourselves. (pp.39-43) For instance, he admits that his
objective view conflicts with the doctrine of separateness of
subjects (according to which a subject is fully and irrevocably
distinct from every other subject), the doctrine of indivisibility
of subjects (according to which no subject can somehow continue to
exist as more than one subject at the same time), and the doctrine
of absoluteness of subjects (according to which it is at any given
time an all-or-nothing matter whether or not one has a given
experience and hence whether or not one survives). He needs to
convince us that we do indeed have a much deeper and more
compelling belief in this objective view than in these doctrines.
Such doctrines are no more than prejudices about ourselves,
prejudices that we must renounce once we see the deep conflict
between these doctrines and our most deeply-held beliefs about
ourselves and the world.

The question of whether or not we should accept the doctrine
of absoluteness seems to be of pivotal importance. Indeed, Unger
considers this doctrine to be the one "that is most influential in
fostering dissatisfaction with the objective view, and in attracting philosophers to alternate views of ourselves." (pp.60-61) It is clear why he needs to refute this doctrine. Clearly, he recognizes an important consequence of his theory. Because physical objects are subject to changes over time, physical objects tend to enjoy at best a physical continuity rather than a perfect identity over time. Clearly, physical continuity is a matter of degree. Hence, if (as he holds) persons are bodies (with appropriately structured brains) and survival consists in physical continuity of the body, then whether or not one continues to exist (i.e. survives) at a later time must also be a matter of degree. Indeed, he insists that we are "gradual creatures." (p.64) Clearly, his objective view conflicts with the doctrine of absoluteness. Hence, in order to convince us to accept his objective view, he must first convince us that the doctrine of absoluteness is unworthy of our confidence in it.

To discredit this doctrine, Unger deploys a number of "spectrum" thought experiments. I will focus on the spectrum of congenial decomposition and the modified physical spectrum, thought experiments that I consider to be representative of his case against the doctrine of absoluteness. Though it may seem that these thought experiments have the same force, the latter does have additional argumentative value. I also consider his treatment of Parfit's physical spectrum, primarily because his description of this spectrum helps to understand his description of the modified physical spectrum.
These thought experiments seem to show that the doctrine of absoluteness conflicts with what Unger calls "principle of minute differences." (p.192) According to this seemingly very plausible principle, a minute physical difference (such as the removal, replacement or rearrangement of a relatively small number of particles or cells) cannot make the difference between a case of survival (i.e. in which the subject continues to exist) and a case of nonsurvival (i.e. in which the subject ceases to exist). Of course, he does not deny that a minute difference in physical continuity can make a minute difference in the degree of survival. Clearly, many of us would find this principle very appealing. If these "spectra" really do reveal a deep conflict between the doctrine of absoluteness and this principle, then many of us would consider these spectra to show this doctrine to be implausible.

Unger thinks that the belief in the "doctrine of absoluteness is...a rather weak belief" from which consideration of these spectra should pry us away. (p.197) These spectra show "that, as we most deeply believe, my survival is a relevantly gradual affair." (p.210) He writes: "While our logic and our language may, or perhaps may not, be all-or-none, we ourselves are not deeply absolute." (p.64)
In the "spectrum of congenial decomposition," the cells of the subject's brain are gradually removed until almost none remain. At each step in the procedure, my least important remaining cell is removed. In the case of a tie, arbitrarily pick one of the least important cells. Perhaps an epithelial cell from my leg would be the first to go. Then the next cell would be selected and removed in this way, and so on. Because the cells of my brain are arguably my most important, these would be the last to go. All pains are taken to make these changes reasonably congenial for me. Life-support systems are provided as soon as my organs start to fail. Whenever I want, I am allowed long periods of rest and/or sleep, during which no cells are removed.

When all that remains of me is half of a brain in a vat, there are probably still experiences. However, when the experiment stops when there are no neurons left, clearly there are no more experiences. The natural question is: "...when did the finite sequence of cases stop containing me?" (p.63) A defender of the doctrine of absoluteness will say that, though we may never know exactly where, this quite momentous event must have happened abruptly somewhere between these two stages. However, Unger denies that "...there [is] occasioned by the removal of just one of these millions of tied neurons, or even by the removal of only a few of them, a metaphysically significant difference..." (p.64) To suppose that there is such a difference seems to conflict with the principle of minute differences. If we insist on drawing a line
between surviving and not surviving, then where we do so is arbitrary and/or conventional.

However, given that what is at issue is the continued existence of the subject of experience, the spectrum of decomposition does not provide a strong case against the doctrine of absoluteness. We should certainly agree with Unger that after only a few cells have been removed, the subject still exists, but when only a few cells remain, the subject no longer exists. Also, he is certainly correct that the doctrine of absoluteness implies that the removal of just a few neurons somewhere in this spectrum must make the difference between the subject existing and not existing. But why should he think this consequence to be so implausible that we should question the doctrine of absoluteness? As more and more cells are removed, the subject's experience will become progressively impoverished. When the removal of just a few additional neurons will result in the subject's ceasing to exist, the level of the subject's experience is already unimaginably low. Almost nothing, then, is lost when these neurons are removed. As one would expect, the subject's experience gradually fades in its quality and intensity as this "spectrum" progresses until there is no experience at all. Perhaps this description may seem so plausible because we are in the grip of what Unger calls the "dimming light model," the view of consciousness as a phenomenon that can grow fainter until a definite point where it vanishes completely. (p.204) It is hard to imagine (and he fails to offer) any other clear picture of what is going on.
Any inclination to see this spectrum as refuting the doctrine of absoluteness is, I think, the result of failure to recognize that what is at issue is survival as a sentient being rather than survival as a person. If a subject is a sentient being, then what is essential to the continued existence of the subject is the continued capacity to have experiences. Personhood is an inessential predicate of sentient beings. Just as people can lose their hair and yet remain the same being, people can lose the capabilities that make them persons (over and above mere sentient beings) and yet remain the same being. Also, just as baldness is a vague predicate of human beings, personhood is a vague predicate of sentient beings. As the "spectrum" progresses, the subject will gradually lose not only his distinctive characteristics but also many of the core characteristics that make him a person. There is no point in the spectrum at which the subject will suddenly cease to survive as a person. However, there is a definite point at which the subject's experiences give out completely and hence at which he ceases to survive as a sentient being. Because it seems quite plausible that a minute physical difference can make the difference between a sentient being surviving to have the most rudimentary of experiences and not surviving at all, the spectrum of decomposition should not convince us to reject the doctrine of absoluteness.

However, because a minute physical difference thus makes the

---

6 Arguably, a dog is a person but perhaps to less of an extent than are most human beings.
difference between a case of survival and a case of nonsurvival, it does seem that the principle of minute differences, as stated, is incorrect. According to a more defensible version of this principle, a minute physical difference cannot make the momentous difference between a case in which the subject continues to have intense and varied experiences and a case in which the subject ceases to exist. The spectrum of decomposition does not show a conflict between the doctrine of absoluteness and this principle as modified.

7. The Physical Spectrum

As Unger constructs it, his "spectrum" of decomposition is actually a series of changes. In a series, something undergoes a sequence of minute physical changes with each such change compounding the preceding ones. In contrast, a true spectrum is a succession of separate cases such that adjacent cases differ only minutely in their physical conditions or in the events involved.

Parfit's physical spectrum is a true spectrum. In this spectrum, case $P_n$ involves the removal of $n$ thousandths of the cells of subject's brain and then their replacement all at once with exact replicas. (pp.234-236) In this spectrum, the degree of physical continuity decreases as $n$ increases. It seems clear that the subject survives in case $P_0$ (i.e. when none of the brain is

---

7 Because I consider the issue of continuity of the rest of the body a needless distraction, I simplify the physical spectrum and modified physical spectrum to involve only changes to the brain.
replaced) but not in case $P_{1000}$ (i.e. when the whole brain is replaced). In the latter case, following the replacement, a different subject is present. The doctrine of absoluteness implies that there is some $N$ such that the result of $P_N$ supports experiences that the subject survives to have but the result of $P_{N+1}$ does not support such experiences. Instead, the experiences that the result of $P_{N+1}$ supports are experiences of a whole new subject. To some, it may seem highly implausible that such a small difference in the number of cells thus removed and replaced can make the difference between the subject's continuing to exist and his being replaced with a different subject.

However, Unger doubts that the physical spectrum adds much of significance to the spectrum of decomposition. He writes:

In any given case in the Physical Spectrum...[there] is just as much and just as little to the person in question [after the removal and before the replacement]...as in the...spectrum of congenial decomposition, where just that many cells are likewise removed but are never replaced. In any given case in the Physical Spectrum, the person will survive if, and only if, the person survives in the corresponding case in our decomposition spectrum. (p. 207)

This obvious claim, together with the above response to the spectrum of decomposition, provides the basis for a reasonable explanation of how such a small difference between cases in the physical spectrum can make a difference in the resulting subjects.

Certainly, the doctrine of absoluteness implies that in the

---

8 In comparing the "spectrum" of decomposition with the physical spectrum, Unger is implicitly reconstructing the former as a true spectrum. Presumably, case n of the result involves the removal of $n$ thousandths of the cells of the brain without replacement.
physical spectrum there is some \( N \) such that the subject survives removal of \( N \) thousandths of the cells of his brain but not \( N + 1 \) thousandths. However, by the reasoning of the previous section, this consequence is not at all implausible. When \( N \) thousandths are removed, the level of consciousness of the subject is reduced to an unimaginably low level. However, because the result does support consciousness and hence the original subject still exists, when these cells are replaced and the replacement cells are assimilated with the cells not removed, the variety and sophistication of the original subject's psychology is restored. However, when \( N + 1 \) thousandths are removed, the result does not support consciousness. In effect, the removal of the slightly greater amount kills the original subject. When these cells are replaced, the result once again supports varied experiences. However, these are experiences of a numerically different subject. As Unger writes: "Since we do not allow for a person to have an interrupted existence, the later additions will do nothing to save him." (p.207) Thus, any division between cases of survival and cases of nonsurvival occurs at the same place in physical spectrum as in the spectrum of decomposition. In the physical spectrum, when \( N \) thousandths of the cells have been removed, the replacement cells enhance the experience of the original subject, but when \( N + 1 \) thousandths have been removed, the replacement cells help to support the experiences of a whole new subject.
8. The Modified Physical Spectrum

Unger suggests a modification of the physical spectrum that he thinks provides a more compelling case against the doctrine of absoluteness. (p.208) In case $P_n$ of his modified physical spectrum, cells of the brain are removed and replaced one at a time until $n$ thousandths of the cells have been thus replaced. However, they are replaced so rapidly that "there will not be enough time for any of the replacing cells to be assimilated into the original person." (p.208) Again, it seems clear that the subject survives in case $P_0$ (i.e. when none of the brain is replaced). However, according to Unger, in the final case $P_{1000}$, in which all of the cells are replaced and none of the replacement cells are assimilated before the remaining cells are replaced, the original subject ceases to exist. After the replacement process is complete and the replacement cells are assimilated with one another, there is a whole new subject. If such is the case, then the doctrine of absoluteness implies that, for some $N$, in case $P_N$ the subject survives but in case $P_{N+1}$ he is replaced with a different subject.

Unger thinks that the modified physical spectrum does create a further difficulty for the doctrine of absoluteness. For him, the most significant difference between this spectrum and the previous ones is that in the earlier spectra "conscious experience eventually gives out" but in this spectrum "there may be full conscious experience from the first case to the last." (p.208) He reasons that the absence of one brain cell between its removal and replacement does not, in itself, impair the ability of the brain to
support experience. We may accept that there is a point in the spectrum of decomposition at which consciousness finally gives out and that at this same point in the physical spectrum the original subject ceases to exist and another subject comes into existence. However, given that a high level of consciousness is supported throughout, it is hard to accept that there's a point in the modified physical spectrum at which the subject is abruptly replaced with another subject. The above response to the spectrum of congenial decomposition does not seem to help here. Unger thinks that it is so implausible that such a momentous difference could result from such a minute difference in physical continuity that we should reject the doctrine of absoluteness. As he prefers to describe the situation, as the spectrum progresses, one subject fades into another.

I do not find this spectrum much more persuasive than the preceding spectra. Indeed, I think that, given Unger's stated understanding of assimilation, his description of the spectrum is inconsistent. Recall that, on his theory, either psychological or biological assimilation can singly satisfy a need for assimilation. Let's suppose that (if indeed there is a sense of biological assimilation that may pertain to our survival) there is no assimilation of this latter sort in the cases of this spectrum. Also recall that what it means for replacement cells to be psychologically assimilated with the existent cells is for these replacements cells to have a chance to contribute to (i.e. to assist the existent cells in) the support of the mental life (the
thoughts, experiences, intentions etc.) of the subject. Given that, in all cases of the spectrum, there is "full conscious experience" at all times during the replacement process, I don't see how he can consistently deny that there is complete psychological assimilation in all cases. In cases in which a large proportion of the cells have been removed and replaced, there are not enough existent cells to support such full experience (and in the later cases not enough existent cells to support consciousness at all). Hence, in such cases, the replacement cells must contribute to the support of such experience. That is, there must be psychological assimilation. Hence, cells that are truly not assimilated do not contribute to the support of experience.

Let's suppose, then, that (as could be the case) the replacement is too rapid for any of the replacement cells to be assimilated. As the spectrum progresses, the number of cells needing to be assimilated grows, and hence the number of cells that are able to contribute to the mental life supported by the brain decreases. As a result, the experience supported after the replacement but before the ensuing assimilation becomes more and more rudimentary until it ceases completely. There is some N such that, after the replacement, in case N the brain supports only the most rudimentary of experience but in case N + 1 there are not enough existent cells to support consciousness and the subject has ceased to exist. When N thousandths of the cells of one's brain are sequentially removed and replaced rapidly, one's consciousness is temporarily greatly impoverished and, when the new cells are
assimilated, the variety of one's experience is later restored. However, when \( N + 1 \) thousandths are thus removed and replaced, one's consciousness is obliterated. The replacement is too rapid for the new cells to be assimilated, and so there are not enough cells working together to be able to support consciousness. In this case and the later ones, the brain is in too much of a state of flux to be able to support consciousness and hence the original subject ceases to exist. In effect, the replacement process kills this subject. Then, when the new cells are assimilated with the existent ones, the result is a whole new subject.\(^9\)

Hence, in this spectrum, there is also a reasonable explanation of how there can be such a division between cases of survival and cases of nonsurvival. Because the minute physical difference between case \( N \) and case \( N + 1 \) does not make the difference between the subject having rich and varied experiences and not surviving at all, there appears to be no violation of the principle of minute differences, as stated.

What Unger seems to have in mind is some sense of assimilation such that the existent cells together with the unassimilated cells can support experience even if the existent cells on their own cannot. Not only does this contradict his own stated understanding of assimilation but I doubt that there can be any such sense of assimilation that pertains to our survival. Certainly, it is hard

\(^{9}\) Because after the replacement but before assimilation what remains of the brain that supports experience is exactly what would remain if the cells were simply removed but not replaced, the division point occurs at the same place as in the earlier spectra.
to see how there could be such a sense of psychological assimilation. What else can it mean to say that the replacement cells are assimilated psychologically with the existent cells other than that the replacement cells assist the existent cells to support experience?

Nor do I see how biological assimilation can be at all relevant to this issue. Biological assimilation is a sort of assimilation that might secure one's survival through changes that the brain may undergo when it does not or cannot support experience (i.e. in the absence of psychological assimilation). I do not see how, in itself, it can be advantageous with regard to the survival of the original subject for replacement cells that already assist in supporting experience to have an extended opportunity to interact biochemically and/or physiologically with the rest of the brain and/or body. Such an opportunity may make no difference to the physical state of, or the physical relationship between, the new and existent cells. I don't see why, when the replacement cells work together with the existent cells to support experience, one would survive any less (why the resulting experiences would be any less one's own) because of a lack of this sort of extended biological assimilation. I can't think of any intrinsic sense of assimilation that could pertain to our survival other than psychological and biological assimilation.

I think that Unger is correct that any notion of extrinsic assimilation is irrelevant to the issue of our survival. Whether or not one survives is a matter of fact that does not depend on
relations to others. If one's survival were something that depends even in part on people's attitudes, then whether or not one survives would be at least in part a matter of convention. It is hard to accept that whether or not we survive could be conventional. Since it seems clear that any sort of assimilation must be intrinsic or extrinsic (or perhaps a combination of these), it appears that there can be no sense of assimilation that pertains to our survival such that the existent cells together with the unassimilated cells can support experience even if the existent cells on their own cannot. Hence, Unger's description of the spectrum simply does not make sense.

Because an extended treatment of assimilation is beyond the scope of this thesis, I do not want to tie my response to this spectrum too closely to a rejection of this notion. What I think is plausible to deny is that, in cases in which a new subject results, there is conscious experience (or even the capacity for conscious experience) continuously present throughout. In such cases, at some times during the replacement process, there is no subject present at all. Recall that if a subject is a sentient being, then what is essential to his continued existence is the continued capacity to have experiences. In those cases in which there is consciousness continuously present throughout, even though the original subject may temporarily lose the attributes that make him a person (over and above a mere sentient being), he continues to exist. When the replacement is complete and the replacement cells are assimilated, his capabilities are restored. However, in those cases in which
there is not this continuous consciousness (or capacity for continuous consciousness), the original subject ceases to exist. When the replacement and assimilation is complete, there is a whole new subject.

One may object that, in the modified physical spectrum (and in the physical spectrum), a minute difference in physical continuity does make the momentous difference between there being and there not being a switch of subjects. However, this minute difference in physical continuity makes only a minute difference (during the replacement process) as regards the existence of the original subject (i.e. the difference between his having the most rudimentary experiences and not existing at all). I think that the principle of minute differences would seem less appealing if it were taken to imply that a new subject of experience cannot come into existence (when the brain can once again support consciousness) after the original subject has ceased to exist. Though the physical spectrum and modified physical spectrum are harder to handle than the spectrum of congenial decomposition, I do not think that these spectra show decisively a deep conflict between the doctrine of absoluteness and the principle of minute differences. Understood properly, the above spectra are not nearly so compelling as Unger thinks.
9. The Case For Absoluteness

So should we accept the doctrine of absoluteness? I concede that no argument can convince someone that justification of egoistic concern (and hence survival in my sense) is all-or-nothing who does not already see it as such. However, I think that there are two major factors that tell in favour of this doctrine.

First, it is hard to imagine what it would be like to survive to a reduced degree. Suppose that my tormentors inform me that there will be a painful experience $F$ about which I should feel specially concerned but to only a reduced degree. How exactly are my expectations to reflect such a prediction regarding my future? In Williams' terminology, I find myself unable to project myself and my feelings of special concern less than fully into a future experience. I think of myself as either having or not having the experience, and I am unable to imagine myself as somehow experiencing this pain only partially. I find myself with no way to think about $F$ as my tormentors have described it. I am unable to conceive of how feelings of special concern could be justified partially.\footnote{For similar reasons, Williams doubts that it can be indeterminate whether an experience involves a given person. (Williams, 1975, pp. 194-196) As I have explained in Chapter 4, I think that there is also an obvious difficulty in projecting one's feelings of special concern into the experiences of more than one subject at a given future time. Hence, this methodology of projection also supports the doctrine of indivisibility (and the view that there can be no survival without identity).}

To convince me that survival is not all-or-nothing, someone would need to provide a model for what it's like to survive
partially. One suggestion is that F could be impoverished in intensity and/or sophistication as may be the experiences in the spectrum of decomposition. Certainly, one may worry less about an anticipated painful experience if one believes that this experience will not be very intense. However, my tormentors have not depicted this experience as weakened but rather as having a weakened special relationship (of the sort that justifies egoistic concern) to my present experience. Another suggestion is that there could be a less than one hundred percent chance that I should be specially concerned about F. However, if what my tormentors have in mind is that myself and a number of my fellow victims will draw straws to determine who is to be tortured, then they have misrepresented my fate. Depending on the outcome of the draw, my special relationship to F will be entirely present or entirely absent rather than weakened. As far as I know, no one has offered any more promising alternatives.

Second, I'm not sure that the (modified) principle of minute differences is defensible and, hence, that any conflict between the doctrine of absoluteness and this principle should discredit this doctrine. After all, a minute physical difference (such as the migration of a minute quantity of arterial plaque) can make the difference between life and death. Of course, in this case, the minute physical difference leads to a widening cascade of physical changes that truly makes the difference between living and dying, whereas the momentous difference between continuing to have varied experiences and not surviving at all, that according to this
principle a minute physical difference cannot make, would not be mediated by any greater physical changes. Nonetheless, I think that this analogy should be given some weight.

For me, the existence of such a sharp boundary in terms of physical continuity between cases of survival and nonsurvival is an acceptable consequence of the doctrine of absoluteness. However, others may find this consequence unacceptable and reject the subjective view as a result. Because qualitative similarity and physical continuity (and arguably any causal relationship between experiences at an earlier time and experiences at a later time) are matters of degree, any reductionist theory implies that survival is a matter of degree. Hence, our commitment to the belief that survival is all-or-nothing is a problem for all reductionist theories, not just Unger’s.

I doubt that the issue of whether or not survival is all-or-nothing can be decisively settled. Fortunately, this absoluteness argument is just one of many arguments against reductionism. Nonetheless, I suspect that our notion of egoistic concern is so inextricably conjoined with the belief that its justification is all-or-nothing that, if it should turn out that this justification would have to be a matter of degree, it would be most reasonable to reject the notion of egoistic concern as hopelessly flawed and to revise many of our associated deep beliefs. It is important to remember that I’m not trying to show that egoistic concern is ever justified. My thesis is only that reductionism cannot explain how egoistic concern can be justified.
Interestingly, Unger himself fails to come fully to grips with the extent to which his theory conflicts with our intuitions, commonsense views, and deep beliefs about survival. For instance, despite his protestations that we are "gradual creatures," he seldom talks as if survival is a matter of degree. Throughout his book, he talks as if there is a determinate fact about whether or not one survives, that is as if survival is all-or-nothing. For instance, he talks as if survival is all-or-nothing when he talks about "barely getting through" and "barely missing." (p.219) In cases in which we lack robust intuitions about whether or not one survives and/or there is a reduced degree of physical continuity, he should clearly recognize that somehow one may survive to a reduced extent.

10. Unger's Methodological Problems

Certainly, most of us do have a great deal of confidence in scientific inquiry, and many of us believe all intelligible questions to lie within its scope. To see us as other than physical objects whose activities are essentially physical processes is to locate us outside of Unger's system and hence outside the scope of scientific inquiry, as it is presently understood. However, most of us also have deep and firmly-held beliefs about ourselves as subjects of experience, beliefs that the above metaphysical doctrines in part reflect. If forced to face the conflict between the objective and subjective views, one is very likely to reject the former rather than the latter. It may seem very attractive to
think of us as objects only until one realizes the conflict of this view with other deep beliefs about ourselves. Clearly, there is little basis for Unger's claim that our supposed deep beliefs clearly favour his view when in fact for most of us our beliefs about ourselves and our survival are hopelessly confused, often contradictory, and in need of careful reconsideration and/or clarification. Moreover, even if, as he thinks, "what we deeply believe" is that we too fall within such an objective order, then this otherwise unfounded belief seems a mere prejudice in favour contemporary scientific dogma. The issue is precisely whether or not we do in part lie outside of this order, and to appeal to such "deep beliefs" is simply to beg the question.

What is needed is a thorough probe of our beliefs about ourselves that forces us to recognize any contradictions among these beliefs and to resolve these contradictions by choosing between those that favour the objective view and those that favour the subjective view. When faced with such choices, it is not at all clear which beliefs we would be more readily prepared to give up. I have two main methodological objections to Unger's use of thought experiments to probe these beliefs.

First, throughout his book there are numerous instances of bizarre cases about which he claims that we have clear intuitions regarding whether or not one survives. In some of these cases, his intuitions one way or the other do not seem at all convincing. For instance, how could we possibly have credible intuitions that one survives when the particles of one's brain are replaced
sequentially but very rapidly but that one does not survive when the quarters of one's brain are replaced at fifteen minute intervals? (p.124) He has misjudged that our dominant responses to these cases are along his lines and indeed that we have dominant responses to (or any deep beliefs about) such cases at all. Applying the Avoidance of Great Future Pain Test certainly does not help to clarify my reactions to such cases. Rather than having clear intuitions or deep beliefs about whether or not one survives in such cases, we are likely to see ourselves as engaged in just the sort of "guessing games" against which he warns. (p.61)

Second, recall that Unger accepts a methodology in which thought experiments may be discounted if they conflict with widely-held philosophical views or with our world view as a whole. I think that this methodology is flawed. Hopefully, our philosophical views are open to revision. How are we to determine whether or not we should maintain these views except by thought experiments and other devices that help us to assess our commitment to them? Unger treats his objective view of ourselves as a widely-held philosophical view and element of our world view, and he seems to think that thought experiments can be discounted if they conflict with this objective view. He seems prepared to accept thought experiments when they deliver the verdict that he seeks but to reject them when they clash with his world view. In this way, he bases much of his case on the prevalent unquestioned confidence in the all-inclusiveness of the objective order. Not only does he fail to appreciate the role of thought experiments in evaluating our most basic beliefs
but he treats his own philosophical and world view as that with which our reactions to the relevant thought experiments must accord. In using his objective view as a constraint on his thought experiments, he begs the question in favour of this view.

One should note that even if thought experiments do show a conflict between the doctrine of absoluteness and the (modified) principle of minute differences, they do not in themselves favour the objective view over the subjective view. The principle of minute differences is certainly an appealing tenet of the objective view. If one accepts the objective view and this tenet thereof, then one will consider the thought experiments that reveal this conflict to refute the doctrine of absoluteness and hence the subjective view. However, equally, the doctrine of absoluteness is certainly an appealing tenet of the subjective view. If one accepts the subjective view and this tenet thereof, then one will consider the thought experiments that reveal this conflict to refute the principle of minute differences and hence the objective view.

Ultimately, one must choose between the commonsense subjective view of ourselves and one's faith in the all-inclusiveness of the objective order. Whether or not one is prepared to accept the objective view depends on whether one gives greater weight to one's belief in an entirely objective world order or to such attractive doctrines as the doctrine of absoluteness. Certainly, we are likely to accept Unger's arguments that the doctrine of absoluteness is false only if we already accept an objective view of ourselves and of reality as a whole. On balance, I think we have a greater and
deeper commitment to this doctrine and the subjective view than to the objective view.

11. Does One Survive Cryotransportation?

Unger is certain that and claims that we believe deeply that one would survive cryotransportation. However, if one does not survive teletransportation, then it is implausible that one would survive cryotransportation. I think that any intuitions or deep beliefs that we would survive cryotransportation are unfounded.

Certainly, it may be physically possible to freeze and later thaw a human brain in such a way that the result can support consciousness. However, we cannot be certain whether one who has the experiences before the freezing survives to have the experiences following the thawing. We cannot be sure that the freezing of the brain does not interfere with some continuity of activity that is needed to secure one's survival. Any deep belief that we may have that this case is a case of survival is based on the unfounded assumption that a clear case of bodily identity is a case of survival. Clearly, from a third-person point of view, this assumption may seem reasonable and under normal circumstances we may be unwilling to abandon this assumption. However, in everyday life, we encounter people whose brains and bodies have had quite ordinary histories and never people with exotic histories involving freezing/thawing and the like. It is clearly a prejudice and indeed begs the question to extend this assumption to such cases. There is no reason to trust what we "deeply believe" about these cases.
Unger's reaction to this case is nothing more than a prejudice, one that may be reinforced by exposure to science fiction. There is even some reason to suppose that one does not survive cryotransportation.

Suppose that in a comparison of cryotransportation and teletransportation, the following experiments are conducted. (A) X's brain and body are superfrozen, flown to Mars in a conventional spaceship (the flight taking slightly less than one year), and finally thawed exactly one year after being frozen. (B) X's brain and body are superfrozen, and then teletransported to Mars, where they are kept frozen, and finally thawed exactly one year after the original is frozen. We may wonder what difference between the original frozen/thawed brain shuttled to Mars and the replica produced on Mars by teletransportation can account for the original surviving in the former but not the latter case. Suppose that, following production on Mars of a replica of one's superfrozen brain and body, this same brain and body (rather than being destroyed) are flown to Mars, and that just before thawing the two frozen and physically indistinguishable (even at an atomic level) bodies are lying on adjacent slabs in a Martian laboratory. It is hard to believe that there is a difference between these bodies that can account for the momentous difference in the ownership of the ensuing experiences. The two frozen bodies are like two indistinguishable pieces of meat in a freezer. If somehow life can be restored to these, it is hard to understand how the one in case (A) can be any more closely related to (and more qualified to be a
We may suspect that there is some process that continues in the frozen body that teletransportation interrupts. However, we can imagine an even more radical sort of freezing (hyper-freezing) in which even the most basic atomic activity (and hence any such process) would cease. If all relevant activity has really ceased, then the difference between these cases with regard to the survival of the original would have to be secured magically. Hence, if one does not survive teletransportation, then it is implausible to suppose that one would survive cryotransportation.

The conclusion that one may not survive cryotransportation presents a serious problem for most physicalist theories (including Unger’s). (However, this conclusion may not conflict with Van Inwagen’s theory. Van Inwagen might accept that the organisms before the freezing and after the thawing are not the same organism. The freezing may end one life, and the thawing may start another.) Clearly, cryotransportation is an extreme case of bodily continuity. While during the normal process of aging there are gradual bodily changes as well as an exchange of particles, there are presumably no such changes while one is superfrozen. Hence, if one does not survive cryotransportation, then one may not survive as a sentient being who has uncontestably one’s numerically same brain and body. If such is the case, then an objective view of ourselves as physical objects must be wrong.
1. Introduction

Because a physicalist theory supposedly presents an analysis of in what survival consists, (i.e. of what constitutes survival) if such an analysis is correct, then it is a necessary truth that one survives in just those cases that satisfy the conditions of the proposed analysis. In particular, one can survive only as someone whose body and/or brain is sufficiently physically continuous with (or, roughly speaking, numerically the same as) one's own. However, in this chapter, I argue that we can conceive of surviving to have experiences that are not associated with our numerically same brain and hence that no physicalist theory can provide such an analysis.

2. One Can Conceive of Disembodied Survival

In this section, I argue that we can conceive of (i.e. that our concept of ourselves and our survival does not rule out) disembodied survival, that is survival to have experiences that are not associated with any body. Hence, it seems, we do not think of ourselves as bodies and our survival as survival of the body. Depending on one's religious beliefs, one may think that disembodied survival actually occurs. One may think that after the death of one's body, one may survive somehow as a disembodied spirit. However, though clearly no satisfactory explanation of how neural activities produce experiences has yet been offered, it is,
I think, a reasonable hypothesis that experiences do result from brain states and/or processes. Thus, I consider such a belief that disembodied survival actually occurs to be implausible but nonetheless coherent.

If indeed disembodied survival is conceivable, then we should be able to imagine roughly what it would be like.1 Certainly, it is hard to imagine how in a disembodied state we could retain certain of our senses. For instance, how could we have tactual and/or bodily sensations without a body? Touch involves contact between one’s body and other physical objects. Bodily sensations, such as pains, seem to have bodily locations. Hence, it would seem that without a body, there could be no pain. Visual experiences tend to represent objects as located in a certain direction and at a certain distance from a point of reference (from one’s eyes and thus from one’s body). On the basis of visual experiences, one should be able to extrapolate one’s position. In this way, vision tends to locate one in space. If, as we may think, without a body one would have no location, then it is hard to imagine how one could have visual experiences as they are normally conceived.

So just what sorts of experience could a disembodied being have? Though while in a disembodied state one may have diminished opportunities to employ one’s powers of reason, this faculty could remain largely intact. It is also tempting to imagine disembodied survival as possibly also involving olfactory and/or auditory

---

1 See Penelhum, 1970, pp.19-36, for a discussion of disembodied perceptions involving the various senses.
experiences. Smell seems to be entirely non-directional, and hearing does not seem so directional as vision. However, there are people for whom this latter sense is accompanied with well-developed directional sensitivities, and the rest of us may have these sensitivities to a lesser extent. However, the directional sensitivities accompanying this sense in my case are so rudimentary that I have no difficulty in conceiving of auditory sensations devoid of all directionality. Perhaps likewise, visual and tactual sensations may be impoverished in their directional sensitivities. Indeed, I think that it is possible to imagine entirely directionless and hence nonbodily visual and tactual experiences. For instance, one could have a free-floating directionless visual experience or perhaps a feeling of an all pervasive agony. Generally, whether or not one could retain all of one's senses, I have no difficulty in imagining what disembodied survival might be like.

Though a disembodied mind is not associated with a body, I think that it is intelligible to suppose that a disembodied mind could have nonetheless a visual (and/or auditory) point of view. Terence Penelhum writes: "...there seems no difficulty in saying of a disembodied person that it might look to him as though there were objects before him which looked to him as they would look to a normal observer under optimum circumstances from a certain position in space." (Penelhum, 1970, p.25) If one's disembodied visual experiences could present themselves as from such a spatial perspective, then they could be just like normal visual
We may think that it is possible to conceive of disembodied survival in which one has the full range of sensations. While dreaming, one may have perceptions of objects even when no actual objects initiate the causal processes leading to these perceptions, and one may have equally illusory perceptions of one’s body. One can imagine a dream that is so vivid that one’s experiences are indistinguishable from one’s normal conscious experiences. In a particularly elaborate dream, one may imagine investigating scientifically the supposed external world and as a result even formulating the hypothesis that one’s body is the cause of one’s experiences. Perhaps disembodied survival may be like such a dream in which one has all the usual experiences without any actual physical basis. One notes that though, of course, it is a reasonable hypothesis that there is an external world, one can’t refute the skeptic who doubts that such is the case. If one can’t be certain that there is material world, then even though it may also seem a reasonable hypothesis that one really has a body and that this body causes one’s experiences (i.e. that one is embodied), one can’t be sure that there are any bodies at all. The fact that one can’t be certain that one is presently embodied shows that disembodied survival must be conceivable.

Strangely, a number of philosophers whose theories do not seem compatible with disembodied survival are sympathetic to some measure with it’s conceivable. As is consistent with his organicist views, Van Inwagen thinks that when one’s life as an
organism ends one "should expect to die."² (Van Inwagen, 1990, p.211) Given Van Inwagen's theory, this would seem to be an analytic truth. Certainly, for the life of an organism to end is for the organism to die. Thus, if one is an organism, when one's life (i.e. the life of the organism that one is) ends, one does indeed die. However, he also states: "Whether one expects to have any experiences after one's death is, of course, a matter of one's religious beliefs." (Van Inwagen, 1990, p.211)

Van Inwagen's attempt to keep the door open in this way to life after death is, I think, utterly inconsistent. If one is an organism, then when the organism that one is dies, one ceases to exist. Presumably, when one ceases to exist, one does so irrevocably. Clearly, it makes no sense to ascribe experiences to something that does not exist. So if one is an organism, it is an analytic and hence necessary truth that one cannot have experiences after one's biological life ends. Given his religious leanings, Van Inwagen probably thinks that in fact one may survive as a disembodied spirit to have certain experiences. However, he cannot even entertain that there can be experiences after the death of the body, unless he doesn't (as he insists) equate one with an organism and one's survival with organic survival. To entertain that one can have experiences after death, he needs to have a sense in which one continues to exist after the death of the body. One can't be both

² Actually, this is what Van Inwagen says he would expect to happen if he were partitioned like Neocerberus. But the point is that partitioning would end a life. (Van Inwagen, 1990, p.211)
an organism that as a necessary truth ceases to exist on its bodily
death and a being that may look forward to life after death. One
cannot maintain such contradictory claims by relegating one of them
to one's religious beliefs and thus shielding it from scrutiny.
Philosophy doesn't end where one's religious beliefs begin.

Another philosopher whose theory of personal identity seems
incompatible with disembodied survival but who is nonetheless
sympathetic to its conceivability is Peter Strawson. According to
Strawson, we are neither pure consciousnesses (as the subjective
view may have it) nor bodies (as the objective view would have it)
but rather persons, entities that have necessarily both physical
and mental attributes. For Strawson, "...the concept of a pure
individual consciousness...is a concept that cannot exist...as a
primary concept in terms of which the concept of a person can be
explained or analyzed." (Strawson, 1959, p.102)³ This concept "can
exist only...as a secondary non-primitive concept, which itself is
to be explained, analyzed, in terms of the concept of a person." (Strawson, 1959, pp. 102-103) Thus, Strawson does not deny that one
may have a pure consciousness. However, a pure consciousness has at

³ Strawson's terminology may lead to some confusion. On the
page of this quote, he employs the terms "pure [individual]
consciousness," "ego," and "subject of experience" as synonyms.
Yet, later he also speaks of "the individual consciousness of...[a]
subject of experience..." thus clearly distinguishing these terms.
(Strawson, 1959, p.114) Though nothing hangs on this
interpretation, it may be reasonable to suppose that he thinks of
"pure ego" and "subject of consciousness" as synonyms but "pure
[individual] consciousness" as referring to a succession of
experiences, to what some philosophers may call "the stream of
consciousness."
best a "logically secondary existence." (Strawson, 1959, p.103)

Though he does not consider the concept of a body and the concept of a pure consciousness to be logically separable, he does seem to think that it is conceivable that one's consciousness could continue after the death of one's body. It follows that "...from within our actual conceptual scheme, each of us can quite intelligibly conceive of his or her individual survival of bodily death." (Strawson, 1959, p.115) Presumably, one's feelings of special concern would be justified for the experiences of one's pure consciousness following the death of one's body, and hence one would survive in my sense to have these experiences. Hence, on his theory, disembodied survival in my sense is conceivable.

However, according to Strawson, the idea of a person never embodied is not intelligible. In order for one to exist after the death of one's body, one must retain the "logical benefit of individuality from having been a person." (Strawson, 1959, p.103) As Strawson writes, "...in order to retain his idea of himself as an individual, he must always think of himself as disembodied, as a former person." (Strawson, 1959, p.103) He could have no idea of himself unless he had been embodied. While disembodied, "he must live much in the memories of the personal life he did lead." (Strawson, 1959, p.116) Eventually, when these memories fade after lengthy disembodiment, he will lose the concept of himself. However, though he may indeed lose the concept of himself, I see no reason why the embodied and hence conceptually advantaged among us cannot think of this unfortunate as still existing and having
experiences. Thus, while still embodied, one would have reason to be specially concerned about an unpleasant experience that one may have after losing the concept of oneself following lengthy disembodiment.4

As Van Inwagen's and Strawson's views attest, there is some acceptance of the conceivability of disembodied survival in the literature on personal identity. There is also a strong independent case that we can conceive of disembodied survival.

3. One Can Conceive of Transfer

Closely related to the objection to physicalism that one can conceive of disembodied survival is the objection that one can conceive of transfer (i.e. survival to have experiences associated with a brain and body numerically distinct from the brain and body with which one's present experience is associated). If transfer is really conceivable, then clearly we do not think of ourselves as bodies or our survival as survival of the body. So, is transfer really conceivable?

I know of very little that is said in the contemporary literature explicitly in defence of the view that transfer is conceivable. However, given the views of certain philosophers, it is reasonable to infer that they would be sympathetic to this view.

4 The issue of whether or not one can survive after losing the concept of oneself is clearly a point of deep disagreement between me and Strawson that I do not attempt resolve in this thesis. The important point here is that, one way or another, he thinks, as do I, that disembodied survival is conceivable.
For instance, it is hard to imagine how Strawson can hold that disembodied survival is conceivable without accepting that transfer is also. If one can conceive of surviving in disembodied form so long as one's memories of having been embodied enable one to retain the concept of oneself, then one should have no difficulty conceiving of transfer. If one can lose a body, then why shouldn't one be able to regain one? Indeed, if after lengthy disembodiment these memories and hence one's concept of oneself were to fade, by regaining a body one could reestablish one's grip on this concept and thus ensure one's survival. By regaining a body, one becomes once again a complete person. Of course, one may think that even if transfer has a place in Strawson's theory, his understanding of survival is so different from my own that his views lend no support to the conceivability of transfer in my sense. However, given my sense of survival, there is a plausible case that we can conceive of transfer.

Suppose that following a wild party the night before, one awakens with a pounding headache. Though cloudy-headed, one remembers much of what happened at the party and regrets the excesses of the night before. Though one's past experiences had left one with a terror of the consequences of excessive imbibing, one had quickly got into the swing of the party, forgotten all caution, and downed one drink after the other. Bleary-eyed, one attempts to focus on the clock. Much to one's surprise, it's already past noon. Fortunately, one has no plans except to meet a friend in the afternoon. One summons the courage to get out of bed
and get dressed, and one heads lazily towards the kitchen. After a half-hearted search of the refrigerator turns up nothing remotely palatable, one puts on the coffee machine. The coffee black and strong seems almost enjoyable. An hour later, after savouring two cups and now feeling more composed, on the way out the door, one passes a mirror and finds to one's shock the face of one's worst enemy. Touching one's face and finding not one's own familiar features but a neatly-cropped beard barely concealing the pockmarked face of one's adversary, one must face the unthinkable: one has one's worst enemy's body. One searches for some way of explaining what has happened. At the party there had been some highly skilled but notoriously eccentric surgeons. Perhaps, after one had passed out, they switched one's brain with that of one's worst enemy - the ultimate practical joke. However, there are no suture marks on one's skull. Now certain that this body, brain and all, is one's worst enemy's, one's horror is indescribable. Suddenly, one awakens feeling shaken but surprisingly fresh and discovers that it was all a dream.

One has imagined (one may say conceived of) transfer. One has imagined what it would be like for there to be experiences (the agonies of a hangover, the terrors of an inexplicable and clearly unwelcome circumstance) about which one should have been specially concerned but that are associated with a brain and body that are

---

5 Perry considers a similar case. See Perry, 1977, pp.3-4. Of course, Perry's case and my own are inspired by Locke's original case of the prince and the cobbler. (Locke, 1975, II, xxvii, 15)
numerically distinct from the brain and body with which one's previous experiences were associated. Of course, to say that transfer is conceivable is not to say that transfer ever occurs but rather that it is intelligible to entertain the possibility that transfer may occur.

4. An Epistemological Objection to Transfer

Bernard Williams offers a number of arguments why as a conceptual fact bodily identity is necessary for personal identity and hence why transfer is inconceivable. (Williams, 1973c) Because he holds that we are bodies and endorses the methodology of projection, he needs to convince us that (despite all appearances) we cannot coherently project ourselves and our feelings of special concern into future experiences associated with a different body. I believe that one of these arguments is relevant and worth considering at this point.

Williams argues that in order to make sense of transfer we would have to appeal ultimately to considerations of physical numerical identity. He thinks that it is incoherent to entertain the possibility of transfer unless it is possible in principle to determine whether a person at one time with a certain body is the same person as a person at another time with a different body. Presumably, in the absence of evidence of numerical identity of their bodies, we would have to fall back on considerations of psychological continuity involving memory and/or personality. He focuses on the view that, in the absence of bodily identity, we can
make sense of personal identity in terms of memory. The apparent memories that one has (and the corresponding memory claims that one makes) of having witnessed previous events and especially of having performed certain previous actions may help to identify the person with these memories as a particular person at the time of the events in question. Because, as he correctly notes, memory claims are not all veridical, we need corroborative evidence for claims to have witnessed past events and performed past actions. Though such evidence need not be directly bodily, he argues that ultimately we must appeal to considerations of bodily identity.

Corroborative evidence could be in the form of reports of other eyewitnesses. However, Williams claims that corroboration of witnessing and doing must appeal to bodily continuity of the other witnesses. He considers the case of Charles who exhibits remarkable psychological similarities to Guy Fawkes and who believes himself to be the same person as Guy Fawkes. Presumably, to corroborate Charles' memory claims, one would have to rely entirely on relevant memory claims of others. But these memory claims would in turn require corroboration, and so on. Ultimately, certain relevant memory claims would have to be corroborated in terms of bodily continuity. Williams concludes that "it is a necessary condition of making the supposed identification on non-bodily grounds that at some stage identifications should be made on bodily grounds." (Williams, 1973c, p.11) It seems, then, that we would be unable to reidentify persons in a world in which transfer was a normal occurrence. He implies that we cannot conceive of such a world.
For Williams, it does not follow directly that we cannot conceive of even a single case of transfer. He recognizes that if Charles has uncanny knowledge of the details of Fawkes' life, then we may have a very strong temptation to say that "Charles is now Guy Fawkes, that Guy Fawkes has come to life again in Charles' body..." (Williams, 1973c, p.8) Certainly, if the tabloid presses were to learn of this remarkable case, they would be inclined to describe it in these terms. However, according to Williams, we are certainly not forced to accept this description of the case, and, on reflection, we may find just about any reasonable alternate explanation (even a quite bizarre one) of the occurrence to be more attractive. Failing to find such an explanation, we might even be inclined to describe the occurrence as mysterious. (I will say more about this in the next section.) Roughly, his argument is that no evidence would provide an adequate basis to conclude in a given case that transfer has taken place, and that therefore it makes no sense to even entertain the possibility of transfer.\footnote{For the purpose of discussion, I have greatly simplified his arguments. According to Williams, it is also because the notion of transfer is so problematic in other ways that we should be reluctant to accept any such evidence that it has occurred. His main argument is that, if Charles could acquire such supposed memories (and other characteristics of Fawkes'), then his brother Robert could too. Because they can't both be Fawkes, it is implausible that either of them (separately) could be Fawkes. Because, as explained in Chapter 4, I do not think that one can have more than one survivor at the same time, I agree that no psychological evidence of transfer can be absolutely conclusive.}

However, any temptation to accept that the conclusion follows from the premise is, I think, due to failure to separate the
metaphysical and epistemological problems of survival. Even if, as Williams thinks, we cannot find adequate evidence to identify persons in a world in which transfer is a normal occurrence, it does not follow that we cannot conceive of such a world. It is even clearer that we can conceive of a single instance of transfer.

Even if we may feel inclined to accept the above inference, Williams' case for the premise is weak. He is certainly correct that we should not accept uncritically Charles' claims to remember being Guy Fawkes and doing what Guy Fawkes supposedly has done. Before we should accept that Charles really is Guy Fawkes, we should want indeed to corroborate Charles' apparent memories and convince ourselves that he could not have obtained such an accurate picture of the events of Fawkes' life by some other means that would be easier for us to accept than his really being Guy Fawkes. If Charles claims to remember a particularly bizarre episode in which no one presently alive would imagine that Fawkes could have been involved, and if an obviously previously undisturbed written eyewitness account of the events were subsequently unearthed that perfectly matched Charles' recollection, then we may start to think that Charles is indeed Guy Fawkes. One notes that this evidence does not involve considerations of bodily identity. Indeed, because the witnesses wrote down their account right at the time of the episode, their identity over time is not an issue.

One may object that we cannot be sure that Charles is not instead one of the witnesses to the episode who is using his special knowledge to convince people that he is Guy Fawkes. There
are some obvious reasons why this objection does not present a serious difficulty. First, Charles may recount several episodes all corroborated in this same way but all with different witnesses. His being able to do so would seem to identify him uniquely. Second, to accept that Charles is one of the long-dead witnesses is already to accept that there can be identity without bodily identity. Why then would we be reluctant to accept that Charles is Fawkes?

One may also object that, like apparent memories of events and reports of purported witnesses, records such as written accounts require corroboration. However, because reports of apparent memories of an event are fallible even when the bodily identity of the reporter and the witness can be established, such memories also require corroboration. To corroborate evidence for a given claim means to obtain additional evidence for the claim. Of course, the more evidence available and the higher its quality, the more probable the claim is rendered. Certainly, evidence of bodily identity may provide corroboration for psychological evidence of personal identity. However, additional psychological evidence can also corroborate psychological evidence. A preponderance of evidence from a range of sources may provide strong (though not conclusive) support for a claim of personal identity. However, according to Williams, without corroboration in terms of bodily identity, no psychological evidence should in itself be sufficient evidence for us to accept a claim of personal identity.
5. The Priority of Psychological Evidence

Despite Williams' insistence, it seems clear that we actually favour psychological evidence of survival over evidence in terms of physical continuity. Certainly, from one's own point of view, psychological evidence (especially evidence in terms of memory) seems more important. If one were to wake (as does the person after the party) with apparent memories of having been a person with a different body, then one would insist that one was the same person as the person formerly with that body.

Unless one expects unique treatment in this regard, one's preference for psychological evidence in one's own case should extend to the case of others. Certainly, if one encounters (as would the friend that the inebriate had planned to meet) a person that one recognizes as a close friend but who claims to be another close friend who has woken up with the body now presented, one would at first suspect being the victim of a bad joke or fear for the friend's sanity. However, if he were to persist in his claims, and to demonstrate remarkable psychological similarity to one's other friend, as well as intimate knowledge of the details of his life, one might eventually become convinced by these claims. A later and similar encounter with a person that one recognizes as the other friend may convince one that his two friends have indeed switched bodies. Any lingering doubt that such is the case would be a result of doubting one's friends' veracity rather than of doubting that if the psychologies associated with their bodies had really switched, they would indeed have switched bodies. Certainly,
if we were presented with a case in which the psychologies associated with two bodies were truly switched, we would be inclined to understand this case as a case of exchanging bodies.

Of course, one may object that such cases never occur and that invariably we do rely on physical continuity in order to reidentify people. Even when a person's physical appearance has changed so dramatically that it is easier to reidentify him on the basis of psychological factors, we would not doubt the verdict that considerations of physical continuity would yield. In fact, in case of doubt, we consider the evidence of physical continuity, such as fingerprints and dental records, as decisive in reidentification. However, one may argue that we are prepared to rely on considerations of physical continuity precisely because cases in which psychological considerations would overturn the verdict of physical continuity never seem to occur. We simply do not tend to encounter cases in which a person with one body claims to have had another body and can present compelling psychological evidence to that effect. Nonetheless, such cases are conceivable, and, in such cases, the psychological evidence would seem to take priority over evidence of physical continuity. Though one would no doubt be perplexed initially by a case in which the two sorts of evidence disagreed, many would find it clear upon reflection that the psychological evidence takes precedence.

One may also object that if we were to find a person who is psychological indistinguishable from someone in the past with a different body, then it would be hasty to draw the bizarre
conclusion that these persons are the same. Perhaps, it would be (as Williams thinks) preferable to say simply that such a phenomenon is mysterious. There are many bizarre phenomena in the world that defy explanation, for instance that numerous people have similar apparent memories of alien abductions. Such an exact psychological similarity would be just one more phenomenon that so far defies explanation. In fact, Williams tentatively suggests that clairvoyance might explain how one's apparent memories may be suitable to be those of someone else without appealing to the identity of these persons. However, one may think that there may be less exotic and more plausible explanations. For instance, one may have these apparent memories as a result of suggestion or delusion.

If, in a remarkable case of psychological similarity, such explanations seem inadequate, then we may simply say that (as with the abduction memories) something has occurred that has not yet been adequately explained.

However, the case of similarities of memories of alien abductions and the case of someone who is psychologically indistinguishable from someone else would be phenomena that clearly differ greatly in their level of mystery. In the case of people with common abduction memories, we may look for an explanation in terms of exposure to accounts of abduction in the media, suggestion, and/or mass hysteria. With a little imagination, we can think of numerous explanations that do not involve extraterrestrial visitations. Even if proliferation of similar cases and careful study thereof were to show the abduction hypothesis to be
plausible, this discovery would not greatly change my world view. There is certainly a high probability that there is intelligent life elsewhere in the galaxy. Such a discovery would only cause me to question my earlier assessment of the probability of its spatial proximity. However, finding someone psychologically indistinguishable from someone in the past is a much more inexplicable phenomenon that would occasion much more serious adjustment of my belief system. Any lingering attraction that I may have to physicalism would be lost. In adjusting to this discovery, I would reconsider many of the objective and scientific elements of my world view. I might even espouse a view of ourselves, that I now find ludicrous, as souls in which memories somehow inhere, a view on which transfer would not just be conceivable but may actually occur.

One may object that the probability of there existing in a vast universe at possibly quite different places and quite different times two beings with an uncanny psychological resemblance may be significant. There is no reason, then, why an observation of this phenomenon should occasion such a dramatic change in our belief system. However, as the following analogy shows, this objection is specious. In the trillions of coin tosses that have taken place, someone almost certainly has observed a sequence of forty heads. However, if while gambling I were to observe forty straight heads, I would be certain that the coin must be weighted. Likewise, even if the probability of there being two such psychologically similar beings were significant, the
probability that we would encounter or hear of such beings would be vanishingly low. Unless, it can be established on independent grounds that personal identity consists in bodily identity, the psychological evidence should be considered the more compelling. Hence, the above epistemological objection to the conceivability of transfer seems weak.

6. A Physicalist Objection to Transfer and Disembodied Survival

It is important to remember that the central claim of physicalism is that we are able to describe reality including one’s survival in physical language. In particular, physicalism claims that we are able to describe one’s survival in terms of the language of physical continuity or numerical identity of physical objects. If, as most of us assume, one survives just so long as one’s numerically same brain retains certain capabilities (i.e. one’s special concern is justified for and only for future experiences to be associated with the same brain as one’s present experience), is it not possible to express any meaningful assertion about survival in terms of the continuing existence and capabilities of brains? How can it be meaningful to speak of transfer of consciousness and disembodied survival if these never occur and are indeed physically impossible? One response to this objection is that it is often perfectly meaningful to assert things

7 To say that transfer and disembodied survival are physically impossible is to say that the universe (everything physical and indeed, if there are such things, everything nonphysical) is as a matter of fact such that these cannot occur.
that are physically impossible. For example, it is meaningful to speak of an object defying the law of gravity. Equally, it should be meaningful to assert that one survives without retaining the numerically same brain.

One may object that there is an important difference between these cases. We have a vivid picture of how an object can defy the law of gravity. Indeed, cartoons often depict this as happening. However, just what it would be like to survive to have experiences associated with another body or with no body at all, especially those that are significantly different from one's accustomed experiences, is not nearly so easy to visualize. Certainly, one can imagine to some extent what the experiences associated with another body might be like, experiences that one may confusedly think of oneself as somehow surviving to have. However, equally one can imagine what the experiences of any other person might be like. One's picture of these latter experiences is no less vivid and, one may think, no different in kind than that of experiences associated with another body that supposedly one would survive to have.

Of course, when one tries to conceive of survival to have experiences associated with a different body or with no body at all, one may imagine that along with the experiences that one survives to have are available memories of one's earlier experiences and/or one's feelings of special concern for these later experiences. However, if the experiences one imagines were really to occur, then these perhaps merely apparent memories would be one's only evidence that one has survived. Such memories would
have an aspect (i.e. the indexical aspect) that represents one as having been involved oneself in what they depict and/or having anticipated these later experiences oneself. However, as already conceded, there is no way to be sure that one’s memories and this aspect thereof are faithful. So there would be no way to be sure that one’s special concern for these experiences was really justified and hence that one has survived. One would not experience directly any relation between the earlier and later experiences that guarantees that the feelings of special concern were justified.

However, the same is true even when the later experiences are associated with the same body as the earlier experiences. Nonetheless, we are able to project our feelings of special concern towards future experiences associated with our same body. We imagine ourselves as having the experiences in question. We do not project these feelings towards the body. We project these feelings directly towards the experiences. We have no more difficulty in projecting these feelings towards the experiences associated with a different body. Indeed, it is exactly as hard to conceive of surviving to have future experiences associated with our same body as to have those associated with another one. Though disembodied experiences may be in some ways different from our accustomed experiences, I don’t see any greater impediment to our imagining ourselves having such experiences. The conviction that we can conceive of transfer and disembodied survival is just as unshakeable as the conviction that we can conceive of surviving to
have later experiences associated with our same brain and body. I conclude that this physicalist objection to transfer and disembodied survival has no force.
Chapter 8
Reference and Survival

1. Introduction

In this chapter, I consider the following possible objection to the claim that disembodied survival and transfer are conceivable and hence that we cannot be bodies and/or brains. Let's grant for a minute that our present concept of ourselves (when we are interested in egoistic concern) is compatible with transfer and disembodied survival. Perhaps, when egoistic concern is at issue, we think somehow of ourselves as the subject of a succession of experiences that need not be associated with the same brain or indeed any brain. However, we may think of ourselves as such because we are unaware of our own essential properties and hence confused about our survival. Given the scientific evidence that there is a neural basis of experience and our belief system as a whole, it may be a plausible hypothesis that we are brains. If we are brains, then, even though it may make perfect sense to care specially about the future experiences associated with one's brain, to suppose that one should be specially concerned about future experiences not associated with one's same brain is simply unintelligible. However, though my response may not convince the physicalist, I will argue that this objection does not seriously undermine the case that transfer and disembodied survival are conceivable. Indeed, to press this objection ultimately begs the question in favour of physicalism.

2. Referential Semantics and the Necessary A Posteriori
Recent developments in the theory of meaning and reference may seem to some to lend support to the above objection. Before the works of Putnam and Kripke on meaning and reference, it was taken for granted that the meaning of a term is lodged as a concept in the head. (Kripke, 1980; Putnam, 1973) Such a "concept" is a complex of beliefs or expectations (the intension of the term) about the nature of the things to which the term refers (the referent or the extension of the term). The intension of a term is the meaning of the term and indirectly secures the reference of the term. That is, something is to lie in the extension if and only if it satisfies the intension.¹ Largely because of compelling arguments due to Putnam and Kripke against this internalist view, it is no longer the dominant view of meaning and reference.

An alternative to such an internalist view of meaning is the referential view (often misleadingly labelled "externalist"), according to which the meaning of a term is determined, at least primarily, by the reference of the term. Putnam argues that the meaning of natural kind terms is lodged, as the referents of these terms, in the external physical world rather than "in the head". (Putnam, 1973) Kripke’s view might be considered a generalization of Putnam’s view. According to Kripke, the meaning of a term (or at least an important aspect of this meaning) is its reference. A term designates its reference directly, that is without the mediation of a concept or description of the referent. In the case of natural

¹ See Putnam, 1973, for a discussion of the assumptions underlying such an internalist view of meaning.
kind terms, the meaning may be lodged in the external world. As with a natural kind term, the meaning of a mental term is its reference. However, the theory does not prejudge the true nature of the referents of these terms (whether the states referred to are physical, irreducibly subjective etc.). In this section, I outline this theory of referential semantics and explain why this view might seem to some to support the above objection to the conceivability of transfer and disembodied survival.

How exactly does a term get its meaning on such a referential view? Consider the case of the term "gold." According to the referential theory, something is gold if and only if it bears the relation "same-S" ("same stuff as" to adapt Putnam's terminology) to the normal instances of what we call "gold," as represented by a number of exemplars. The term "gold" gets its meaning from these exemplars. I will treat "same-S" not just as an intra-world relation that relates samples of substances in the actual world but also as a cross-world relation that may extend to samples in different possible worlds. Gold is whatever bears "same-S" to the exemplars in the actual world.

We may suppose that whatever is the Lockean real essence of gold is what we mean by "gold." If, as we would expect, these

---

2 Kripke treats natural kind terms in much the same way as he treats proper names. He considers a proper name to be a tag that does not have associated with it any descriptive conditions that as an analytic matter the referent of the name must satisfy. I do not claim to represent Kripke's view in detail, only to sketch a theory of referential semantics that helps to understand and resolve the above objection.
exemplars are samples of an element that has atomic number 79, then this real essence is what in fact has this atomic number. Any stuff without the right microstructure is not gold because it doesn’t bear “same-S” to the right local stuff. Because our pre-atomic concept of gold involved the surface properties of gold but not the correct microstructure, not surprisingly this concept did not determine the extension of the term. Our pre-atomic idea of gold was incomplete or even inaccurate and needed to be filled in or corrected by empirical discovery.³

Though our pre-atomic concept of gold in terms of its surface properties was inadequate, this concept (which we may think of as consisting of descriptive conditions defeasibly associated with the term) would have enabled us to tentatively identify a sample of a solid as gold and hence as something that might serve as an appropriate exemplar. In this way, even though it may not reflect the complete and accurate nature of the referent, our concept (of the referent of a term) may indirectly “fix the reference” (Kripke’s phrase) of the term. (Kripke, 1980, p.135) (In this sense, there is an internal as well as an external aspect to meaning.) So long as this concept enables us to select appropriate and sufficiently representative exemplars, in using the term we may still succeed in referring to the referent. In pre-atomic times, we were able to apply the term “gold” without knowing the true nature

³ Of course, the concept of gold depends to some extent on the speaker. For simplicity, I ignore the difficulties of any supposed division of linguistic labour. I assume that at any given time our complexes of beliefs about the referent of a term agree.
Kripke noticed that this theory of meaning has major consequences for the theory of necessary truth. According to Kripke, a designator is "rigid" if it has the same referent in every possible world. (Kripke, 1980, p.48)\(^4\) For example, because something in another possible world is gold just in case it bears same-S to the exemplars of gold in the actual world, "gold" is a rigid designator. It is a mistake to think that it is a mere contingent fact that gold has atomic number 79 in the actual world. Rather, gold has this atomic number in all possible worlds, and hence it is a necessary fact that it does so. Because we can only determine the nature of gold by empirical investigation, that gold has this microstructure cannot be known a priori. Hence, it is a necessary a posteriori fact that gold has atomic number 79. (Kripke, 1980, pp.123-125,138)

Though there is no possible world in which gold does not have atomic number 79, we can imagine circumstances in which it would be rational to suppose that gold does not have this nature. Indeed, in pre-atomic times, when we all thought of gold as whatever has its well-known surface features and had no idea about its microstructure, we were in just such circumstances. It is important to distinguish and not to equivocate between two senses of

\(^4\)To say that a term has the same referent in different possible worlds is to assert that there is a cross-world identity relation (such as same-S) between the referents in these worlds. As Kripke understands such relations, they may relate individuals (persons, particular experiences etc.).
conceivable: epistemically possible, and metaphysically possible. An assertion is epistemically possible if it is not contradictory given our concepts of the referents of the terms involved. An assertion is metaphysically possible if there is a possible world in which it is the case. In pre-atomic times, it was conceivable in the sense of epistemically possible that gold does not have atomic number 79, but it was not (and is not) conceivable in the sense of metaphysically possible that gold does not have atomic number 79. That is, it is possible that our concept of gold might not accord with what we now believe to be its true microstructure, but (given that our belief is correct) it is not possible that gold might not have this microstructure.

Some philosophers believe that the referential theory favours a physicalist view of ourselves and our survival. As Nagel writes: "...the concept of the self does not tell us fully what kind of things we are...Our idea of ourselves is one whose exact extension is determined in part by things we don’t necessarily know simply in virtue of, or as a condition of, having the concept: our true nature and the principle of our identity may be partly hidden from us." (Nagel, 1986, p.39) That is, we cannot infer our true nature from the subjective concept of ourselves that most of us presently possess. As with gold, our concept of ourselves is incomplete or inaccurate and needs to be filled in or corrected by empirical discovery regarding our true nature. To wrongly suppose that our concept of ourselves reveals the conditions of our survival leads to the illusion that we could exist independently of our brain and
body. That we can conceive of transfer and disembodied survival (in the sense that these scenarios do not conflict with our subjective concept of ourselves and our survival) does not show that there is any possible world in which we could survive to have experiences associated with another brain or not associated with a brain at all. That is, transfer and disembodied survival may be epistemically but not metaphysically possible.

Gary Wedeking objects to my position on just this basis. According to Wedeking, what I claim to be our concept of ourselves and to be compatible with transfer (and I think to reflect our true nature) is just a complex of beliefs about the referent of "me." This "concept" allowing for transfer is based on a mistake about what we are that derives from confusing epistemic possibility with metaphysical possibility. According to Wedeking, the reference of "me" is fixed by various observable aspects of my mental life. However, it is reasonable to suppose that the reference of "me" is the Lockean real essence of myself, which is in fact my brain, or the relevant (i.e. causally productive) aspects thereof. That is, when we refer to ourselves, we are referring to what in fact plays this causal role in our life. Wedeking thinks that it may be reasonable to construe Unger as holding such a position when he talks about our notion of ourselves reflecting our best general world view (a scientific world view according to which everything including ourselves is objective). Though I will not argue that the

---

5 Wedeking communicated this objection to me both in conversation and in writing.
notion of a hidden self is in itself objectionable, I will respond
(in Section 6) to Wedeking's objection by providing a plausible
case that if egoistic concern is ever justified, then (in conflict
with his position) transfer and disembodied survival are
metaphysically (rather than merely epistemically) possible. To
reject this case out of hand is to beg the question in favour of
physicalism.

3. Reduction and the Identity Theory

Recall that the basic idea behind any sort of reduction is
that things of one sort are shown to be "nothing but" (Searle's
phrase) things of another sort. (Searle, 1994, p.112) (As I have
explained in Chapter 2, I understand "nothing but" in an
ontological sense.) For instance, on the basis of empirical
discovery and resulting theoretical advancements, it is now widely
accepted that heat is nothing but molecular motion. Likewise, a
reductionist with regard to experiences may think that experiences
are nothing but brain states. A reductionist with regard to persons
may think that persons are nothing but bodies and/or brains.

How exactly are such reductions established? Heat was once
thought of as something that feels a certain way. Because our
feelings of heat have been explained ultimately in terms of
molecular motion, we now accept that heat is nothing but molecular
motion. That heat is nothing but molecular motion does not imply
that there is no heat, that the notion of heat should somehow be
eliminated. To establish that this phenomenon is nothing but (and
hence to identify this phenomenon with) molecular motion is to locate this phenomenon in the molecular view of matter.

Before tackling the problem of whether the true nature of a person is a brain, we need to consider whether the true nature of an experience, such as pain, is a brain state. According to the identity theory, mental states are nothing but brain states. For instance, pain may be nothing but C-fibre stimulation. Certainly, we have discovered a close relationship between brain states and experiences. As most of us accept, experiences are caused by brain states and processes. Indeed, there is said to be strong evidence of the constant conjunction of pain and C-fibre stimulation. The identity theorist may think that, by uncovering this evidence, we have discovered that pain is in fact C-fibre stimulation.⁶

The referential theory may incline one towards this theory. Our present concept of an experience may be as something of a purely and irreducibly subjective nature that could exist apart from the brain, but this concept may not be a full or accurate picture of the nature of an experience. Perhaps we have formed such

---

⁶ As Searle notes, this proposed identity cannot be correct empirically. He writes: “C-fibres are a type [among other types] of axon that transmits certain sorts of pain signals from peripheral nerve endings to the central nervous systems.” (Searle, 1994, Chapter 2, Note 5) The neural processes that are directly responsible for sensations of pain occur in certain regions of the brain itself. Because this empirical mistake is widespread in the philosophical literature (and indeed in Kripke’s monograph), I will continue to talk of the claimed identity between pain and C-fibre stimulation, but this identity claim should be thought of as shorthand for an identity claim between pain and what is indeed responsible for pain in the human (or, more generally, the vertebrate) brain. I am indebted to Phil Hanson for pointing out this mistake.
a concept of an experience because the true nature of an experience may be hidden from us from the subjective perspective. Because examination of our present concept of an experience does not reveal any necessary a posteriori connection between the mental and the physical, we may be misled into thinking that an experience could exist apart from the brain. It may be epistemically but not metaphysically possible that an experience could exist apart from the brain. Perhaps the true nature of an experience may be something (a brain state) that cannot exist apart from the brain.

I should stress that an identity theorist may accept the thesis that particular mental states are particular physical states (i.e. the token-token identity thesis) but deny the thesis that types of mental states are types of physical states (i.e. the type-type identity thesis). For example, a type-type claim may be that pain is C-fibre stimulation. A token-token claim may be that this particular pain is a particular brain state (which may be an instance of C-fibre stimulation). The token-token view is now the more widely accepted version of the identity theory.  

4. Kripke’s Argument Against the Identity Theory

Kripke himself argues on the basis of his referential view that experiences are not brain states. (Kripke, 1980, pp.144-155)

---

A token-token identity theorist need not deny that there are types of mental states. Surely, pain is such a type. This theorist may claim that types of mental events are disjunctions of the physical states (some of which may be of the same physical type) that are mental states of this type.
I focus for now on his argument that type-type identity claims, such as the claim that pain is nothing but C-fibre stimulation, are false.

According to Kripke, "pain" and "C-fibre stimulation" are rigid designators. It seems clear that "pain" has the same reference in every possible world in which there is pain. Because C-fibre stimulation is a well-defined neural event, "C-fibre stimulation" also has the same reference in every possible world in which there are C-fibres to be stimulated. Hence, if "pain" and "C-fibre stimulation" have the same reference in the actual world, then these designators have the same reference in every possible world. If pain is C-fibre stimulation in this world, then pain is C-fibre stimulation in all possible worlds. That is, if pain and C-fibre stimulation were identical, then this identity would be necessary. It would be wrong to think that it is a contingent property of a brain state that it is a mental state. We cannot accept that pain could exist separately from C-fibre stimulation without accepting that pain is necessarily distinct from such stimulation.

Clearly, pain appears to us as something that could exist entirely independently of any brain state. We can imagine experiencing pain even if our brain were to have no such state, and (as my arguments for the conceivability of disembodied existence show) we can imagine experiencing pain even if we lacked physical states altogether. We can also imagine experiencing no pain even if our brains have this state. That is, we might not be in pain even
though the C-fibres of our brain are stimulated, and we might be in pain even though the C-fibres of our brain are not stimulated. If we think in terms of possible worlds, we may suppose that there are possible worlds in which there is C-fibre stimulation without pain, and vice versa. So, it seems that pain and C-fibre stimulation cannot be identical in any possible world, even in the actual one.\footnote{Most materialists would not find the claim to be at all objectionable that physical states other than brain states (and C-fibre stimulation in particular) could (perhaps in aliens and in terrestrial organisms with different sorts of central nervous systems) also be pains. The possibility of multiple realizability of types of mental states is one of the attractions of the token-token view. Following Kripke, I focus for now on the (instructive, if no longer attractive) type-type claim that pain is in fact C-fibre stimulation.}

However, the identity theorist may even find some encouragement in Kripke's argument. The identity theorist may readily accept that pain and C-fibre stimulation coincide in all possible worlds. Perhaps, it is because we have an incomplete and inaccurate concept of pain that we so readily suppose that pain could exist without C-fibre stimulation, and vice versa. If the identity theorist can show that this feeling of contingency is explainable and hence illusory (that is, explain away these feelings), then there is reason to resist our inclination to think of pain as separate from any brain state.

The case of the heat/molecular motion identity provides a hint of how the identity theorist might do so. "Heat" and "molecular motion" are also rigid designators. So this identification would also have to be necessary. Though at one time people thought of
heat as something that feels a certain way and had no inkling that it had anything to do with motion of particles, we now accept that heat is nothing but molecular motion. The apparent possibility of heat not turning out to be molecular motion was merely epistemic and hence an illusion. Perhaps, likewise, the apparent possibility of pain not turning out to be C-fibre stimulation might be thought to be an illusion. If the identity theorist can show that this is an illusion, he would thus provide a good reason why the premise that pain could exist separately from C-fibre stimulation is false. However, Kripke argues that the materialist cannot explain away in this way the appearance of contingency of the connection between mental and physical states.

According to Kripke, there is an obvious way to explain the appearance of contingency of certain necessary a posteriori truths. Suppose that "A" and "B" are rigid designators and that it is a necessary a posteriori truth that A = B. If the qualitative features of A and B on the basis of which we recognize phenomena as referents of these designators coincide only contingently, then this contingency produces in us the illusion of contingency of the statement "A = B." Indeed, if such is the case, then two conscious beings could be in exactly the same epistemic situation, could use the same designator to refer to the underlying phenomena, even though these phenomena may be very different.

In the case of heat and molecular motion, we recognize heat by the sensation S that heat produces in us. However, sentient creatures on a different planet may not experience S in the
presence of molecular motion but instead in the presence of light waves of a certain frequency. In the presence of these waves, these creatures would be in the same epistemic situation as we are in the presence of molecular motion, they could use the word (i.e. "heat") that we use to refer to the phenomenon underlying S, and yet it would not be molecular motion which causes in them sensation S.

Can we say something on these lines to explain away our feeling that the identity of pain and C-fiber stimulation is merely contingent? Kripke argues that the case of pain and C-fibre stimulation is not analogous to the case of heat and molecular motion. Molecular motion could exist without being felt as heat, without producing sensation S. However, if pain were C-fibre stimulation, then C-fibre stimulation could not exist without being felt as pain. If a type of experience is a type of brain state, then this brain state cannot exist without being felt as the type of experience in question. As Kripke notes, "In the case of molecular motion and heat there is something, namely the sensation of heat, which is intermediary between the external phenomenon and the observer. In the mental-physical case, no such intermediary is possible, since here the physical phenomenon is supposed to be identical with the internal phenomenon itself." (Kripke, 1980, pp.151,152) It is possible to have sensation S (and hence be in the same epistemic situation as we are) in the absence of heat and not to have S (and hence not be in the same epistemic situation as we are) in the presence of heat. However, there can be no such intermediary in the case of pain. As Kripke writes, "To be in the
same epistemic situation that would obtain if one had a pain is to have a pain; to be in the same epistemic situation that would obtain in the absence of pain is not to have a pain." (Kripke, 1980, p.152) The appearance of contingency of the connection between pain and C-fibre stimulation cannot be explained away in the same way as the appearance of contingency of the connection between heat and molecular motion.

As Kripke diagnoses the difference, the reference of the rigid designator "heat" is fixed (i.e. determined "for the real world and for all possible worlds") by an accidental property of its referent, its ability to produce sensation S in certain beings. (Kripke, 1980, p.132) In contrast, the reference of the rigid designator "pain" is determined by "the property of being pain itself, by its immediate phenomenological quality." (Kripke, 1980, p.152) We can explain away the appearance of separability only if we pick out the phenomenon by an accidental rather than an essential feature. Clearly, we pick out pain by its essential phenomenological nature. Any phenomenon that we recognize in exactly the same way that we recognize pain must be pain. Because our feelings of contingency of the connection between pain and C-fibre stimulation cannot be explained away in this manner, these feelings provide good reason to believe that pain cannot be (identical with) C-fibre stimulation.

Clearly, not everyone will find such Kripkean arguments decisive. Indeed, even Kripke regards "the mind-body problem as wide-open and extremely confusing." (Kripke, 1980, footnote 77,
The materialist may object that even though it may be true that we cannot explain away the appearance of separability by treating the sensation of pain as an accidental property of pain, there may be other ways to handle this appearance of separability. For instance, Brian Loar suggests a way to account for the appearance of separability (between pain and C-fibre stimulation) that is compatible with the identity of experiences and brain states. (Loar, 1997) I do not claim to be able answer this

According to Loar, one conceives something (a property, entity, process, event etc.) under a concept, or equivalently one uses a concept to refer to (i.e. to fix the reference of a term referring to) the thing in question. One can conceive pain under two different concepts: the obvious phenomenal concept and the physical concept (as a type of brain state). Loar accepts Kripke's claim that the phenomenal concept of pain does not conceive pain under an accidental mode of presentation, that this phenomenal concept conceives (and picks out) its referent "directly and essentially." He also accepts that phenomenal concepts are not conceptually reducible to physical concepts in the sense that for something to fall under a physical concept does not entail that it falls under the corresponding phenomenal concept or vice versa. However, he rejects what he calls Kripke's Semantic Premise: An identity statement linking conceptually independent concepts that are used to refer to something can be true (i.e. the concepts fix the same reference) only if at least one of the concepts conceives the thing in question under an accidental mode of presentation.

However, there are two major of problems for Loar. First, to my mind, Loar has not adequately explained (in terms of phenomenal concepts being recognitional and recognitions being gestalt and perspectival) how we can be unable to "see a priori" that two concepts that conceive something essentially and directly pick out the same thing. Surely, if two concepts "capture the essence" of something, then we should be able to see this a priori. I do not find his claim persuasive that this complaint equivocates on "capture the essence." Second, if, as Loar accepts, for something to fall under a physical concept does not entail (or otherwise enable us to see) that it falls under the corresponding phenomenal concept, then we are unable to explain in physical terms how an experience can "feel like this." He accepts this "explanatory gap," but he does not think that this gap is a problem for physicalism. I think that the existence of this gap is fatal to this view. An in depth treatment of Loar is beyond the scope of this thesis.
objection to the satisfaction of the materialist. Nor do I deny that there may be other objections to the above Kripkean argument. I remind the reader I am not tackling the mind-body problem in this thesis. Like Searle, I take it for granted that Kripke’s anti-materialist argument is decisive. (Searle, 1994, p.117) I mention this objection here primarily so as not to leave the reader with the impression that there is no materialist response to Kripke’s argument.

I don’t see any serious difficulty in adapting Kripke’s argument to the case of supposed token-token identities. Suppose that A is a particular experience (say the painful spinal spasm that I am presently feeling) and B is the corresponding brain state (say the present state of my brain or the relevant state of a part thereof). As above, “A” and “B” are rigid designators, so if A were identical with B, then this identity would be necessary. A presents itself to me as of a certain qualitative nature. However, because (from my subjective point of view) I have no access to the underlying brain state, A does not present itself to me as a brain state. Thus, A seems to me as something that could exist apart from the brain state in question and indeed any brain state. That is, any dependence of A on B seems merely contingent. Because according to the token-token theory, the experience is (essentially and necessarily) the brain state, the experience is essentially such as it presents itself to me. In experiencing A, I have access to the essence of this experience. Anything that now presents itself to me exactly as does this pain must be this pain. Thus, the apparent
contingency of the connection between A and B cannot be explained away by claiming that the brain state only presents itself to me as the experience in question but isn’t. Thus, by Kripke’s reasoning, there is good reason to suppose that A cannot be B. That is, a token experience is not a token brain state. Hence, there seems to be a strong case that the identity theory (type-type and token-token) is false, that mental states are not brain states.

I conclude that there is a strong case that it is not just conceivable in the sense of epistemically possible but also conceivable in the sense of metaphysically possible that an experience can exist apart from the brain with which it is associated. If a particular experience can exist either in association with a brain or apart from any brain, I can see no reason why an experience could not be (why it should not be metaphysically possible for an experience to be) associated with a brain that is numerically distinct from the brain with which it is actually associated.

5. Kripke’s Argument That We Are Not Bodies and Doubts About This

Kripke offers a similar modal argument that persons are not bodies. According to Kripke, Descartes argued that a “person or mind is distinct from his body, since the mind could exist without the body.” (Kripke, 1980, p.144) On the basis of his referentialist semantics, Kripke defends this argument. He argues that you cannot accept the premise that the mind could exist without the body without accepting the conclusion that a person or mind is distinct
from his body. Because we readily accept the premise, there is a strong prima facie case that a person is not a body.

According to Kripke, a proper name is a paradigm case of a rigid designator. For instance, "Descartes" is the rigid designator of a certain person. We may also consider "Descartes' body" to be a rigid designator. Hence, if Descartes were identical to Descartes' body, then the identity of the referents of these rigid designators would be necessary. That is, Descartes and Descartes' body would not just coincide in the actual world but indeed could not (i.e. in any possible world) exist separately. If a person is a body in the actual world, then that person must be that body in any world. Because it seems clear (and my arguments for the conceivability of disembodied survival and transfer seem to show) that a person could conceivably exist apart from his body, a person is not (i.e. even in the actual world) a body. Roughly, Kripke’s argument is that an identity between a person and a body would have to be necessary, and we have reason to believe that this identity could not be necessary.

Unfortunately, this argument is far from decisive and indeed could be turned on its end to yield the very opposite conclusion. A physicalist may argue that because (as our scientific world view may incline us to suppose) a given person is a particular body in

---

10 By "Descartes' body," I do not mean whatever body Descartes may have in a given possible world in which he exists but rather the body that he does in fact have. Thus I stipulate that in another possible world, "Descartes" body refers to whatever (if anything) bears "same body as" to his body in the actual world.
the actual world, that person must be that body in any world. The apparent contingent coincidence of a person and his body in the actual world is in fact necessary. Hence, in no sense can disembodied survival and transfer be possible.

Interestingly, in the same monograph, Kripke entertains, though rather tentatively, a position - his view of the necessity of personal origin - that apparently conflicts with the view that persons are not bodies (and/or brains). The considerations that he raises may lead us to doubt that persons are really so separable from their bodies and thus to doubt that transfer and disembodied survival are conceivable. He argues that it is intuitively plausible that one could not have come from a different egg and sperm than the ones from which one did in fact originate. It seems reasonable that a table could not have been fashioned from a different piece of wood than that from which it has in fact been fashioned. (Kripke, 1980, pp.113,114) Why, then, should we expect that a person could have had a different biological origin?\textsuperscript{11}

Largely because of such considerations, Kripke finds "Cartesian dualism" unattractive. He writes: "If we had a clear

\textsuperscript{11} These considerations may also lead us to doubt that specific mental states and specific brain states are really as separable as the above argument against the token-token view seems to suggest. If a person could not have a different origin, then perhaps a particular experience could not have a different origin, that is not result from the very brain state from which it does in fact result. Arguably, if a person could not have a different biological origin, then that person's first experiences could not have a different basis (i.e. not be associated with the brain/body with which they are associated). How, then, could his later experiences, be associated with a different brain or with no brain at all?
idea of the soul or the mind as an independent, subsistent, spiritual entity, why should it have to have any necessary connection with particular material objects such as a particular sperm or a particular egg?" (Kripke, 1980, footnote 77, p.155) What exactly does he mean?

One may characterize Cartesian dualism as the view of the self as a separately-existing subject of experience (possibly pre-existing and outliving the body), that, during one’s biological lifetime, interacts with one’s body and brain. Presumably, the origin of a person is the origin of his soul and is not to be equated with (and indeed may predate) the union of the sperm and egg in question. Clearly, Kripke’s view that a person could not have a different biological origin does seem counter to Cartesian dualism, as characterized. If one is a separately-existing subject, then why couldn’t that subject have associated with a different body from the one with which it did in fact originally associate?

But does this rejection of this Cartesian view present a problem for the view that transfer and disembodied survival are conceivable? Even if a person could not have a different origin, he could still have a different future. Kripke thinks that even though, for instance, Queen Elizabeth could not have had a different origin, “One can imagine...that various things in her life could have changed....” (Kripke, 1980, p.111) He does not insist that the person’s subsequent experiences would have to be associated with the same body and hence that transfer and disembodied survival are inconceivable. Rather, though he does not
make his position entirely clear, he seems to agree with Sprigge that it is conceivable that Queen Elizabeth could even be changed into a swan, arguably a case of transfer. (Kripke, 1980, p.111) However, no matter what Kripke's opinion, some may think that if you couldn't have a different biological origin, then you couldn't have such a radically different future.

I think that there is good reason to resist the view that one could not - in a sense relevant to this thesis - have had a different biological origin. Certainly, it may be necessary that the same table could not have come from a different piece of wood and perhaps that the same body could not have had a different origin. However, it is not so obviously necessary that our experiences (even our first ones) could not be associated with a different body, and hence it is not so obviously necessary that a person (in the sense at issue) could not have a different biological origin and/or a radically different future. Perhaps, our confusion about the self is the result of a reasonable inclination to think of bodies, like tables, in this way and an equivocal temptation to think of ourselves as bodies. Perhaps it is true that if a particular sperm had not fertilized a particular egg, then one would never have come into existence. However, this is only to say that the world is such that the same person could not have a different biological origin (just as, for instance, C-fibre stimulation could not result in a salty taste), that it is physically impossible to have a different biological origin. However, it does not follow that it is metaphysically impossible
for one to have a different biological origin, and it is this latter sense of possibility that is at issue. One notes that there is nothing inconsistent about Kripke's (perhaps quite plausible) general intuitions about the importance of origins with Cartesian dualism, which may be understood to imply the metaphysical impossibility of one's soul having a different origin.

I am inclined to agree with Kripke that "the fact that it is hard to imagine me coming from a sperm and egg different from my actual origins...[seems] to indicate that we have no clear conception of a soul or self." (Kripke, footnote 77, p.155) Most of us have confused and even self-contradictory views of ourselves. However, given this confusion, to suppose that any intuitive feeling that one could not have a different biological origin is at all decisive is to beg the question. Indeed, though this feeling and perhaps the prevalent scientific world view may tempt us to think of ourselves as bodies, given the case that our concept of ourselves is compatible with transfer and disembodied survival, I think that on reflection we are more strongly inclined to believe that we are not bodies (and/or brains).

6. Are Transfer and Disembodied Survival Metaphysically Possible?

Given the above modal argument (in Section 4) that a particular experience is not a particular brain state and hence that the experiences about which we should care specially can conceivably exist apart from the brain with which they are in fact associated, I think that there is a case that transfer and
disembodied survival are metaphysically possible. Even though the referential view may incline some to doubt that transfer and disembodied survival are conceivable (i.e. metaphysically possible), careful consideration of this view may just as reasonably seem to support the claim that these are conceivable. I offer the following argument, based on the referential view, in support of this metaphysical possibility.

Suppose that E is one’s present experience, that F is a future experience to be associated with one’s brain, and that any feelings of special concern for F that may accompany E are justified. As argued above (in Section 4), it is conceivable (in the sense of metaphysically possible) that any experience (and in particular F) could be associated with no brain at all or arguably with a different brain than that with which actually it is to be associated. That is:

(1) there are possible worlds in which F will be associated with no brain at all and possible worlds in which F will be associated with a numerically different brain than that with which actually it will be associated.

Such can be the case even if E is associated with the same brain with which actually it is associated.

However, one may doubt that there is any reason to believe that, in such possible worlds, F is such that feelings of special concern for F that may accompany E are justified. Because F could exist in a possible world without E having existed, being such that feelings of special concern that may accompany E for the experience in question are justified is not an essential property of F.
Nonetheless, it seems reasonable to suppose that, if one should be specially concerned about a future experience in the actual world, then one should be specially concerned about the same experience in any world in which one exists. It seems plausible that whether or not one who has E should be specially concerned about F depends entirely and directly on the experiences themselves. That is, in any possible world in which both E and F exist, any feelings of special concern for F that may accompany E are justified. Thus, it seems reasonable to accept:

(2) if the person who has E in this world is justifiably concerned about F, then the person who has E in any world in which F exists is justifiably concerned about F.

Though not everyone will accept this premise, it should make perfect sense to anyone who holds the commonsense view that ownership of an experience is essential to the experience, and that one’s ownership of a future experience justifies one’s special concern for this experience. (The reflective plain man’s view, articulated in Chapter 1, Section 2, accords with this view.) In particular, in this world, he who has E should be concerned for F because F will be his experience. If the ownership of an experience is essential, then E and F will have the same owner in any world in which both these experiences exist. Thus, in any such world, the one who has E is justified in being specially concerned about F. An experience about which one who has E is not justified in feeling specially concerned may be qualitatively indistinguishable from F but cannot be the numerically same experience as F.

From (1) and (2), it follows that there are possible worlds in
which one's feelings of special concern for a future experience may be justified even when this future experience is to be associated with a numerically different brain or with no brain at all. That is, one could survive as a sentient being to have experiences that are associated with a different brain or with no brain at all. Hence, in light of the referential theory of meaning, there is a case that transfer and disembodied survival are not only conceivable in the sense of being epistemically possible but also conceivable in the sense of being metaphysically possible.

Of course, physicalists may deny the premise that if the one who has E is justifiably concerned for F in this world, then this is so in any world in which both E and F exist. They may deny that the justification of this concern depends only on the essences of these experiences. Perhaps this justification depends on other aspects of E and F, such as the entities that have these experiences and how these entities are related. In particular, physicalists may insist that, for this concern to be justified, E and F must be associated with the same brain.

Though I do not deny that some relationship between E and F may be necessary for this concern to be justified, I do not see what grounds the physicalist could have to favour this sort of relationship. For instance, it may the case that (as the reflective plain man's view would have it), for this concern to be justified, E and F must be related in belonging to the same self. As shown in the previous section, we are certainly no more inclined to believe that this relationship should be physical than of some other
nature. Of course, the physicalist may consider it ontologically extravagant to postulate a self over and above the body. However, considerations of ontological economy come into play only when the theory with the sparser ontology can account for the relevant facts just as well as can the theory with the richer ontology.

As claimed above, the reflective plain man's view is able to explain how egoistic concern can be justified: i.e. He who has E should care about F because it will be his experience.\(^\text{12}\) Hence, to show that physicalism is preferable (on ontological grounds) to this view, the physicalist needs to explain, in terms of the association of E and F with the same brain, why the one who has E has reason to care specially about F. To insist that the needed relationship is physical, without accepting the need for such explanation, is simply to beg the question in favour of physicalism. The burden of explanation is thus on the physicalist.

I conclude that the above objection to my anti-physicalist argument based on the conceivability of transfer and disembodied survival can be reasonably met. Given the present facts, we have no compelling reason to accept Wedeking's account of the self.

One may object that (even if the notion of egoistic concern is intelligible) egoistic concern for future experiences might never be justified in the actual world or perhaps in any other possible

\(^{12}\) Because I too am troubled by the ontological commitments of this view, I certainly do not endorse this view. I consider it an important question what minimum (and hopefully more modest) ontological commitment is needed to explain how egoistic concern can be justified.
world. However, recall that I am not claiming that egoistic concern for future experiences is ever justified, only that physicalism cannot explain how it could be justified. To accept that egoistic concern is sometimes justified (say for future experiences to be associated with one’s same brain) is to saddle physicalism with this burden of explanation. As we will see in the next chapter, it is doubtful that physicalism can meet this burden.
1. Introduction

Despite any difficulties for the identity theory, a satisfactory explanation of how brain states can have a subjective aspect would perhaps convince me that experiences are brain states. Likewise, a satisfactory explanation of why we should care specially about our future experiences, in terms of their association with our brain, would perhaps convince me that we are brains. I concede that such an explanation would undermine my antiphysicalist arguments based on the supposed conceivability of transfer and disembodied survival. In this chapter, I argue on the basis of two seemingly reasonable claims encompassed by Butler’s principle that no such explanation can be given. I conclude that there is a strong case that our survival cannot be analyzed in physicalist terms.

2. Butler’s Principle and Physicalism

Recall, from Section 2 of Chapter 4, that what should be uncontroversial about Butler’s Principle is that if survival as a sentient being is to be analyzed in terms of a certain relation, then either it should be self-evident or we need to say why one should care specially about whether or not this relation obtains. This uncontroversial part encompasses Claims (1) and (2) that I have attributed to Butler. According to Claim (1), unless it is
self-evident that we should be specially concerned about the terms of a supposed analysis, whoever offers the analysis has the burden of explaining why the terms of the analysis should matter in this way. According to Claim (2), an explanation of why we should care specially about the terms of an analysis must be given in still other terms.

Though (1) seems to be an obvious requirement of rationality, some physicalists speak as if this requirement is unreasonable. For instance, though they do not seem to consider it self-evident that the association of a later experience with the same brain as an earlier experience (or more generally physical continuity between what realizes these experiences) justifies feelings of special concern for the later experience that may accompany the earlier experience, physicalists may deny that it is reasonable to expect a "deep reason" in terms of physical identity or continuity for such concern. Perhaps association of a future experience with the same body as one's present experience, though not a deep reason, is a reason for some concern. Physical continuity is all we have, and it must somehow be good enough. If the claim here is that egoistic concern is never justified, and that we have to settle for physical continuity (in terms of which we may perhaps explain how some other sort of concern is justified), then this claim may be defensible. The theory would have to identify this species of concern and how it could be justified in terms of physical continuity. (As I explain in Section 3 below, it is hard to imagine what this species could be.) However, if this claim is an attempt to somehow
accommodate the notion of egoistic concern within a physicalist world view by denying the need to account for how it may be justified, then to accept this claim is to renounce the need to explain and indeed to understand. To deny the need for such a "deep reason" is to deny the need to give any reason at all. Claim (1), which is central Butler's principle, is simply a reasonable demand for an explanation of what is not obvious.

I believe that this very reasonable claim, together with Claim (2), is enough to discredit physicalism. Recall that for one who has an earlier experience E to survive to have a later experience F is for feelings of special concern for F that may accompany E to be justified. If survival of one who has an earlier experience E to have a later experience F is to be analyzed in terms of the association of these experiences with the same brain, then either we need to say why this association justifies any special concern for F that may accompany E or it has to be self-evident that such is the case. Now, it may be that the regress ends at this point (and indeed never gets off the ground), that it is clear why the association of F with the same brain as E justifies feelings of special concern for F that may accompany E. However, this is certainly not self-evident, at least not to me. If survival is to be analyzed in terms of physical continuity, then we need to explain why this physical continuity justifies these feelings of special concern. We must explain why one should care specially about future experiences to be associated with one's same brain.

An explanation of why physical continuity justifies this
special concern cannot be given in terms of physical continuity. According to claim (2), we need to explain in other terms why this physical continuity justifies this concern. Thus, having exhausted the explanatory power of physical continuity, we must seek an explanation in other terms or deny that egoistic concern is ever justified. It seems that physicalism is unable to explain how egoistic concern may be justified and hence that there is no hope of analyzing survival in my sense in physicalist terms.

I believe that the fact that physicalism cannot explain how egoistic concern can be justified, together with the considerations that favour conceivability of transfer and disembodied survival and the difficulties for representative physicalist theories, provides a strong case against physicalism. It seems that if we accept a physicalist view of ourselves and our survival, then we must deny that special concern is ever justified. Thus, if we are to maintain a concept of ourselves as beings who should be specially concerned about certain future experiences, then it seems that we cannot be bodies or brains.

With no room left for an explanation in terms of physical continuity of how egoistic concern can be justified, we may appeal to a qualitative relation (i.e. a content-based relation or a relation of physical similarity between the brains and bodies with which the experiences are associated). If the justification of egoistic concern could be explained in terms of such a qualitative relation, then it would be this qualitative relation, rather than physical continuity, in which survival would consist. Physical
continuity would be important only in as much as it promotes such a qualitative relation. However, as I have argued in Part I, the justification of egoistic concern cannot be explained in such qualitative terms. The Butlerian regress does not end in such terms. If the justification of egoistic concern cannot be explained in terms of such a qualitative relation, then this egoistic concern cannot be explained in reductionist terms.

Because there is a strong case that no physicalist and no qualitative analysis of survival as a sentient being can be satisfactory, I conclude that survival as a sentient being cannot be analyzed in reductionist terms. Generally, reductionist theories are unable to explain how egoistic concern for future experiences is ever justified. One who accepts that such concern may be justified must reject reductionism.

3. As Good As It Gets

As I have argued in Section 6 of Chapter 4, in a world in which egoistic concern is never justified and hence no one ever survives in my sense, it would be a very reasonable convention to identify someone who is qualified qualitatively to fill a person's economic, social, and/or human niche as a survivor of this person. If the difficulties of accounting in reductionist terms for survival as a sentient being incline us to doubt that one can ever survive in my sense, then perhaps we should content ourselves to make sense of this conventional sense of our survival (what I will call the external sense of survival). At least, one's survival in
this sense may be significant to others.

As I have argued, physicalism is unable to analyze survival as a sentient being. However, it is clear that physical continuity has a role in securing one’s survival in this external sense of there being someone who continues to fill one’s niche. One’s physical continuity (perhaps the continued existence of one’s same brain) is the best guarantee that there will be someone who will have the qualitative features to continue to fill this niche. However, it is clear that this continuity is not that in which this survival consists. Physical continuity is important for survival in the external sense because it tends to preserve the qualitative relations (psychological and physical) in which survival in this sense does consist. It is these qualitative relations, not the physical continuity itself, that matter in this external sense. Clearly, in itself, physical continuity is not a good basis for convention regarding survival. If X resembles Y but is not physically continuous with Y, and X is physically continuous Z but does not resemble Z, then it makes more sense to treat Y than to treat Z as a survivor of X. Perhaps, of a number of equally qualitatively qualified candidates we may pick the one that is physically continuous with the original to be the same as the original, but this choice would be arbitrary. If egoistic concern is never justified, then we may think that the holding of qualitative relations that enable one to fill such a niche is “as good as it gets.”

However, physicalists may insist that physical continuity of
the brain and/or body is, instead, somehow "as good as it gets." In order to make sense of this claim, physicalists need to explain how physical continuity in itself can matter. If neither survival as a sentient being nor survival in the external sense consists in physical continuity, then it is hard to imagine how physical continuity can in itself really matter. Because physicalism cannot explain how egoistic concern can be justified, physicalism can offer no more reason to care about a future experience associated with one's same body than about a future experience associated with a numerically distinct body that enjoys the same qualitative relations to one's present experience. Any attachment that one may feel to one's own body in itself over a numerically different but qualitatively indistinguishable body is indeed purely sentimental. There seems to be no sense in which one might consider physical continuity as good as it gets. Unlike the qualitative view, physicalism does not seem able to provide an analysis of any interesting sense of survival. An appreciation of these difficulties should make physicalism much less attractive than many now find it.
CONCLUSION
1. Nonreductionism

Though it is probably not possible to refute skepticism about survival in the sense at issue, many of us are deeply committed to the belief that our feelings of special concern for future experiences are sometimes justified and hence that one may survive in this sense. Because, as I have argued, no reductionist theory can provide an analysis of survival in this sense, if we accept this belief, then we must take a nonreductionist view of survival. Recall that a nonreductionist theory is a theory according to which one's survival over time and any facts about one's survival cannot be fully described in terms of physical relations and/or qualitative relations. If egoistic concern is ever justified, then we must content ourselves to say that it is an unanalyzable fact that such concern is sometimes justified. Butler seems to have thought that one's identity is somehow unanalyzable or primitive in this sense. Feelings of special concern for F that may accompany E are justified because there is an unanalyzable relation between E and F, that is a relation that is neither qualitative nor physical (i.e. does not depend on physical continuity of the brains/bodies with which the experiences are associated). (This is not necessarily to deny that this relation between E and F consists in E and F belonging to the same immaterial subject of experience.) One should care specially about future experiences to be associated...
with one's brain not because they will be associated with one's brain or because they may be appropriate sequels to one's present experience but because they will bear this relation to one's present experience.

In this final chapter, I broadly enumerate the various nonreductionist alternatives. Certainly, there is a wide range of possible nonreductionist theories. Because I consider theories that do not accord one's brain some role (whether direct or indirect) in producing experiences about which one should be specially concerned to be highly implausible, I will consider only theories (interactionist theories) according to which the brain does have some such role.

There are two main types of nonreductionist theories: pure ego theories and what I will call contingent physical continuity theories. According to a pure ego theory, the justification for one's special concern for a future experience is that the future experience belongs to one's ego. According to a contingent physical continuity theory, it is a contingent fact (one true at least in the actual world) that X survives as Y just in case there is a certain physical continuity between the brains and/or bodies of X and Y. However, X's survival as Y does not consist in this physical continuity. On both of these types of theories, it is the continued existence of a certain entity (a metaphysical one in the case of

---

1 The term "pure ego theory" was popularized by H.P. Grice. (Grice, 1942, p.76) However, as Grice notes, this term was probably introduced by C.D. Broad.
pure ego theories, and a physical one in the case of contingent physical continuity theories) that is responsible for one's survival.

Clearly, because a nonreductionist theory denies that one's survival consists in physical and/or qualitative relations, any such theory will have ontological commitments over and above physical entities and single experiences. Such a theory may have a commitment to a persistent entity to which the experiences involved belong or to a transtemporal relation between these experiences (such as the unity relation of a transtemporal unity encompassing these experiences). It is meaningful to assert that one survives in the sense at issue, and we are unable to do so without the use of terms that refer to such things. At least one such term must be treated as primitive and its referent included in a complete ontology. For instance, a term that refers to the owner of the experiences (as does "ego") may serve as such a term.

2. Pure Ego Theories - The Substance View

A pure ego theory identifies one with an entirely and irreducibly immaterial entity, one's ego or soul. The relation between temporally separate experiences that justifies special concern consists in the experiences belonging to the same ego. According to such theories as they are normally understood, we are (in Parfit's words) "separately existing entities." (Parfit, 1984, p.210) Not only cannot one's ego be identified with or reduced to the states and activities of one's brain, but the ego does not even
depend on the brain for its existence. That is, an ego is actually capable of existing apart from the brain. However, this is not to deny that one's brain has a role, at least during one's biological lifetime, in producing one's experience. Parfit calls this view "the Cartesian View." (Parfit, 1984, p.223)

Because of the strong empirical evidence that the brain has an important role in determining one's character traits, preserving one's memories, and shaping one's thoughts, it makes sense to deny that one's distinctive psychology (including one's memories, intentions, beliefs, and character traits) inheres in one's ego. One's ego may serve only as owner of one's experiences (and thus to explain why one should care specially about certain future experiences). On such a theory, there may even cease on the death of the brain to be experiences about which one should be specially concerned, even if the ego should continue to exist. That is, one may survive in the sense at issue only so long as one's brain survives and supports consciousness. Richard Swinburne proposes a neo-Cartesian theory along these lines. (See Shoemaker and Swinburne, 1984, pp.22-34, especially p.33. Swinburne also suggests an alternative view that the ego ceases to exist when the brain irrevocably ceases to support experience but that "there remains the logical possibility" of its coming into existence again.)

Though pure ego theories may have the advantage of conforming to some of our deep beliefs (prejudices reductionists might think) about ourselves and our survival, there are serious problems with these theories. Though one has access to one's experiences, one has
no access to any such mental substance. Moreover, the nature of such substance is utterly obscure. Also, for an ego theory to explain why one should care specially about certain future experiences to be associated with one's brain is theoretically convoluted. One should care about the future experiences because the ego will continue to associate with the brain and the ego owns (i.e. is the bearer of) the future experiences in question. Moreover, this labyrinthine explanation is plagued with two major obscurities. It is totally mysterious how an ego associates (and interacts) with a brain, much more so than how a brain can directly cause an experience. It is also unclear how an ego can own an experience. One may also be inclined to agree with Locke that the divergence of one's consciousness (and the experiences that might justifiably be the object of one's special concern) from one's mental substance (i.e. one's ego or soul) is just as conceivable as its divergence from one's physical substance (i.e. one's brain and/or body). (Locke, 1975, II, xxvii, 13) If Locke is correct, then pure ego theories are unable to account for one's survival.

Generally, the unanalyzable and immaterial entities whose existence pure ego theories posit conflict with the largely scientific world view that is now widely accepted. Not surprisingly, most modern philosophers are reluctant to include egos in their basic ontology and tend to dismiss such theories.
3. Contingent Physical Continuity Theories

On a contingent physical continuity theory, it is a contingent fact (i.e. one true in the actual world) that one survives just so long as one's brain and/or body enjoys sufficient physical continuity. Physical continuity can include continuity of matter and also activity (such as continuous support of consciousness etc.). It is reasonable to suppose that experiences result from brain states. Perhaps, future experiences about which one is justified in feeling specially concerned result from brain states that are physically continuous with the brain states that support one's present experiences. Even though one's survival does not consist in this physical continuity, this physical continuity may provide (i.e. be causally responsible for) one's survival.

Because on a contingent physical continuity theory, as a matter of fact, no nonphysical entity exists separately from the brain and/or body, such a theory does not count as a substance view. Such theories seem to be variants of what Parfit calls the "further fact view," a view that

...denies that we are separately existing entities, distinct from our brains and bodies, and our experiences...[but]...claims that personal identity [or survival] is a further fact, which does not just consist in physical and/or psychological continuity. (Parfit, 1984, p.210)

Contingent physical continuity theories recognize that one's survival does not just consist in the continued existence and

---

2 On an initially attractive theory of this sort, it is a contingent fact that X survives as Y just in case Y has the numerically same brain as X.
capabilities of one’s brain. Because such theories do not claim to be able to analyze or describe survival in terms of physical and/or qualitative continuities, they are nonreductionist theories. The appeal of contingent physical continuity theories is obvious. They seem to be compatible with our deep beliefs and intuitions about survival, and indeed to account for what many of us consider obvious facts of our existence, and yet do not seem to conflict as seriously as do pure ego theories with an objective and scientific world view.

Of course, if survival does not consist in physical continuity, then we need to specify in exactly what survival does consist. The obvious suggestion is that one’s survival consists in the continuation of one’s stream of consciousness. (One’s stream of consciousness is the total temporally-ordered and unified succession of one’s experiences. I think most of us have a clear appreciation of the nature of a stream of consciousness.) That is, to say that special concern for a later experience F that may accompany an earlier experience E is justified is just to say that E and F belong to the same stream of consciousness. It is because of one’s privileged access to one’s present experience that one should be specially concerned about its contents. Likewise, membership of future experiences in the same stream of consciousness as one’s present experience provides a privileged (though of course delayed) access to these experiences that justifies one’s special concern for these experiences. The relation between an earlier experience E and a later experience F that
justifies feelings of special concern \( F \) that may accompany \( E \) is the bare psychological relation of their belonging to the same stream of consciousness (i.e. the transtemporal unity relation of a stream of consciousness). Because survival cannot be analyzed in reductionist terms, this relation does not consist in a physical relation (such as the association of the experiences with the same brain) or in a content-based (more generally, qualitative) relation. The unity relation between an earlier experiences \( E \) and a later experience \( F \) (and hence survival of one who has \( E \) to have a \( F \)) consists in a content-independent psychological relation between \( E \) and \( F \).

Because this transtemporal unity relation is irreducible, the theory should treat the stream of consciousness as a metaphysical primitive and thus include streams of consciousness in its basic ontology. Nonetheless, it is reasonable to suppose that the brain is causally responsible for experiences and for their belonging to the same stream of consciousness. Physical continuity of the brain with which \( E \) and \( F \) are associated is important only in as much as it helps to secure this relation between \( E \) and \( F \). I will call this view that survival consists in the continuation of the stream of consciousness but is provided in this way by physical continuity "the contingent physical continuity view."\(^3\)

There are some clear advantages to this view. First, though

---

\(^3\) R. Sikora develops a theory along these lines in his paper. Sikora’s work has inspired this whole discussion of contingent physical continuity theories.
streams of consciousness are not mere physical processes, they are things that we naturally suppose to exist. Indeed, one has quite direct access to one's stream of consciousness. Unlike pure ego theories, this view does not invoke some obscure, ghostly and separately existing metaphysical entity to account for one's survival. Second, there is an obvious way in which a stream of consciousness can own an experience (i.e. simply by containing the experience). Third, the above Lockean objection to the substance view does not apply to this theory. The divergence of one’s consciousness from one’s stream of consciousness seems clearly inconceivable. Fourth, if we include streams of consciousness in our basic ontology, then there is no need for an additional ontological commitment to experiences. An experience is just a brief segment of a stream of consciousness. Fifth, this view is compatible with skepticism about our survival. If, as the skeptic may think possible, streams of consciousness never have an extended transtemporal unity, experiences would still count as short streams of consciousness. This view is also compatible with the possibility that we do not survive apparent gaps in consciousness.

Clearly, there remains the serious objection to the contingent physical continuity view that the nature of the interaction between the stream of consciousness and the brain is mysterious. However, likewise, it is mysterious how a physical brain state can cause an experience. Though the problem of explaining this interaction is certainly formidable, hopefully it is not as insuperable as the difficulties of explaining how a brain can interact with an ego and
how an ego can own an experience.

4. Problems for the Contingent Physical Continuity View

Though the view that survival consists irreducibly in the continuation of the stream of consciousness which cannot as a matter of fact exist apart from the brain has certain obvious attractions, some may feel that this view suffers from some equally obvious difficulties. Though a treatment of these difficulties is beyond the scope of this thesis, it is useful to briefly enumerate these difficulties and sketch how these might be resolved.

First, some philosophers and scientists have argued on the basis of empirical evidence that the experience associated with the brain at a single time may be disunified or, more problematically, only partially unified. Clearly, if there are doubts about the unity of an experience at a time, then there should be even more serious doubts about the transtemporal unity of experience. However, not only is it doubtful that the notion of such a partial unity is even coherent but also there is serious controversy about whether the scientific research supports these conclusions. (See Hurley, 1994, Lockwood, 1989, 1994, Marcel, 1993, 1994, for various perspectives on this issue.)

Second, some may think that this view suffers from two (possibly related) metaphysical difficulties. If streams of consciousness are not individuated in terms of the contents of their component experiences or in terms of physical continuity, then they seem to have no criterion of individuation. To answer
this objection, one needs to defend haecceitism with regard to streams of consciousness (the view that streams of consciousness can have a primitive identity). (See Adams, 1979, for a defence of haecceitism in general.) One may also object that streams of consciousness must be either properties of brains (and hence cannot conceivably exist separately from the brain) or separately existing entities (i.e. substances). Hence, one needs to find an appropriate metaphysical status for streams of consciousness.

Third, apparent gaps in consciousness seem to be a serious problem for this view. Though we can doubt that we survive apparent gaps, it may count seriously against a theory of survival if it cannot explain how we might survive apparent gaps. Certainly, one rationale for positing the existence of egos (as owners of experiences) is to explain how one can survive such gaps. One possible solution to the problem of gaps is to suppose that they are merely apparent. As Unger notes, Descartes held that while apparently unconscious, one is in fact dreaming continuously. (Unger, 1990, pp.15, 45) However, this suggestion does not explain why when one is awakened suddenly from a deep sleep (during which one does not exhibit rapid eye movement), one does not remember dreaming. In an unpublished thesis, Wayne Henry suggests that during an apparent gap in consciousness, even if one does not have structured dreams that one could possibly remember, one may still

---

4 See Descartes, 1976c, pp.278-280. Note also that Locke took it to be a clear commitment of Descartes' view of thinking as an essential property of the mind that we must be conscious continuously. (Locke, 1975, II, i, 9)
have a very minimal level of consciousness. (Henry, 1987)

5. Conclusion

I leave the reader with three alternatives to reductionism. One alternative is the skeptical view that one never survives over time in my sense. Because we cannot be sure that egoistic concern is ever justified, skepticism about survival in my sense should be considered viable. Those who reject reductionism but believe that one may survive in the sense at issue may choose between two nonreductionist views: the substance view according to which survival consists in the persistence of a pure ego and the view that survival though metaphysically primitive is likely causally dependent on physical continuity. Both of these views have ontological consequences. One can deny these consequences only if one is prepared to deny that egoistic concern is ever justified.
Bibliography


