To my parents, Ron and Barbara Franke, my grandparents, Bill and Olga Jerman,
my brother, Jerry Jeske, and the memory of Sam.
# TABLE OF CONTENTS

**Abstract**  
4

**Acknowledgements**  
6

**Chapter 1: Introduction**  
7

**Chapter 2: Non-Reductionism and the Reductionist Alternatives**  
19  
Section 1: Are We Separately Existing Entities?  
19  
Section 2: The Reductionist Alternatives  
50

**Chapter 3: Identity and What Matters**  
76  
Section 3: Can Persons Switch Bodies?  
76  
Section 4: Should We Care If We Get New Bodies?  
110

**Chapter 4: The Grounds for Special Concern**  
144  
Section 5: Introduction  
144  
Section 6: The Problem of Special Concern  
146  
Section 7: An Impersonalist Account of Special Concern for One’s Future Self  
156  
Section 8: Extending the Impersonalist Account Into the Interpersonal Case  
160  
Section 9: Alternatives to the Impersonalist Account  
169  
Section 10: The Rejection of the Virtue-Promoting View in the Intrapersonal Context  
174  
Section 11: The Rejection of the Virtue-Promoting View in the Interpersonal Context  
182  
Section 12: Self-Referential Altruism and Self-Concern  
190

**Chapter 5: Persons, Compensation, and Utilitarianism**  
196  
Section 13: Introduction  
196  
Section 14: Utilitarianism and the Separateness of Persons  
198  
Section 15: Reductionism and the Separateness of Persons  
206  
Section 16: Psychological Reductionism and Compensation  
213  
Section 17: Identity and Compensation  
227

**Works cited**  
230
The Persistence and Importance of Persons

by

Diane Jeske

Submitted to the Department of Linguistics and Philosophy on July 22, 1992 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Philosophy

Abstract

In this thesis I will defend a Reductionist criterion of personal identity, and show that that criterion supports certain commonsense claims about the morality and rationality of special concern and about the morality of the distribution of goods among persons.

In Chapter 1 (Introduction), I will introduce the problem of personal identity across time. Persons seem to persist through certain changes, but we do not think that they could persist through any and all changes. Thus, by examining a range of both ordinary and science fiction cases, we can try to determine the necessary and sufficient conditions of personal identity. We will see that a plausible theory of personal identity supports and motivates the commonsense claim that we should be specially concerned about our future selves and intimates.

In Chapter 2 (Non-Reductionism and the Reductionist Alternatives), I will argue that we should reject a Non-Reductionist theory of personal identity, where Non-Reductionism is the view that persons are separately existing entities. Non-Reductionism is unable to provide a coherent account of why persons persist through certain changes and not others, thereby committing us to a general skepticism about personal identity. Therefore, we should conclude that persons are not separately existing entities. I then describe the principal forms of Reductionism, and thereby motivate the claim that we should not accept a purely physical account of personal identity.

In Chapter 3 (Identity and What Matters), I discuss the question whether persons can get new bodies. I present certain cases that suggest that we should reject our initial intuition that persons can swap bodies. I then consider whether we should accept the claim that it is not our identity that matters in our survival, and suggest that Reductionists can continue to claim that identity matters.

In Chapter 4 (The Grounds for Special Concern), I defend the commonsense view of special concern. Commonsense holds that my special concern for my future self and intimates is justified by the fact that my future self and intimates stand in special relationships to me. I claim that this commonsense view can be motivated and supported by the plausible Reductionist criteria discussed in Chapter 3.
show that the commonsense conception is more plausible than the Impersonalist view that tries to ground special concern in terms of impersonal values.

In Chapter 5 (Persons, Compensation, and Utilitarianism), I discuss Parfit’s claim that if we were to deny that identity is what matters and accept that psychological continuity is what fundamentally matters in our survival, then we should be more inclined to find Utilitarianism plausible. I show that Parfit’s view of personal identity and what matters actually supports a moral theory in which the relationships in which persons stand to one another have intrinsic significance, rather than supporting the person-neutral theory of Utilitarianism.

Thesis Supervisor: David Brink

Title: Associate Professor of Philosophy
ACKNOWLEDGEMENTS

Many people deserve my gratitude for help that they have given to me on this thesis. One person, however, stands out as having given an extraordinary amount of time and effort, and that is David Brink, my thesis supervisor. David introduced me to the topics with which the thesis is concerned, and it was in large part his careful approach to and enthusiasm for these topics that generated my own interest. David has provided extensive and detailed comments on every draft of every chapter of the thesis, comments that have forced me to constantly rework and restate my positions. Most importantly, he has always taken my work seriously by taking the time to provide responses on behalf of my opponents (among whom he was often included) and by suggesting ways for me to defend my own positions. Any virtues the thesis may have are due in large part to his constant help and incisive, challenging comments. Also, David has given support and encouragement at times when I was most in need of them. All in all, David has provided a model for me of what a mentor and philosopher can and ought to be.

The other members of my thesis committee, Josh Cohen and Bob Stalnaker, also deserve special thanks. Josh helped me to see issues that I had not seen before, and, in several places, he showed me that I had ignored important alternative positions and lines of argument. His comments helped to improve substantially the final draft. Bob agreed to join my committee at the last moment and I greatly appreciate his reading the thesis and providing interesting and provocative comments at the defense.

I would also like to thank Judy Thomson. She provided valuable assistance by commenting on early drafts and helping me to get the thesis off the ground.

Many friends and colleagues in the MIT philosophy department provided support and advice, both philosophical and emotional, without which the last four years would have been for less interesting and bearable. For that, I thank Lenny Clapp, Tim Dilligan, Tim Hinton, David Hunter, Tracy Isaacs, Rob Stainton, Jason Stanley, Ed Stein, and Daniel Stoljar. Special thanks go to Tim Dilligan for his valuable help in getting the thesis formatted correctly. And, of course, I thank my family to whom this thesis is dedicated. Without their love and support, none of the following would have been possible.
Chapter 1
Introduction

In The Metamorphosis Kafka tells us that Gregor Samsa woke up one morning to find that he had turned into a giant beetle. Gregor had gone to bed the previous evening with the body of a man, but in the morning he has a new body, the body of a bug. This bug that his family now sees remembers Gregor’s life and has thoughts and beliefs similar to those that Gregor had the previous evening. As we read Kafka’s story, we sympathize with the plight of a man who has become entrapped in such a horrid body.

Of course, I feel very confident that, given the laws of physics and human physiology, I need not worry about waking up in the body of a beetle. Nonetheless, Kafka’s story seems entirely coherent. Initially, it seems possible in some sense for a person to undergo the terrible changes that poor Gregor undergoes. It does not seem that the opening statement of The Metamorphosis is a contradiction. So a person seems to be able to undergo immense physical changes and still remain the same person.

Suppose, however, that we alter Kafka’s story somewhat. Later in the day on which Gregor’s family finds the beetle in his bed, Gregor’s body comes back home. The man who occupies this body remembers Gregor’s life and has thoughts and beliefs similar to those that Gregor had the previous evening. His family is naturally overjoyed that Gregor has come home, that he has not turned into a
beetle after all. The reader will wonder how that beetle came to be in Gregor’s bed, and, more importantly, how that beetle came to have Gregor’s memories and character. There seems to be no question whatsoever that that beetle is not Gregor.

Perhaps, then, we should not accept Kafka’s description of the events that transpired at the Samsa home on that fateful morning in *The Metamorphosis*. Perhaps Kafka should have told us that, one morning, a giant beetle woke up in Gregor’s bed and, oddly enough, that beetle was under the false impression that it was Gregor. After all, this beetle has no more claim to being Gregor than does the beetle in my variation on Kafka’s story. How wrong we were to suppose that a person could come to have the body of a beetle. Kafka’s story is really about a beetle that mysteriously comes to have a particular person’s memories and beliefs.

Maybe it is mistaken to think that a person could possibly acquire the body of a beetle. But let us change Kafka’s story one more time. Suppose that the person who wakes up in Gregor’s bed one morning has Gregor’s memories and beliefs but Winston Churchill’s body. Similarly, the person who wakes up in Winston Churchill’s bed has Churchill’s memories and beliefs and Gregor’s body. We will probably be inclined to think that Gregor and Churchill have exchanged bodies. If, for some reason, we do not want to say that, our next best option seems to be to say that Gregor has somehow become just like Churchill and *vice versa*. Either the persons involved have undergone immense physical change or they have
undergone immense psychological change.

Of course, persons never undergo these sorts of changes. But, when we try to find the correct criterion of personal identity, we are trying to find the necessary and sufficient conditions for being the same person across time. So we need to decide whether persons could *possibly* undergo the sorts of changes that I have described. Is a person the sort of entity that could acquire a new body? Is a person the sort of entity that could acquire an entirely new character and personality? We may wonder why we should be concerned about these questions given that persons never do actually acquire new bodies or entirely new characters (at least not over night). But persons do fall into comas in which most of their mental lives cease, and persons do suffer total amnesia. Can persons survive in these cases? In these cases some element of ordinary survival is missing. Bizarre science fiction examples can help us to see what is a necessary condition of survival rather than only contingently an element of ordinary survival.

Moreover, even if persons do not get new bodies or completely new personalities over night, they do undergo drastic physical and psychological changes. I am very different now than I was at six and I will change even more by the time I am seventy. Nonetheless, it does seem that I was once a girl of six and that I will (hopefully) be a woman of seventy. My body can double in size and I will still survive, but if my brain or heart stops functioning I will not survive. Why is it that I survive certain changes and not others? How much change can I undergo
and remain the same person?

We will evaluate various criteria of identity by seeing how they deal with both ordinary change and bizarre science fiction sorts of change. If a criterion is unable to accommodate the persistence of persons through ordinary sorts of change, we will have some reason to reject that criterion. Similarly, a criterion should be able to accommodate our strongest intuitions about the science fiction cases. But, if we are to rely on our intuitions about bizarre cases, we need to look at a wide range of cases in order to try to eliminate inconsistencies. Many of our intuitions may be clouded by an extrapolation from the ordinary run of cases. We want to be careful not to mistake what we take to be good evidence of personal identity in ordinary cases for the actual necessary and sufficient conditions of personal identity. In other words, we do not want to mistake our epistemological criteria of identity for a metaphysical criterion of identity, where an epistemological criterion will tell us not necessary and sufficient conditions of identity, but, rather, what can reliably function as our evidence of personal identity. An adequate theory of personal identity will provide an account of how our epistemological criteria are related to the metaphysical criterion, but, nonetheless, some considerations may be relevant to determining one but not the other (see Chapter 3).

Let us return to the first case of Gregor. As I said, we do initially have the intuition that Gregor now has the body of a beetle. Most of us probably would accept, at least initially, a Non-Reductionist explanation of our intuition. A
Non-Reductionist about persons would say that somehow Gregor's soul (or spirit, mind, or essence) has managed to leave Gregor's body and now inhabits the body of a beetle. Gregor, of course, goes where his soul goes. The evidence that we have for the migration of Gregor's soul is the fact that the beetle now has Gregor's memory and character. At least some Non-Reductionists, however, would hold that the soul could become dissociated from a person's character and memories. Hindus, for example, believe that persons are routinely reborn as beetles and other insects, but they do not believe that beetles remember their past lives.

But what is the relationship between a person's psychology and his soul? Can a person's soul endure any and all psychological and physical change? In Chapter 2 I will argue that we should reject Non-Reductionism about personal identity because it fails to provide an adequate answer to these questions. It provides no explanation either of how persons can survive certain changes and not others or of how our psychological and physical evidence of identity is related to the metaphysical conditions of a person's persistence. Reductionist views, on the other hand, do provide such an account because they claim that one or more of the elements of our survival that function as our evidence of identity are the necessary and sufficient conditions of personal identity; Reductionists hold that continuity of psychology and/or continuity of the body constitute the identity of persons across time. In Chapter 3, however, we will see that it is very difficult to arrive at any Reductionist view that accommodates all of our strongest intuitions. Part of the
difficulty is that many of our intuitions are affected by one particular extrapolation from the ordinary cases; we continue to assume that it is rational for a person to care about her own survival.

For consider a case where I am offered a choice: I can die tomorrow or I can choose to live for another fifty years. Of course, there will be situations where I will seem justified in choosing to die tomorrow. I would rather die tomorrow than live for fifty years in constant excruciating pain. I would rather die tomorrow than live for fifty years with a homicidal pathological disorder. But, all else being equal, I am justified in caring that I survive. Notice, however, that in ordinary cases, once I cease to exist, my mental life does not continue on. Suppose, however, that my body and brain were destroyed, but that my mental life were to continue on in a replica of me. This is certainly not like ordinary death. My replica will finish my thesis and support my family, things that would not get done if I died and my mental life ceased. Is my replica me? We may be inclined to think so, because I should not dread replication as I would dread getting hit by a bus. But maybe my replica is not me, and replication is a case in which I cannot care that I do not survive.

I think that most of us have an extremely strong intuition that I would survive as the person with my body in a case where I was replicated without having my brain and body destroyed. Moreover, we also seem to have a strong intuition that it would be rational for me to care that the replica die tomorrow and that the
person with my body live for another fifty years rather than *vice versa*. But why should I have that preference? Can I really care about this particular body? If not, why care more about the person with my body? If she is me, can the mere fact that I am identical with her matter? Perhaps my intuition that she is me is even just an extrapolation from ordinary cases of survival. These are questions that I will address in Chapter 3. Although I think that I will survive as the person with my body and that, therefore, I am justified in caring more about her than about the replica, I think that this is a very difficult conclusion for which to offer any conclusive arguments. Why does identity matter if one component of identity, continuity of body, does not seem particularly important? If identity just consists in certain physical and/or psychological continuities, and those continuities seem unable to support the sort of concern that we previously had for identity, we might conclude that we should cease to be concerned about identity. My concern for my future selves, then, should be the same sort of concern that it is appropriate for me to have for any future person with whom I am psychologically continuous. Or a more extreme view would hold that concern for future selves is unjustified.

We might, however, continue to claim that identity matters and claim that our concern with identity is to be explained in terms of something other than the sorts of continuities with which the Reductionist analyzes personal identity. In other words, it is possible that concern about identity is not reducible to or analyzable in terms of the constitutive elements of identity yet is still justifiable and rational.
I think that we can explain why identity matters in terms of the sorts of projects around which we structure our lives. I have a great many private projects, projects such that if they are to be realized I must survive and realize them. For example, I want that I finish my thesis; for this project to be realized, I must survive and finish my thesis. These sorts of projects actually form the core of our relationships to our future selves. Thus, a certain form of concern is actually partly constitutive of my relationship to my future self. We can notice, also, that many of my projects involve certain other persons, namely my intimates, my friends and family members; for example, I want to take care of my child. These projects form the core of my relationship to my child, so a certain form of concern is also partly constitutive of my relationship to my child. So in Chapter 4 I suggest that special concern can be grounded by the special relationships in which we stand to certain persons, and that such an account is much more plausible than an account that attempts to ground special concern on impersonal values such as the promotion of virtue.

My account of special concern leaves several questions unanswered. It provides an account of how to justify special concern for certain persons to whom we already stand in certain special relationships. Thus, unlike the account that grounds special concern in terms of impersonal values, it provides no answer to the question about why we should form such special relationships in the first place. Why should I stand in this sort of relationship rather than some weaker sort of
relationship to my future self? Why form friendships with other persons? If we need to appeal here to an account of value to explain why these relationships should matter, is it plausible to suppose that the account of special concern is not value-dependent? Given the implausible results yielded by the value-dependent account of special concern, I think that we have sufficient motivation for accepting that our reasons for forming friendships are different in kind from our reasons for being specially concerned about those persons with whom we are already friends. We can hope that the account of why we should form special relationships will offer further support for the claim that, once formed, those relationships provide intrinsic reasons for concern. Because special concern is partly constitutive of special relationships, an account of why we should enter such relationships will also be an account of why we should develop special concern.

If we do accept the account of special concern that I offer, there are still questions within the account that need to be addressed. First, as I point out at the end of Chapter 4, my account seems to lead to the conclusion that my duties to myself are stronger than my duties to any of my intimates, even my children. Thus, it seems that it is wrong of me to bear a certain burden rather than having an equal burden imposed on my child. This is a very unattractive conclusion, but it is one that I do not have the resources to deal with here. It may be that we can resolve the difficulty if we give an account of what form special concern should take in various sorts of relationships. In so doing we will be giving a partial
characterization of the very nature of the relationships, because these forms of concern are partially constitutive of special relationships. In characterizing what I owe my friend and how that should be balanced against what I owe myself, we will be characterizing friendship itself. This issue, however, is beyond the scope of the present project.

In Chapter 4 I am concerned primarily with special concern for future selves and intimates. But I will indicate that we also stand in special relationships to other persons, in particular, to our fellow countrymen. If these special relationships can provide an account of familial duty and of friendship, perhaps they can also provide an account of political duty and obligation. If so, the models of family and friendship may provide interesting parallels to the case of political community. Conversely, the case of political community may help us to understand better the nature of family and friendship. What we take as given in one case may help us to resolve unclear issues in another case. Again, however, this is an issue that I leave for another time.

My account of special concern is motivated by the claim that certain projects that are part of our relationships to our future selves constitute a basic form of concern for our future selves. Derek Parfit has argued that we should abandon those private projects. Although I reject his argument for that claim in Chapter 3, I think that it is important to see what implications such a view might have. Parfit thinks that if we hold, as Reductionists, that our identity is not what
matters while psychological continuity is what matters, we should find
Utilitarianism an attractive moral theory. I will argue that Parfit’s view in fact
supports a moral theory in which the relationships in which persons stand to one
another are significant *in and of themselves*. We should accept that special
relationships play an important role in determining duties and obligations whether
or not we accept that identity matters.

The chapters that follow, then, are an attempt to show that the most plausible
Reductionist criteria of personal identity support and motivate many
commonsense claims about morality, in particular that our strongest duties are to
the persons with whom we interact most often. This may be surprising, because
most of us, before we begin to philosophize, probably accept some version of
Non-Reductionism, the view that I reject in Chapter 2. If the correct account of
personal identity is closely linked to commonsense claims about morality, how is it
that most of us accept that commonsense morality but not Reductionism? This
question raises interesting issues about the exact relationship between normative
issues and metaphysical ones, and about how a theory about one set of issues can
function as justification for a theory about the other set of issues. Perhaps we
should take it as further evidence for the theory of personal identity that I defend
that it is able to offer support for certain commonsense moral claims. After all,
our understanding of the nature of persons derives not only from metaphysical
conceptions but also from certain normative conceptions. Therefore, we should not
be quick to think that all of our claims about the importance of persons must be firmly grounded on claims about what constitutes the persistence of persons.

Claims about the rational and moral importance of persons may be prior to or independent of any metaphysics. Nonetheless, the fact that a certain metaphysical theory of personal identity can support those claims shows that we can have a unified theory of the persistence and the importance of persons.
Chapter 2
Non-Reductionism and the Reductionist Alternatives

1 Are We Separately Existing Entities?

1.1 Introduction

Yesterday I decided to write a paper on personal identity. Today I am starting to write a paper on personal identity. The person who made the decision to write the paper and the person who is now writing are the same person. In other words, the person who made the decision is identical to the person who is now writing. What does it mean to say that the former person and the latter person are identical? In virtue of what facts is it true both that I decided to write and that I am now writing? In this chapter I want to consider how a Non-Reductionist about personal identity would answer these questions. As I said in Chapter 1, most of us accept, implicitly, some version of Non-Reductionism.

All Non-Reductionists agree that the facts that constitute a person's identity across time are, in some way, independent of the observable changes in a person's body and psychology. Most Non-Reductionists hold that a person's identity across time is determined by the existence of a separately existing immaterial substance that underlies those observable changes. So the Non-Reductionist seems to be able
to easily accommodate the persistence of persons through change. Nonetheless, we
will see that the Non-Reductionist has no adequate account of why persons persist
through certain changes and not others, and that, without such an account, the
Non-Reductionist is committed to a general scepticism concerning our judgments
of personal identity. Such considerations should motivate us to try to find a
plausible Reductionist view of personal identity. So after a discussion of
Non-Reductionism, I will describe various forms of Reductionism and motivate the
claim that we should accept some version of Psychological Reductionism rather
than a purely Physical Reductionist view.

1.2 The Distinction Between Reductionism and Non-Reductionism

Questions of the form “What makes it the case that person x is the same person as
person y?” can be answered in two different ways: one can respond as a
Reductionist or one can respond as a Non-Reductionist.¹ While a Reductionist
“claim[s] ... that the fact of a person’s identity over time just consists in the
holding of certain more particular facts” (Parfit 210;²), Non-Reductionists, on the
other hand, claim that personal identity involves some ‘further fact’. What is a
further fact, and how does such a fact differ from more particular facts? In order

¹There may, of course, be other ways of answering this question, but I think that all of the cur-
rently or historically prevalent views can be regarded as either Reductionist or Non-Reductionist.
²All references to Parfit are to Parfit 1984, unless otherwise indicated.
to answer this question, we need to consider the principal forms of Reductionism and Non-Reductionism.

**Reductionism** Reductionists differ over whether personal identity consists in certain psychological facts or whether it consists in certain physical facts. A Psychological Reductionist analyzes personal identity in terms of psychological phenomena and their interrelationships. Psychological phenomena include psychological events or states such as memories, experiences, beliefs, values, and intentions. These phenomena can stand in certain relations to one another, in particular, relations of similarity or causal relations. For example, when I vote for candidate x, I still have the belief that he is the best person for the job, a belief that I formed at some earlier time. My belief that candidate x is the best person for the job may lead me to form an intention to vote for candidate x, and my experience of seeing candidate x speak is the cause of my memory of seeing candidate x speak. When a ‘sufficient’ number of the mental events or states of person x and the mental events or states of person y stand in these relations to one another, we say that x is *psychologically connected* to y. If x is psychologically connected to y, and y is psychologically connected to z, then x is *psychologically continuous* with z. The Psychological Reductionist claims that person x is *identical with person y if and only if x and y are psychologically continuous*. (This criterion will be developed and modified in section 2.2.)

The Physical Reductionist, on the other hand, analyzes personal identity in
terms of physical or bodily phenomena and their interrelationships. The cells of a human body are constantly being lost and replaced. Nonetheless, this process is gradual. If a person lost the vast majority of his cells simultaneously, then, according to a Physical Reductionist, he would not survive. In order to survive, a person’s body must trace a continuous path through space and time, and must have its matter replaced gradually. In other words, his body must be both spatio-temporally and materially continuous. If person x’s body is both spatio-temporally and materially continuous with person y’s body, then we will say that person x is physically continuous with person y. The Physical Reductionist claims that person x is identical with person y if and only if persons x and y are physically continuous. (This criterion will be developed in section 2.1.)

I have presented only a rough sketch of the main forms of Reductionism. There are many variations on Psychological Reductionism and Physical Reductionism. For example, certain versions of Psychological Reductionism have clauses specifying how psychological continuity must be caused in order for it to be sufficient for identity. Certain variations on Physical Reductionism specify that x and y need have only physically continuous brains, not physically continuous bodies, in order to be identical. And, of course, one may combine Psychological Reductionism and Physical Reductionism, thereby claiming that physical and psychological continuity are both necessary and only jointly sufficient conditions of personal identity. But all of these various forms of Reductionism analyze personal
identity in terms of the same sorts of facts that are relied upon by the
straightforward versions of Physical and Psychological Reductionism that I have
described in this section.

Non-Reductionism  In the *Discourse on Method*, Descartes wrote

> I knew that I was a substance whose whole essence or nature is solely
to think, and which does not require any place, or depend on any
material thing, in order to exist. Accordingly this 'I' - that is, the soul
by which I am what I am - is entirely distinct from the body (36;
emphasis my own).

In this passage Descartes takes a very common intuition, the intuition that I could
exist apart from my body, as supporting the claim that I must be some sort of
spiritual or intellectual substance, a separately existing entity that is distinct from
any of the physical phenomena used by the Physical Reductionist to explain the
persistence of persons. The persistence of the person, according to the Cartesian
Non-Reductionist, is determined by the persistence of the soul. This version of
Non-Reductionism is committed to substance dualism; it holds that there are two
kinds of substances, one mental and the other physical. Persons have physical
bodies only contingently, whereas each is identified with the mental substance that
is his soul. A person persists as long as and only as long as his soul persists.

It is clear how the Cartesian Non-Reductionist differs from the Physical
Reductionist, but how does his view differ from Psychological Reductionism? The
Psychological Reductionist need not posit any mental substance underlying
psychological continuities, and he claims that it is these continuities rather than some mental substance which determine personal identity across time. The Cartesian Non-Reductionist claims that experiences must be had by “a substance whose whole essence or nature is solely to think”, that such a substance is necessary if a set of experiences is to be united as the experiences of a single person. Persons are not bundles of or constructions out of experiences and other psychological events and states. Rather, persons are substances that have, underlie, and bind together experiences.

Despite philosophical worries about substance dualism, Cartesian Non-Reductionism is still a very intuitive account of what persons are and of how they persist. Most of us do think that we are entities that could acquire new bodies or even become disembodied, and Cartesian Non-Reductionism, unlike Physical Reductionism, accommodates these intuitions. We also have the intuition that we are entities that are distinct from any of our particular mental states, and are even distinct from all of our mental states taken collectively. Cartesian Non-Reductionism captures this intuition that seems to be in conflict with Psychological Reductionism. Thus, if the Cartesian view turns out to be a good account of personal identity, we will have at least one good reason to be attracted to substance dualism.

Not all Non-Reductionists, however, are committed to substance dualism. Instead of explaining personal identity in terms of the simple, unanalysable notion
of a mental substance, one might try to explain personal identity in terms of a simple, unanalysable notion of a physical substance. Although we think that we could be disembodied, we have also the conflicting intuition that persons persist as long as and no longer than their bodies do. This latter intuition is what motivates Physical Reductionism. But a Physical Non-Reductionist, unlike a Physical Reductionist, will deny that we can give a reductive criterion for human bodies in terms of material and spatio-temporal continuities. ‘Same body’, like ‘same soul’, cannot be analyzed. It is the ultimate notion in terms of which we can explain personal identity.

Despite Physical Non-Reductionism’s greater economy with respect to kinds of substances it posits and despite the body’s seeming to be more familiar than is the soul, I think that Physical Non-Reductionism is a less attractive alternative than is Cartesian Non-Reductionism. The soul is typically taken to be an indivisible substance that is not subject to change, so we can understand why the notion of the soul is not further analysable. Human bodies, on the other hand, like all physical objects, are, it would seem, divisible and subject to change. Therefore, one is naturally led to wonder why the notion of a body is not analysable in terms of the physical continuities appealed to by the Physical Reductionist. Thus, in what follows I will concentrate on Cartesian Non-Reductionism. ³

³In Chapter 3 we will see that Bernard Williams may hold some form of Physical Non-Reductionism.
Both Cartesian and Physical Non-Reductionism agree that “personal identity is something ultimate, unanalysable in terms of such observable and experiencable phenomena as bodily continuity and continuity of memory” (Swinburne 26). But both views also add a claim about what sort of substance, mental or physical, determines personal identity. Another form of Non-Reductionism, what Parfit calls the ‘Further Fact View’, makes no claim about what sort of substance a person is to be identified with; rather, it holds only that personal identity is not to be analysed in terms of mental or physical continuities. The Further Fact View, then, is no more than a denial of Reductionism. Because such a view is only properly a response to the failure of Reductionism as an account of personal identity, I will leave it aside. Only if we were to find Reductionism implausible should we consider the Further Fact View.

1.3 The Distinction

Now that we have a basic idea of what the main forms of Reductionism and of Non-Reductionism are, we need to see what it is that characterizes a view as being either Reductionist or Non-Reductionist. Let us now return to the distinction between ‘more particular facts’ and ‘further facts’ and try to give it some more substance. It is clear that physical and psychological phenomena are more particular facts, while the existence of Cartesian Egos or souls or unanalysable bodies are ‘further facts’. So far, however, the distinction is unhelpful; it comes to
no more than the distinction between the types of facts that Reductionists use to analyze personal identity and the types of facts that Non-Reductionists use. What we need to know is how facts such as psychological continuity and physical continuity differ from facts such as the existence of Cartesian Egos.

One possibility is that we should view 'more particular facts' as 'familiar facts' and 'further facts' as 'non-familiar facts', where familiar facts are ones that we already know about or believe in the existence of before we come to analyze personal identity. Consider the Psychological Reductionist's account of personal identity. He analyzes it in terms of facts of whose existence we are already convinced and about which we already have some understanding. We accept that there are experiences, beliefs, and memories, and that these can stand in relations of similarity or causal connectedness to one another, and we already have some idea of what sorts of entities experiences, beliefs, and memories are. And the Physical Reductionist asks us only to accept that there is matter that undergoes change such as gradual replacement of parts and movement through space and time. So Reductionists add nothing new to our ontology when they analyze personal identity. The Cartesian Non-Reductionist, on the other hand, analyzes personal identity in terms of facts with which we are not already familiar; he adds new entities to our ontology. Cartesian Egos are not entities that we all think that we understand. The Further Fact View is simply the denial that personal identity can be analyzed in terms of anything that we already know about or can come to
know about. The Physical Non-Reductionist starts with what looks like the familiar notion of the human body but when he claims that it cannot be analysed in terms of continuities, it becomes clear that this sort of body is "of a kind that is not yet recognized in the theories of contemporary physics" (Parfit 210).

Another way of getting at the difference between further facts and more particular facts is to see further facts as being in some way distinct from the evidence that we use, or could possibly use, in making at least third-person judgments of identity. I have pointed out that there are two senses of the term 'criterion', one epistemological and the other metaphysical. An epistemological criterion spells out a method by which we can determine questions of identity, whereas a metaphysical criterion spells out the actual necessary and sufficient conditions of personal identity. Any epistemological criterion should refer to facts about physical and psychological continuities; these continuities are our evidence of personal identity (except perhaps in first-person judgments - see section 1.3.2 below). If we are unsure about whether x is identical with y, we find out by checking facts about physical and psychological continuity.

Reductionists appeal to the same facts in their metaphysical criteria as are appealed to in our epistemological criteria. Non-Reductionism, on the other hand, is precisely the view that the facts which are the necessary and sufficient conditions of identity are distinct from the facts which we take to be good evidence of personal identity. Some further fact distinct from the facts which we take to be
evidential is what determines personal identity. The Non-Reductionist will agree that facts about continuities are evidence of personal identity, but he denies that those continuities are the necessary and sufficient conditions of identity. In section 2.2.3, I will consider whether this is a coherent and plausible position.

1.4 Why Should We Suppose That There Are Further Facts?

1.4.1 The Subject of Experiences Argument

What reason do we have to posit separately existing spiritual substances? Recall that Descartes suggested that it is easy to imagine oneself separated from one's body. This thought prompts him to ask

But what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory experiences (83).

What Descartes says in this passage seems clearly true. I perform certain actions, I feel certain emotions, and I perceive certain sensations. But, more than that, I seem somehow to ‘stand behind’ my actions, emotions, and sensations. Even though experiences belong to me, I am not my experiences; rather, I am an entity that has experiences, I am a subject of experiences. Just as I only contingently have a body, so also do I only contingently have any particular experiences. A subject of experiences is, in some important sense, independent of the experiences
Reid objected to Locke's version of Psychological Reductionism, because, he contended, Psychological Reductionism cannot accommodate the truth of the claim that persons are subjects of experiences. A Psychological Reductionist analyzes personal identity in terms of experiences and other psychological phenomena and their interrelationships. But, Reid claims, all of our experiences or "operation[s] of our minds" are "successive in their nature"; in other words, an experience that I had yesterday is necessarily different from one that I am having today. A subject of experiences or the 'self', however, must have a "[c]ontinued uninterrupted existence" (109; in Perry 1975). Therefore, Reid concludes, there is a manifest problem in analyzing personal identity in terms of successive experiences:

Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling; [i] I am something that thinks, and acts, and suffers. My thoughts, and actions, and feelings, change every moment; they have no continued, but a successive existence; but [ii] that self, or I, to which they belong, is permanent, and has the same relation to all the succeeding thoughts, actions, and feelings which I call mine (109; in Perry 1975).

Reid is making two claims in this passage. First, he reiterates Descartes' point that a person is something that has experiences and is not to be identified with any or even with the entire collection of her experiences. Second, he says that an

---

4I will concentrate on Psychological Reductionism because that is the form of Reductionism that Reid had in mind. As long as Reid's argument is not conclusive against any one form of Reductionism, then it does not constitute a good argument for Non-Reductionism.
identification of a person with her experiences would commit us to the view that persons do not persist. Experiences are fleeting things, constantly giving way to new and different experiences. But a person is something that has each of these experiences in turn and yet remains the same person. I stand in the ownership relation to each of my experiences, and the same I stands in that same relationship to each momentary experience.

Let us begin with Reid’s first point. Does the Psychological Reductionist deny that a person is a subject of experiences, something that has experiences and is yet distinct from those experiences? Hume (164, 168; in Perry 1975), of course, did explicitly deny that a person is anything more than a bundle of perceptions and sensations and he therefore concluded that the notion of a subject of experiences is an illusory one. But a Psychological Reductionist can agree that a person is nothing beyond a collection of experiences and yet not draw the conclusion that there are no subjects of experience. Consider again the Psychological Reductionist criterion of personal identity: person x is identical with person y if and only if x is psychologically continuous with y. This criterion says that if person x has certain experiences and person y has certain other appropriately related experiences, then x is identical with y. So the criterion grants that persons are entities that have experiences, that they are subjects of experiences. The Psychological Reductionist claims that what it is for a subject of experiences to have a given experience is for that experience to be a member of the collection that is that particular subject of
experiences. Subjects of experience are distinct from experiences, because
collections of experiences are distinct from their members.

This Psychological Reductionist response does not seem entirely satisfactory.
As Parfit says, "[a] Reductionist can admit that ... a person is what has
experiences, or the subject of experiences. This is true because of the way in which
we talk" (223). Because the Psychological Reductionist agrees that there are
persons, and our concept of a person is the concept of a subject of experiences, the
Reductionist can also agree that there are subjects of experiences. The
Reductionist tells us that persons are collections of appropriately related
experiences; therefore, she concludes, subjects of experiences are such collections.
But the intuitive conception of subjects of experiences is a conception of entities
that are independent of momentary experiences in a way that collections are not
independent of their members. So the Psychological Reductionist, while
accommodating the mere claim that we are subjects of experiences, denies that
subjects of experiences are what, intuitively, they seem to be.

What Psychological Reductionism denies is that subjects of experiences are
further facts or souls that 'stand behind' experiences and perceptions. This notion
of subjects of experiences is intuitively compelling, and it is the intuition that
motivates Non-Reductionism. So to point out that Psychological Reductionism
accommodates subjects of experiences but not in this intuitive manner is just to
point out that Reductionism is not Non-Reductionism. Non-Reductionism, then,
still owes us a reason to accept the intuitive picture of subjects of experiences. This reason is supposed to be supplied by Reid’s second claim. Reid says that persons or selves are permanent entities, entities that persist through time. But then persons must be entities that are distinct from and independent of psychological phenomena, because psychological phenomena are only momentary. Experiences constantly give way to new and different experiences. If we do not posit some sort of further fact, we will have no way of accounting for the persistence of persons.

The Reductionist, however, can explain and accept the fact that persons are temporally extended entities. A person is not to be identified with one particular perception, but with a temporally extended series of appropriately related experiences. If experience e is appropriately related to experience f, then e and f are co-personal. Because e and f may not be co-temporal, the person, as a bundle of appropriately related experiences, can persist through time. If the person is the collection of co-personal experiences, then the person is related to each experience in the same way; the person is the only maximal collection of co-personal experiences of which each experiences is a member. So it does not seem that persons must be further facts for it to be the case that persons are more than momentary entities.

The only problem, then, that the Subjects of Experiences Argument presents for Reductionism involves Reductionism’s inability to capture the intuitively compelling picture of a person as something which stands behind the constantly
changing stream of experiences and perceptions. In section 1.4, however, we will see that this picture creates some problems for the Non-Reductionist. But now we need to look at another argument for Non-Reductionism.

1.4.2 The Argument from Introspection

The Non-Reductionist still needs to provide a compelling reason to posit a separately existing entity. Butler claimed that such a reason is provided by an understanding of how each of us acquires knowledge of her own persistence. He writes that “by reflecting upon that which is myself now, and that which was myself twenty years ago, I discern they are not two, but one and the same self” (Butler 106; in Perry 1975). He goes on to claim that

though the successive consciousnesses which we have of our own existence are not the same, yet are they consciousnesses of one and the same thing or object; of the same person, self, or living agent. The person, of whose existence the consciousness is felt now, and was felt an hour or a year ago, is discerned to be, not two persons, but one and the same person; and therefore is one and the same (102; in Perry 1975).

Butler begins by reiterating the point stressed by Reid, that each of my experiences is necessarily not identical with the succeeding experiences of self-awareness. Nonetheless, Butler says, each of my experiences of self-awareness is an experience of being aware of one and the same persisting self. Moreover, I am aware of that self as persisting. Each time I have an introspective experience of being self-aware, I know that it is the same self of which I was aware in the preceding introspective state of self-awareness.
How are these facts about self-awareness supposed to support the claim that our identity through time must involve a further fact? These claims seem to support the view that in introspection we are given non-inferential knowledge of our own continued existence. I am aware that the person of whom I am now self-aware is the same person of whom I was self-aware previously. Cartesian Non-Reductionism offers a simple explanation of how this knowledge of persistence can be given to me through introspection. If I am an entity that stands behind my experiences, then each of my experiences of self-awareness is an experience of being this particular persisting entity. Each of the experiences is, of course, different, but they are all experiences of being the same persisting separately existing entity. Thus, introspection provides me with non-inferential knowledge of my own identity. Merely by introspecting on my own states of self-awareness, I can know that the self of which I am aware today is the same self of which I was aware yesterday.

*Prima facie* it seems that Reductionism cannot accommodate non-inferential knowledge of self-identity. (Again, here, I will concentrate on Psychological Reductionism.) What is it, according to Psychological Reductionism, for me today to be identical with me yesterday? It is for my experiences today to be appropriately related to my experiences yesterday, for me today to be psychologically continuous with me yesterday. The relation between my experiences is not, however, something that can be given in experience. Each experience is different from the succeeding one, and each state of awareness I have
can only be an awareness of my present state, not of that state's relation to any other state. Therefore, if Psychological Reductionism were true, I could not have knowledge of my own persistence as a result merely of introspection. Rather, I would have to infer my persistence from facts about continuity. But, then, if Butler is correct in claiming that I can have non-inferential knowledge of my own persistence, then Psychological Reductionism is false.

It is true that I do seem to have knowledge of my own identity through time, and I never check facts about psychological continuity in order to acquire such knowledge. But introspection can be a reliable source of knowledge about personal identity without introspection's allowing me access to the necessary and sufficient conditions of identity. Memory gives me access to past experiences, and because memory is typically reliable, it will usually be true that I am who I think I am, that the person remembering is identical with the person who had the experience being remembered. Of course, there can be failures of memory and we often make mistakes. When introspection is not reliable, we may need to go to great lengths to check our own identities; we will need to investigate psychological continuities. But the mere possibility of error does not impugn the Psychological Reductionist criterion. In normal cases we can trust introspection to provide knowledge of personal identity. So the Reductionist must deny that we have non-inferential knowledge of our own identity: our knowledge is based on claims about the reliability of memory or about continuities. Nonetheless, the Reductionist can
explain the force of Butler's claim. We do appear to have non-inferential
knowledge in so far as we typically are justified in relying directly on introspection
as a source of knowledge about our own identities.

Although the Reductionist must deny Butler's claim, it is not clear that
Non-Reductionism fares any better on this point. For there certainly are cases in
which one person thinks that he is aware, in succeeding consciousnesses, of two
distinct persons, or in which one person in succeeding states fails to be aware of
any persisting self. For example, if a man goes insane, he may come to think that
he is Napoleon; he will mistakenly believe that his present experience is an
experience of a self that was at Waterloo. Or an amnesiac will be unable to discern
by introspection that his current state of self-awareness is an awareness of a self
that had a particular experience before the automobile accident that caused the
amnesia. Again, the mere possibility of epistemological failure does not impugn
the Cartesian Non-Reductionist criterion. But whereas the Reductionist had some
recourse in such cases of failure, we have to wonder what recourse is available to
the Non-Reductionist. If introspection fails to provide me with knowledge of my
identity, how can I, if Non-Reductionism is true, acquire such knowledge? I
cannot, of course, have any direct empirical knowledge that my self now is identical
with some previous self, because such identity is determined by the persistence of
an entity that is independent of any empirical phenomena. The Non-Reductionist
will have to say that I should rely, as does the Reductionist, on the evidence
appealed to by third persons in making judgments about my identity; he will have
to appeal to facts about psychological and physical continuity. But for the
Non-Reductionist the self is somehow independent of these continuities. So can the
Non-Reductionist consistently appeal to such continuities as evidence of personal
identity? This is one of the questions that I will address in the following section.

1.5 Indeterminacy and the Epistemology of Personal Identity

Derek Parfit argues, on the basis of his Combined Spectrum example, that we
should reject Non-Reductionism because it cannot allow that it can be
indeterminate whether I survive in some case. I think that, at least prima facie,
this fact might actually support Non-Reductionism, but I will describe Parfit’s
example and argument because they are suggestive of some further problems for
the Non-Reductionist.

I will begin with a description of the Combined Spectrum. Consider the
following range of possible cases which involve variations in the degrees of physical
and psychological continuity which hold between me and a future person:

Case 1: In this case nothing is done to me. Therefore, I am fully
continuous both physically and psychologically with the resulting
person. This is just the normal case.

Case 2: Case 2 differs from Case 1 in that in Case 2 scientists replace a
couple of my body cells with new ones and they alter a couple of my beliefs or memories or other psychological states.

We can imagine many such cases and in each Case \( j \) the resulting person is slightly less connected, both physically and psychologically, to me than the resulting person was in case \( j-1 \). We have at the end of the Spectrum,

Case \( n \): In this case, the scientists again begin with me, but they destroy my body and then make an exact replica, both physical and psychological, of Greta Garbo. (Parfit 236-237)

I have described a range of possible cases. We are not to imagine that first the scientists replace a few of my cells, and then a few more, and eventually they destroy my body. Rather, there are many cases: in the first nothing is done to me, in the second scientists only replace a few of my cells and stop, in the third scientists replace a larger number of cells and stop, etc. In each case the changes are instantaneous or as nearly so as possible.

Now, suppose that we were to consider each case and to ask about the resulting person in each case ‘Is she identical with me?’. It seems quite clear that in Cases 1 and 2 and cases close to the beginning of the Spectrum, the resulting person is me: surely I can lose a few body cells and a couple of memories or beliefs and yet survive. It also seems quite clear that in Cases \( n, n-1 \), and other cases near the end
of the Spectrum, the resulting person is not me: after my body is destroyed why should we think that the replica of Greta Garbo is me rather than Greta Garbo or an entirely new person who has just been created? But what about the cases in the middle of the Spectrum? In those cases, is the resulting person me?

If we say that I survive in the cases in the near end of the spectrum but not in the cases in the far end, what options do we have about the middle range of cases? First, we could say that there is some case j in the middle of the spectrum such that I survive in j but not in j+1. This option is plausible only for the Non-Reductionist, not for the Reductionist. Non-Reductionists typically claim that the soul is indivisible; either one has a particular soul or one does not have that particular soul. Having a soul cannot be a matter of degree. Thus, the Non-Reductionist will say that in each case of the spectrum either the resulting person has my soul or she does not. Because it seems that I do not survive as the resulting person at the far end of the spectrum while I do survive at the near end, it must be the case, according to the Non-Reductionist, that there is some case in the middle range that is the first case where the resulting person does not have my soul.

Parfit claims that it is implausible to suppose that there is such a first case:

It is hard to believe ... that the difference between life and death could just consist in any of the very small differences described above [in the Combined Spectrum]. We are inclined to believe that there is always a difference between some future person’s being me and his being someone else. And we are inclined to believe that this is a deep difference. But between neighboring cases in this Spectrum the
differences are trivial.

Parfit concludes that it is implausible to suppose of any case \( j \) that I survive in \( j \) while I do not survive in \( j+1 \). The best option for the Reductionist, then, is to say that it is *indeterminate* whether I survive in the middle range of cases.

In order to motivate the idea of indeterminacy, let us consider another case where most of us already are compelled to admit indeterminate cases. Take the spectrum of color that ranges from orange to red. At the near end of the spectrum are colors that are clearly orange while at the far end the colors are clearly red. But what about the colors in the middle of the spectrum? It does not seem right to say that they are orange nor does it seem right to say that they are red. We usually call such cases 'borderline cases'. There is just no saying whether these colors are orange or red; in other words, it is indeterminate whether they are orange or red. In this color spectrum, I think that most of us would agree that it is implausible to suppose that there is a sharp borderline such that on one side of it all of the colors are orange but on the other side of it all of the colors are red. Any borderline that we might draw would seem entirely arbitrary.

Now let us return to the Combined Spectrum. The Reductionist claims that identity is determined by physical and psychological continuity. In the near end of the spectrum, there are clearly sufficient connections to guarantee my survival while in the far end there are clearly not enough. In the middle range of cases, there is no saying whether there are enough connections; these cases are borderline.
cases. The notion of psychological continuity is vague. It involves the idea of ‘sufficient’ connections holding between \( x \) and \( y \), but it can be indeterminate whether there are sufficient connections just as it can be indeterminate whether a particular color is orange or red. Having connections can be a matter of degree, and how can it be that for some \( j \) there are sufficient connections in \( j \) but not in \( j+1 \) when the difference between \( j \) and \( j+1 \) is trivial? Given the implausibility of there being a sharp borderline in the Combined Spectrum and given the Non-Reductionist’s inability to accommodate indeterminacy (having a particular soul is not a matter of degree), Parfit concludes that the Combined Spectrum is an argument against Non-Reductionism.

In order, then, for Parfit’s Combined Spectrum Argument to be a good argument against Non-Reductionism, both of the following claims must be true: it must be implausible that there is a sharp borderline in the spectrum, and the Non-Reductionist must be unable to accommodate indeterminacy. Let us begin with the claim that it is implausible to suppose that there is a sharp borderline somewhere in the spectrum. Parfit’s reason for thinking this supposition implausible is that it would commit us to the claim that the difference between the resulting person’s being me and her not being me would consist in the very small difference between neighboring cases. Such a small difference is not a deep enough difference to constitute the difference between my surviving and my not surviving. I think, however, that this is a reason only for the Reductionist, not for the
Non-Reductionist, to find it implausible that there is a sharp borderline somewhere in the spectrum. According to the Reductionist, identity consists in the connections described in the spectrum, so any difference between my surviving and my not surviving must involve some difference in the connections. Parfit is right to claim that that difference could not be as trivial as that existing between neighboring cases. The Non-Reductionist, on the other hand, denies that identity consists in connections; rather, he claims that identity is determined by the persistence of a soul. So if there is a sharp borderline somewhere in the spectrum, the difference in neighboring cases, according to the Non-Reductionist, is anything but trivial, because in some case $j$ the resulting person has my soul while in case $j+1$ the resulting person does not have my soul. Because the soul is indivisible, it is not all implausible for the Non-Reductionist to claim that there is such a case $j$ somewhere in the spectrum.

And, if the Non-Reductionist can plausibly hold that there is a sharp borderline somewhere in the spectrum, then I think that this is a more intuitively attractive option than the Reductionist option of claiming that it is indeterminate whether I survive in the middle range of cases. In order to motivate the indeterminacy option, I used the example of borderline cases on the spectrum of colors from orange to red and pointed out that it is plausible to say, in those cases, that it is indeterminate whether the given color is orange or red. But the issue of whether I survive seems entirely different from the issue of whether a certain color is orange.
or red. For consider that I am told that the resulting person in one of the middle cases will suffer great pain. Should I dread that pain? Should I be indifferent to it? The answers to these questions seem to depend upon the answer to the question about whether I survive. If it is indeterminate whether I will survive, there seems no way of understanding how I should view the resulting person. It just seems that either I will suffer pain or I won’t. For these reasons we do not want to be too quick to accept Reductionism and indeterminacy (cf. Swinburne 18-21). (I will consider how to deal with indeterminacy and future-directed self-concern in Chapter 3.)

Parfit, however, has another reason for rejecting the Non-Reductionist option of a sharp borderline. He claims that “[i]t is ... hard to believe ... that there must be such a sharp borderline, somewhere in the spectrum, though we could never have any evidence where the borderline would be” (239). If we were Reductionists, we could have no way of telling which small difference between neighboring cases is the difference between my surviving and my not surviving. If we were Non-Reductionists, it seems that our problem would be more acute. At least the Reductionist would know that the difference between life and death consists in one of the small differences in connections. That, however, is what the Non-Reductionist denies. For the Non-Reductionist the borderline occurs where there are neighboring cases such that in the first the resulting person has my soul while in the next she does not. But that change, it seems, has nothing to do with
any small changes in connections. Because the soul stands behind and underlies changes in psychological phenomena, it is difficult to see why any small change in connections would cause or indicate the loss of the soul. So not only can we have no evidence where the borderline is, we cannot even discern what sort of thing would constitute evidence for the borderline.

Now an even more serious problem arises for the Non-Reductionist. If my persistence is not determined by changes in degree of continuity, then how can we be sure that I do survive in the near cases of the spectrum and that I do not survive in the far cases? We are inclined to say that I can survive the sorts of changes that occur in the near cases and that I cannot survive the sorts of changes that occur in the far cases. We are unsure about whether the changes in the middle cases are such that I can survive them. The Reductionist has an account of why I can survive certain changes but not others; he says that my identity is determined by degree of continuity, which is in turn determined by whether changes such as those described in the spectrum occur. The Non-Reductionist theory of identity, on the other hand, does not offer an account of how such changes are related to the persistence or non-persistence of the soul. Therefore, the Non-Reductionist has no account of how we can know whether the resulting person in any of the cases of the spectrum is me. Non-Reductionism, then, seems to commit us to a general skepticism about personal identity; at least in the case of third person judgments of identity, the basis for our judgments is no indicator
at all as to whether or not there is personal identity. (And introspection would be little help in the middle range of cases. What is at issue here is whether the resulting person really is whoever it is that she thinks herself to be.)

The Non-Reductionist might respond that he posited the existence of a soul precisely to be able to account for and to underlie the sorts of changes described in the spectrum. Persons do persist through time and through drastic psychological and physical changes. To account for the persistence of a person through such changes, one must suppose that there is some entity such as the soul behind the changes and not subject to the changes. But, then, the Reductionist can respond, why not suppose that the soul is still behind the drastic changes at the far end of the spectrum and that the resulting person is me? After all, the Non-Reductionist allows that the soul can survive death. Part of the motivation for Non-Reductionism lies in the fact that it can allow for persons to become disembodied. So if the soul can survive a drastic change like death, why not suppose that it also survives the drastic changes of the Combined Spectrum? The Non-Reductionist does not offer us a theory of the persistence of the soul, so he offers us no way of determining what sorts of changes are changes that would cause the soul to cease to exist. The Non-Reductionist, then, has no reason to reject the counterintuitive claim that I survive as the resulting person in the cases at the far end of the Spectrum, but he also has no theory concerning which changes are indications of the persistence of the soul. On the other hand, the
Reductionist can account for the persistence of the person through change; for the Reductionist identity is determined by whether or not changes are governed by the sorts of regularities used to define psychological and physical continuities. We seem to have no need for the soul, especially if it makes it the case that we could never have knowledge of personal identity. The Non-Reductionist fails to offer an account of why certain changes seem to provide good reason for claiming that a certain person has ceased to exist.

The Non-Reductionist may respond by saying that connections are evidence, even on his view, of personal identity. As we saw in section 1.2.1, the Non-Reductionist, unlike the Reductionist, makes a much sharper distinction between the metaphysical and epistemological criteria of identity. The Non-Reductionist will say that the persistence of the soul is the necessary and sufficient condition of personal identity, but our way of knowing about personal identity involves the sorts of connections and changes described in the Combined Spectrum. To point out that for the Non-Reductionist there is a sharp divide between the sorts of facts involved in his metaphysical criterion and the sorts of facts involved in our epistemological criteria is just to point out what differentiates Reductionism from Non-Reductionism.

But how, we can ask, are the epistemological and metaphysical criteria related? In order for our use of the epistemological criteria to yield knowledge, connections must be in some way tied to the persistence of the soul. Perhaps the
Non-Reductionist can offer some theory that says that we cannot account for the laws and regularities that typically govern psychological continuity unless we posit the soul as a regulating principle. If the soul were what regulates the continuities that we observe, then we could conclude that when continuities are no longer present, the soul must have ceased to exist or, at the least, have departed from the body. This is the sort of claim that the Non-Reductionist would have to defend if he wants to avoid a complete skepticism about personal identity. But I think that it is very difficult to see why, even if the soul is what regulates continuities, the presence of continuities need indicate the persistence of one and the same soul.

Parfit points out that

[a]s both Locke and Kant argued, there might be a series of such entities that were psychologically continuous. Memories might be passed from one to the next like a baton in a relay race. So might all other psychological features. Given the resulting psychological continuity, we would not be aware that one of these entities had been replaced by another. We therefore cannot know that such entities continue to exist (223).

Why might a series of souls not serve just as well as one persisting soul? I doubt whether the Non-Reductionist can give satisfactory answers to questions such as these (cf. Shoemaker 1984 124, 151).

I said above that the soul is typically taken to be an indivisible entity. But a Non-Reductionist need not hold that having a particular soul cannot be a matter of degree. If one could have a particular soul to a greater or lesser degree, then the Non-Reductionist could agree that, in the middle range of cases, it is
indeterminate whether I survive. Degrees of the soul, according to this version of Non-Reductionism, would correlate with degrees of continuity. The main problem with this approach is that, without some explanation of how it is that degrees of the soul are consistently correlated with degrees of continuity, the Non-Reductionist view is starting to look *ad hoc*. The Non-Reductionist seems to be searching for some way to incorporate both a Reductionist criterion of identity and the thesis that personal identity consists in some further fact. We should conclude that the further fact is simply superfluous.

So, at the very least, the Non-Reductionist needs to show why psychological continuities require an underlying immaterial regulating principle and why a series of such principles would not do as well as a single persistent one, or why degrees of a divisible soul are correlated with degrees of observable continuities. Without arguments for either of these claims, it appears that Non-Reductionism commits us to a general skepticism about personal identity, leaving us unable to understand why persons can survive certain changes but not others. I think that we now have enough reason to try to develop a plausible Reductionist view, despite its commitment to indeterminacy about survival. We will see in Chapter 3 that indeterminacy is not really a problem for Reductionism, that it coheres well with a plausible theory about what matters in our survival.
2 The Reductionist Alternatives

I have already briefly described two forms of Reductionism, Psychological Reductionism and Physical Reductionism. In this section I will develop and motivate these views, make some necessary modifications to the criteria discussed above, and suggest some alternative versions of both views. I will attempt to motivate the claim that psychological continuity is at least a necessary condition of personal identity. The next chapter will be devoted to a discussion of whether psychological continuity is also a sufficient condition of survival.

2.1 Physical Reductionism

Although Psychological Reductionism is the dominant view among philosophers today, I am going to begin with a discussion of Physical Reductionism. I think that Physical Reductionism is more easily motivated, while Psychological Reductionism is only appealing after we see some of the inadequacies of the Physical view. For consider: I have a body, I have the same body today that I had yesterday, and I am quite sure that I will have this body tomorrow. Even Descartes, our paradigm Non-Reductionist, admitted the pull of some form of a Physical view: “I considered myself, firstly, as having a face, hands, arms, and the whole machine made up of flesh and bones, such as it appears in a corpse and which I designated by the name of body” (104).

So it is obviously true that I and all other persons have bodies. But, of course,
having a body is not the same as being identified with a body. Books have covers, yet we certainly think that a book can have its cover replaced and still remain the same book. Even though, on the one hand, “there is the feeling that my identity cannot possibly be the identity of a body I can clearly imagine myself exchanging for another body, or even imagine myself losing altogether” (Wiggins 45), there is also, on the other hand, the feeling that persons are related to their bodies in a much more intimate way than the way in which books are related to their covers. So, in some sense, it seems very natural to suggest the following as an identity criterion for persons:

(1) person $x$ is identical with person $y$ if and only if $x$ and $y$ have the same body.

In offering criterion (1) we have analyzed the identity of persons in terms of the identity of certain physical objects, namely human bodies. If we accept (1), then we are also committed to the following claims: a person cannot get a new body, a person cannot survive without his body, and two persons cannot share one body.

The Physical Non-Reductionist could accept (1) as it stands. But, as I said above, human bodies seem to be like other physical objects in so far as it seems possible to give their identity conditions in terms of observable continuities. The Physical Reductionist, then, goes on to analyze the notion of ‘same body’. Bodies are physical objects, so the identity conditions for bodies will parallel, in certain respects, those for chairs, watches, ships, houses, etc. But living bodies are also
significantly different from artifacts. I am not going to go into these issues in depth; rather, I will only attempt to give some idea of what sorts of continuities might be involved in the persistence of human bodies.

There are two factors which *prima facie* seem relevant to the identity of physical objects: (i) spatio-temporal continuity, and (ii) material continuity.

Physical continuity will involve both (i) and (ii). Let us begin by considering (ii). For some objects we might think that if that object is to persist, it cannot lose or gain any matter. Locke (35; in Perry 1975) claimed that heaps and piles are such objects. It is clear, however, that human bodies are not like heaps and piles in this respect. Human bodies acquire entirely new cells every few years (people eat, dispose of waste, lose weight, gain weight, etc.), but human bodies endure for longer than a couple of years. It is clear that at least in any normal case, a person has the same body throughout his entire life. So for human bodies, physical continuity involves the gradual change of bodily matter. The change must be gradual. Complete, instantaneous replacement of all or most cells would certainly constitute the creation of a new body. How gradual must the change be? How many cells can be simultaneously replaced without destroying the original body? These are questions which will come up again, but for now we can simply say that any replacement of cells cannot involve a ‘significant portion’ of the body’s matter without destroying that body and creating a new one.

Now, briefly, consider (i) spatio-temporal continuity. For an object to be
spatio-temporally continuous is for it to trace a continuous path through space and time (cf. Parfit 203). For example, an object cannot disappear in one spot and then reappear suddenly in a different spot not contiguous to the first unless it traces a path between the two spots. I suppose that the question about temporal gaps is more debatable - can we disassemble an object and then reassemble the same object?5 (Parfit 203-204). Again, I will not answer these questions although I am inclined to think that a human body is the sort of object that cannot have any spatial or temporal gaps in its history.

With these preliminaries we can restate (1):

Physical Reductionism: person x is identical with person y if and only if person x’s body is physically continuous with person y’s body.

As I have remarked, if we accept Physical Reductionism we can accept that person x is identical with person y even if person y’s body does not have all of the same cells as person x’s body. We want to allow such cases since it is an empirical fact that the cells of a person’s body are constantly being lost and replaced with new ones, although it is never the case that all cells are ever simultaneously lost and replaced (Parfit 204). But Physical Reductionism also allows that I could survive an operation in which all of my cells are gradually replaced with new ones until all of my old cells have been destroyed and new ones put in their places. A Physical Reductionist could stipulate that the replacement of cells must occur as it does in

---

5 Notice that if we answer ‘yes’ we will have to allow for spatial gaps as well if the object is disassembled and reassembled at spatially discontiguous points.
the normal case, by some sort of 'natural' process perhaps, in order for physical continuity to be sufficient for personal identity. However, we do want to allow that a person could get artificial limbs or a liver transplant and yet have the same body. I think that the Physical Reductionist would do better by stressing the gradual replacement of parts; if the operation only takes a minute we would be less inclined to say that I survive than if the operation takes a year.

Notice that we started with (1) and then reformulated it as Physical Reductionism. We set out to capture the intuitive notion of having the same body and arrived at the notion of a physically continuous body. A question then arises as to whether having the same body is just to have a physically continuous body. In the next chapter, I will suggest that cases of physical splitting force us to answer no to this question. What I want to notice here is that much of the initial appeal of the Physical Reductionist view derives from the intuitive idea of one person, one body, and vice versa. Williams points out that "if I am asked whether the person in front of me is the same person as one uniquely present at place a at time t, I shall not necessarily be justified in answering 'yes' merely because I am justified in saying that this human body is the same as that present at a at t" (1). However, we do normally feel confident in identifying persons on the basis of bodily criteria. If I am trying to pick the man who stole my wallet out of a police line-up, I do not say "That is the body that took my wallet, but I cannot be sure that it is the person who took my wallet". I conclude from the fact that it is the
same body that it is the same person.

However, such considerations can only serve to motivate the Physical view. They cannot be used as an argument because, as I have said, we have to be careful not to assume that our evidence of personal identity constitutes necessary and sufficient conditions of identity. The Physical view does offer a simple explanation of why some of our evidence is good evidence. But I often resort to other than bodily facts in identifying persons. I may ask them questions about their past, etc. Also, we often manage to identify persons whom we have never seen or cannot presently see, as when I recognize whom I are talking with on the telephone even if her voice sounds different. So perhaps persons only keep the same bodies throughout their lives as a matter of contingent fact. Before we consider this possibility, we need to consider one more issue within Physical Reductionism.

**Body or Brain?** Consider the following case taken from Shoemaker (1963 23-25). Brown and Robinson enter the hospital for a special type of brain operation which requires the surgeons to remove the brains of Brown and Robinson from their respective heads. When it comes time to replace their brains in their heads, Brown’s brain is accidently put into Robinson’s body and *vice versa*. The body which contains Robinson’s brain then promptly ceases to function. Let us call the surviving person, the one with Brown’s brain and Robinson’s body, ‘Brownson’. Brownson will claim to be Brown, he will remember Brown’s life, he will love Brown’s wife, etc. Is Brownson Brown or Robinson or
neither? Brownson has most of Robinson's body; in fact, he has all of it except for the brain. Does he, then, have a sufficient amount of Robinson's body to make him physically continuous with and, thus, according to Physical Reductionism, identical with Robinson?

I think that most of us, at least initially, have the intuition that Brownson is identical with Brown. This intuition may lead to a problem for the Physical view. Brownson has very little of Brown's body, so, on a straightforward interpretation of 'physical continuity', Brownson is not physically continuous with Brown. One suggestion is that the Physical Reductionist should concentrate only on the brain, not on the entire body, that the Physical criterion should say that person x is identical with person y if and only if x and y have the same brain. Of course, human brains, like human bodies, lose cells over time (although some brain cells are never lost), so the criterion would have to be modified to: person x is identical with person y if and only if x's brain is physically continuous with y's brain.

But what is the motivation for concentrating on the brain? In the case of Brownson, it is not the brain per se that inclined us to say that Brown is Brownson, but rather the fact that Brownson has Brown's memories, values, and beliefs, in other words, the fact that he is psychologically continuous with Brown. For suppose that Brownson had Brown's brain and yet claimed to be Robinson, had Robinson's memories, etc. We would not then be at all inclined to say that Brownson is Brown. So the shift from continuity of the body to continuity of the
brain is implicitly motivated by a concentration upon psychological features of persons. What appeared to be motivation for a focus on brain continuity is really motivation for some form of Psychological Reductionism. Apart from its being the carrier of psychological continuity, it is hard to see why one might not offer as a criterion of identity instead of continuity of the brain perhaps continuity of the heart or continuity of the tongue. If a Physical Reductionist allows that persons can have heart transplants or lose their tongues, surely he should also allow for brain transplants. He has no purely Physical Reductionist way to justify concentration upon the brain. 6

So the Physical Reductionist should hold that Brownson is Robinson because Brownson has the vast majority of Robinson’s body. But our intuitions about the Brownson case indicate that it seems plausible to suppose that a person could get a new body. After all, as I said, the reason why we claim that Brownson is Brown has to do with the fact that Brown’s brain carries with it Brown’s psychology. So what if we could transfer the psychology without transferring the brain matter? Let us consider just such a case as a way of motivating Psychological Reductionism.

6In “Human Beings” Mark Johnston does hold that “one would go where one’s brain goes and that one could survive as a mere brain” (78). He tries to justify his concentration upon the brain rather than on other organs, but his emphasis upon continued mental life of a certain sort points to more than a Physical Reductionist motivation for his view.
2.2 Psychological Reductionism

To begin, we need to return to the intuition mentioned by Wiggins: he says that he can clearly imagine himself exchanging his body for a new one. Recall that the Physical view has as a consequence the claim that a person cannot get a new body. So Physical Reductionism clashes with a very common intuition, an intuition which I am going to use as motivation for Psychological Reductionism.

Let us now expand on Wiggins’ remark. In order to do this, consider the following case, taken from Williams’ “The Self and the Future” (in Williams Problems of the Self):

Body-Switch: You and I are captured by a mad scientist who wants to use us as guinea pigs to test his new invention. I am placed in one booth, booth A, and you are placed in a different booth, booth B. Let us call my body ‘A-body’ and your body ‘B-body’. The scientist then pulls a switch, producing the following changes: A-body emerges from booth A, but when we question this A-body person, she makes memory claims which fit your past life, her statements express desires, beliefs, and character traits that we previously associated with you, and her mannerisms and personality resemble very closely those that you had before the experiment. Similarly, the B-body emerges from booth B and makes memory claims which fit my past life, etc.

\[7\text{All references to Williams are to this work unless otherwise indicated}\]
From which booth do I emerge after the experiment? I think that many of us are inclined to think that I emerge from booth B and that you emerge from booth A. In other words, our intuition is that the experiment has caused you and me to switch bodies. If this description of what happens is the correct description of what happens, then, in this case you and I acquire new bodies - I get your body and you get mine. Is such a thing possible?

Of course, Body-switch is not now technologically possible, because no scientist has a machine that could cause you and me to exchange bodies. Body-switch may not even be physically possible; the laws of human physiology and psychology in this world may be such as to make persons’ changing bodies impossible in this world. We, however, are interested in whether Body-switch is metaphysically possible. Because we are interested in finding necessary and sufficient conditions of identity, we want to know whether body switching could happen in some possible world. If it can, then Physical Reductionism would be an unsatisfactory metaphysical criterion of personal identity.

Notice that we make our judgement in Body-Switch on the same basis as we initially made our judgment in Shoemaker’s Brownson case which I discussed above. We seem to think that what a person remembers, believes, and values, determines who that person is. So both the Brownson case and Body-Switch should motivate us to consider that perhaps the conditions of identity are psychological and not physical. I will now describe a view which, at least in one of
its forms, would accommodate our intuitions about Body-Switch and Brownson.

This view is Psychological Reductionism.

2.2.1 Locke and the Memory Theory

The Memory Theory and its Initial Plausibility  I want to begin by
discussing the version of Psychological Reductionism presented by John Locke.
Although this version of Psychological Reductionism is extremely inadequate, it
provides a good starting point for discussion, because the criticisms that have been
made of Locke's theory help to motivate a plausible form of Psychological
Reductionism, a form which still has Locke's theory at its core. Also, a discussion
of Locke's theory will allow us to discuss some issues that will clarify some issues
in later chapters.

Locke begins by saying that a person is a "thinking intelligent being, that has
reason and reflection, and can consider itself as itself, the same thinking thing, in
different times and places; which it does only by that consciousness which is
inseparable from thinking" (39; in Perry 1975). Because a person is a thinking
thing, his identity is due to his consciousness; therefore, it seems natural to define
personal identity over time in terms of memory, that is, in terms of consciousness
of past actions, thoughts, and experiences: "as far as this consciousness can be
extended backwards to any past action or thought, so far reaches the identity of
that person" (Locke 39; in Perry 1975). Locke, in fact, seemed to hold the view
that for person x and person y to be the same person, either y must remember having \textit{all} of the experiences, thoughts, and actions of x or x must remember having \textit{all} of the experiences, thoughts, and actions of y (Locke 46; in Perry 1975).

But this view is clearly false. I do not remember everything that I thought or did yesterday or even this morning; Locke’s view would lead to the conclusion that only persons with absolutely phenomenal memories persist for any significant duration. So we need to alter his criterion to the following:

\textbf{Memory Criterion:} \textit{person} z \textit{is identical with person} y \textit{if and only if} x remembers or can remember quite a few of the experiences, thoughts, actions, etc. of y or vice versa.

The phrase ‘can remember’ allows us to say that x could be identical with y even if x is asleep. We do not want it to be the case that x must actually be now remembering an experience of y in order for it to be the case that x is identical with y. The term ‘quite a few’ in our criterion is vague. In certain cases it will be clear that enough memory connections exist to satisfy the criterion; in other cases, such as cases of severe amnesia, it will be clear that there are not enough memory connections. In cases such as those in the middle range of cases in the Combined Spectrum, it will be indeterminate whether or not there are sufficient memory connections (see section 2.3).\footnote{It will differ from case to case what constitutes a sufficient number of memory connections. For example, suppose that x remembers having a certain number of the thoughts, etc., of y and that that is all that x remembers. In that case, we might conclude that \(x\approx y\). However, suppose...}

61
Locke’s view captures the simple intuition that we are who we think we are. As we saw in section 1.3.2, we trust the deliverances of memory to provide us with knowledge of personal identity. If I remember being at the mall on Tuesday, that is at least *prima facie* evidence that I am identical with someone who was at the mall on Tuesday. When Brownson remembers Brown’s life, we are inclined to think that Brownson is Brown. Memory, then, seems central in determining who a person is.

**Reid’s Objection** There are, however, two famous objections to Locke’s theory, objections first presented by Butler and Reid. I will begin by discussing Reid’s objection (Reid 114ff.; in Perry 1975). Reid asks us to consider the case of an army officer who, as a child, was beaten for stealing fruit from an orchard. Then, in his first battle, he leads an attack on the enemy, and, late in life, he is made a general. When he was in his first battle, he remembered or could remember the beating that he received as a child as well as other things that he did as a child; therefore, on Locke’s theory, the young officer is identical with the child. Similarly, the general can remember taking the standard and performing other actions done by the brave young officer, so the general is identical with the young officer. The general, however, does not remember being beaten as a child; in fact, the general has no memories of childhood, so the general, according to Locke, is not identical with the child. But, since the general is identical with the officer and the officer is that x also remembers having an equal number of the thoughts, etc., of z, where z is not identical with y; in this case, we would certainly hesitate in saying that x=y, or in saying that x=z.
identical with the child, by the transitivity of identity, the general is identical with the child. Thus, we have a contradiction: the general is identical with the child and the general is not identical with the child.

In Reid's example, we have the old general (G), the young officer (O), and the child (C). By hypothesis G does not have any memories of the actions or thoughts of C, i.e. G has no direct memory connections to C, but G has direct memory connections to O and O has direct memory connections to C. But, as we have seen, since G=O and O=C, the transitivity of identity forces us to conclude that G=C. Because the memory criterion yields the conclusion that G is not identical with C, we have to revise Locke's memory criterion so that it yields that G=C. Instead of direct memory connections, we ought to consider memory continuity, where memory continuity is a relation between persons at times. Memory continuity is the ancestral of direct memory connectedness:

\[
\text{person } x \text{ at time } t \text{ is } \text{memory continuous with } \text{person } y \text{ at time } t' \text{ if and only if } x \text{ and } y \text{ are the endpoints of a series of persons at times such that each person in the series is memory connected to the preceding person in the series.}
\]

For example, if \( x \text{ at } t \) is connected to \( z \text{ at } t' \) and \( z \text{ at } t' \) is connected to \( y \text{ at } t'' \), then \( x \text{ at } t \) is continuous with \( y \text{ at } t'' \) even though \( x \) is not connected to \( y \). The advantage of continuity over connectedness is that continuity, like identity, is a transitive relation. We saw that even though G is connected to O and O to C, G
may yet not be connected to C. But because G is continuous with O and O with C, G is continuous with C. Thus, memory continuity provides a suitable criterion of identity. Our modified Lockean criterion will be as follows:

Revised Memory Criterion: person x at time t is identical with person y at t' if and only if x is memory continuous with y.  

We started out with the intuition that memory and personal identity are intimately connected: this intuition was the motivation for the simple Lockean account. But after these revisions can we still appeal to our initial intuition? We all know that memory can be an erratic thing: at various times of our lives we are able to remember varying previous events, and which events we remember depends upon other factors besides merely the temporal placement of the events. The Revised Memory Criterion allows memory to wander more freely, as it were, and that, I think, makes the Revised Criterion more intuitively pleasing than the original Memory Criterion.

Butler's Objection

Now let us turn to an objection first raised by Butler. Butler charged that Locke's account of personal identity is circular: "one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in

---

9 If one takes persons at times literally as temporal parts of persons, then the Revised Memory Criterion would need to be restated as a criterion of the co-personality of temporal parts of persons. I will continue to speak of persons at times while remaining neutral as to whether persons at times are temporal parts of persons.
any other case, can constitute truth, which it presupposes” (100; in Perry 1975).

Butler can be interpreted as saying that the concept of memory presupposes personal identity. When we say that x remembers doing a, we are implying that x is identical with the person who did a. So we cannot analyze personal identity in terms of memory-continuity because memory presupposes personal identity; Locke’s account of personal identity is circular.

In order to understand Butler’s worry more clearly, we need to consider what is involved in a person x’s remembering an event E (my account here follows that of Shoemaker 1984, 81ff.): (i) *z seems to remember event E*, or we can say that *z has an ‘apparent memory’ of E*(81). (ii) *apparent memories must be like, in certain respects, the events of which they are apparent memories*. For example, the content of an apparent memory of my birthday party will be different than the content of an apparent memory of my trip home to Chicago. But, as Shoemaker points out, if my birthday party was just like yours, my apparent memory of my birthday party will be, in many respects, like your apparent memory of your birthday party. Surely, though, I do not have a memory of your party just because your party was like mine. So we add that (iii) *z’s apparent memory of E must be caused in some appropriate way by his having been involved in E.*¹⁰ So, even though my party was like yours, it is my party that I am remembering because my memory was caused by my having attended *my* party. Simply put, it is my party that I remember.

¹⁰This is what Shoemaker (1970) calls the ‘previous awareness condition for remembering’(269).
because it was my party that I attended. And now we see why memory presupposes personal identity: our third condition makes reference to the person who is doing the remembering and says that she must have been involved in E. Therefore, memory cannot be used to explain personal identity.

Butler is quite right: memory presupposes personal identity. Let us, then, define a new notion which is like memory but does not presuppose personal identity, and we will call this notion, after Shoemaker (1970), 'quasi-memory'. The first and second conditions above also hold for quasi-memory, but condition three is different: (quasi-iii) for x to have a quasi-memory of E, it must be the case that x's apparent memory was caused by someone's, not necessarily x's, having been involved in E.\(^{11}\) I can have a quasi-memory of your party if my apparent memory is caused in some way by your party; perhaps our neurosurgeon can transfer information from your brain into my brain. Quasi-memory does not presuppose personal identity, so we can avoid Butler's charge of circularity if we define personal identity in terms of quasi-memory instead of in terms of memory. Notice that memories are a subclass of quasi-memories; they are quasi-memories in which the someone who was involved in E is the person who has the apparent memory of E. But it is important that I can quasi-remember the experiences or actions of someone else. If I, however, quasi-remember a sufficient number of x's experiences, then I am identical with x, that is my quasi-memories are actual memories.

\(^{11}\)There is a weaker previous awareness condition on quasi-memory than there is on memory (Shoemaker 1970, 271).
But avoiding Butler's charge of circularity is not enough: we want to retain the initial intuitive pull of Locke's theory. I said that many of us believe there to be some intimate link between memory and personal identity: this belief could be explained by the fact that memory presupposes personal identity. Do we have reason to believe that quasi-memory should be part of the analysis of personal identity? Suppose that I quasi-remember a sufficient number of person x's experiences to satisfy the criterion of personal identity, or suppose that I am quasi-memory-continuous with x. This means that my quasi-memories are qualitatively just like memories, that x actually had the experiences that I remember, and that my quasi-memories are causally dependent on x's experiences. Of course, it is true that I can quasi-remember somebody else's experiences: this is precisely the difference between memory and quasi-memory. Yet I think that it is clear that if I had sufficient quasi-memory connections to x, we would have evidence that I am identical with x, that my quasi-memories are actual memories (cf. Shoemaker 1970, 279-280). As Parfit says,

In our statement of our revised...Criterion, we should not claim that, if I have an accurate quasi-memory of some past experience, this makes me the person who had this experience. One person's mental life may include a few quasi-memories of experiences in some other person's life... our criterion ignores a few such quasi-memory connections. We appeal instead to overlapping chains of many such connections (222).

So we can avoid Butler's charge of circularity and still retain the intuition behind Locke's initial account by appealing to the concept of quasi-memory.
More Problems  Even though we have avoided the charges that Locke’s account is circular and that it violates the logic of identity, I do not think that we yet have a plausible criterion of personal identity. Consider the following case: suppose that I am quasi-memory-continuous with person x. However, also suppose that there is no other type of psychological continuity between me and person x. X and I have different beliefs, preferences, ideals, goals, character traits, and mannerisms, and my character traits are not in any way a development of x’s, that is it is not the case that if x had not been the way she was that I would not be the way that I am (cf. Shoemaker 1984 90). In other words, the psychological traits (other than quasi-memory) of x seem to bear no relation, either of causal dependence or of similarity, to my psychological traits. Is quasi-memory alone enough to guarantee identity? Consider the reversed case, where I do stand in these other psychological relations to x but I am not quasi-memory-continuous with x. We normally call this sort of case a case of ‘amnesia’, and “it seems conceivable that someone should survive total amnesia, total loss of memory” (Shoemaker 86). If someone could survive amnesia, then our revised Lockean criterion must be wrong.

Let us return to our intuition that memory and personal identity are intimately linked. If one were asked to expand on this intuition, one might respond with a counterfactual such as the following: if I had not had an experience E then I would not be able to remember E. If it were the case that person x’s apparent memory of experience E was in no way dependent upon the experience of person y, then we
would not take that apparent memory as evidence of the identity of x and y; for this reason, we included a clause about causal dependence in our of ‘quasi-memory’.

Now consider a case where person x shares some psychological trait p with person y. We may have good reason to believe that if person y had not had trait p, then person x would not have had trait p, and it “is precisely when the circumstances are such that evidence of similarity is evidence of such a causal or counterfactual dependence that evidence of similarity is evidence of identity” (Shoemaker 1984 90). Of course, I may have trait p because my mother had trait p and I have been greatly influenced by my mother; yet, it is clear that we would not take this to be evidence that I am identical with my mother. In fact, I may have a great many of my psychological attributes as a result of my interaction with some other person. So the sort of causal dependence between psychological states that is an indication of identity is more than a mere counterfactual dependence. In order for the relevant sort of dependence to be present, the content of the later psychological state must be in some sense a development of the earlier states, a development in which the earlier states determine and control the later states. Also, we must consider such psychological similarity in conjunction with other facts, including how many psychological connections exist; clearly, one or two such connections will not do. So we now need to develop a new criterion of identity, one that considers other psychological connections besides memory connections.
Psychological Reductionism  Psychological Reductionism takes into consideration other sorts of psychological connections besides memory connections. Our first task, then, is to make clear what other sorts of connections we might reasonably suppose help to constitute our identity over time. For example, there is an obvious counterfactual dependence between an intention to do a and the actual later performance of a; I would not have done a if I had not intended to do a. Also, person x’s desires and beliefs can be dependent on the desires and beliefs of person y. For there to be a psychological connection, x need not have the same beliefs and desires as y; rather, x’s beliefs and desires could be causally or counterfactually dependent upon those of y. For example, x may believe that candidate z is the best person for the job because y had certain general beliefs about what makes a candidate the best candidate for a particular job. Similarly, x’s desire to vote for z may be dependent upon y’s desire to vote for candidates of a certain sort. We can find similar relationships if we consider psychological phenomena such as values, ideals, and goals. So psychological connections are not merely, or only, connections of similarity; they are connections of counterfactual dependence. If person y had not had the experiences or traits that he had, then person x would also have had relevantly different experiences or traits (Shoemaker 1984 89-90).

The first criterion which we considered, the Memory Criterion, defined personal identity in terms of direct memory connections: person x is identical with person y if and only if person x has ‘a sufficient number of’ direct memory connections to
We then saw that we needed to appeal to the transitive relation of memory continuity. So Psychological Reductionism should appeal to a parallel notion of *psychological continuity*:

**Psychological Reductionism:** *person x at time t is identical with person y at time t' if and only if x at t is psychologically continuous with y at t'.*

Psychological Reductionism, then, avoids any Reid-like objections, and it represents a significant improvement over the Revised Memory Criterion. Now let us consider some modifications and variations of Psychological Reductionism.

**The Cause of Psychological Continuity**

Consider the following case: scientists wipe away all of my memories, inducing a state of amnesia. Then they give me new memories, memories (or, quasi-memories) that exactly fit the life of Eleanor Roosevelt. This is done by means of a small disk on which they stored some information that was removed from Eleanor Roosevelt's brain. I said above, in discussing the Lockean theory, that we take it to be good evidence that I am the person who did x if I can remember (or, quasi-remember) doing x. In this case, however, although I have an apparent memory of marrying FDR, we may not be inclined to take that as evidence that I am the person who married FDR. If we did not know about the scientist's tampering with my brain, we might be more inclined, although not strongly inclined, to take my apparent memories of Eleanor
Roosevelt’s life to be actual memories. Once we know how I came to have the memories, however, our intuitions about the case change to some extent.

We can construct similar sorts of examples for every sort of psychological trait or state: if the scientist were to give me all of Eleanor Roosevelt’s desires and beliefs by inserting more disks in my brain, would I then be Eleanor Roosevelt? Suppose that the scientist were to make it the case that I am psychologically continuous with Eleanor Roosevelt as she was on the day that she died. Do we want to say that Eleanor Roosevelt now inhabits my body? If we think that it matters how I came to be psychologically continuous with Eleanor Roosevelt, we will think that Psychological Reductionism needs to be modified. Perhaps psychological continuity is only the criterion of personal identity if it has been caused in the right way. For consider the following case: suppose that Eleanor had never in fact married FDR. The scientist might still have given me the apparent memory of having married FDR. So my psychological states seem in some way too independent of Eleanor’s life, and we may then feel uncomfortable in saying that I am identical with Eleanor.

So let us modify Psychological Reductionism:

Psychological Reductionism person x at time t is identical with person y at time t' if and only if (i) x at t is psychologically continuous with y at t', and (ii) this psychological continuity has the right cause.

As Parfit says, we then get three versions of Psychological Reductionism:
These differ over the question of what is the right kind of cause. On the Narrow version, this must be the normal cause. On the Wide version, this could be any reliable cause. On the Widest version, this could be any cause (207).

In order to get a clearer idea of how these three versions of Psychological Reductionism differ, let us reconsider Shoemaker's Brownson case. Recall that Brownson has Brown's brain, Robinson's body, and is psychologically continuous with Brown. On the Widest version of Psychological Reductionism, Brownson is Brown because Brownson is psychologically continuous with Brown and that is sufficient for identity: the Widest version does not care how Brownson came to be psychologically continuous with Brown. The Wide version will agree that Brownson is Brown, but it agrees because Brownson is psychologically continuous with Brown and this psychological continuity is the result of Brownson's having Brown's brain. We can assume that brain transfers are reliable causes of psychological continuity because the brain is the normal cause of psychological continuity. Suppose, however, that no brain transfer was done, and Brownson has just mysteriously become psychologically continuous with Brown. The Widest version would still say that Brownson is identical with Brown, but the Wide version would not agree. Mysterious occurrences do not qualify as reliable causes or procedures.

Now consider a third Brownson case: in this case no brain transfer was done, but the scientists have a reliable way of transferring information from Brown to Robinson's brain, and we again call the resulting individual 'Brownson'. Brownson
is psychologically continuous with Brown, and, on both the Widest and the Wide versions, Brownson is Brown: the scientists’ procedure is reliable. But, on the Narrow version, Brownson is not identical with Brown; such a transfer of information certainly is not the normal cause of psychological continuity whatever the normal cause is. What would the Narrow version say about the case where Brownson has Brown’s brain? Most philosophers who discuss the Narrow version take it to be the case that the brain is the normal cause of psychological continuity and, thus, that Brownson is Brown if he has Brown’s brain.

In the following discussion of Psychological Reductionism, I will focus on the Wide version of the view. Any considerations that I will advance for thinking that the Wide version is unsatisfactory will suggest that the Widest version is completely unsatisfactory, so I will focus my discussion on the strongest, most plausible version of Psychological Reductionism. Narrow Psychological Reductionism is actually a version of what I will call ‘Mixed Reductionism’, and I will discuss that view in the next chapter.

Fission Before we move on, we need to discuss one more revision of Psychological Reductionism, a revision motivated by the following case:

Fission: Suppose that my two sisters and I are in an automobile accident. My body is damaged beyond repair but my brain continues to function. My sisters’ bodies are undamaged but their brains have ceased to function. At the hospital where we are taken after the
accident, the surgeons remove my brain, split it in half, and put each
half in one of my sisters' now debrained bodies. We can assume that
half of a brain is sufficient to guarantee psychological continuity (Parfit
253ff.).

In this case, there are two resulting persons each of whom has half of my brain and
is psychologically continuous with me. All of the versions of Psychological
Reductionism, the Wide, Widest, and Narrow versions, will yield that I am
identical with both of the resulting persons. But identity, unlike psychological
continuity, is a one-one relation. Because it seems clear that the resulting persons
are two distinct persons, I cannot be identical with both of them. So the
Psychological Reductionist concludes that I am not identical with either of the
resulting persons, because there is no more reason to suppose that I am one of the
resulting persons rather than the other. The Psychological Reductionist criterion
now reads:

Psychological Reductionism: \( x \) at time \( t \) is identical with \( y \) at time \( t' \) if
and only if (i) \( x \) at \( t \) is psychologically continuous with \( y \) at \( t' \), (ii) this
continuity has the right, or, a reliable cause, and (iii) continuity has
not taken other than a one-one form.

How plausible is Psychological Reductionism? That is the question to which we
now turn.
3 Can Persons Switch Bodies?

3.1 Introduction

I began my discussion of Reductionism by describing the Physical Reductionist criterion of personal identity. The Physical Reductionist criterion seemed attractive initially because of its simplicity. It reduced the problem of personal identity to one instance of the more general problem about the persistence of physical objects. Then, however, we saw that mere physical continuity does not seem to be enough to guarantee the survival of a person. Robinson was physically continuous with Brownson, and yet we were not inclined to think that Robinson survived as Brownson. And the Body-switch case seemed to show that physical continuity is not even a necessary condition of personal identity. I then described Psychological Reductionism, a view which traces persons via similarity and causal dependence of psychological states, thereby preserving our intuitions about the Body-switch case. But how plausible is Psychological Reductionism? To answer this question, we need to ask how much weight should be given to our intuitions that persons can get new bodies over competing intuitions that we have in other cases that I will discuss in this chapter. We will find that our discussion of identity
conditions forces us to discuss what matters in our survival.

3.2 Williams’s Argument Against Psychological Reductionism

As I have said, the primary and most compelling motivation for the Psychological Reductionist view is our strong intuitions about the Body-switch case. Recall that, in that case, person A goes into one booth and person B goes into a second booth. A scientist then activates his machine. As a result, when the person with A’s body steps out of the first booth, she is fully psychologically continuous with person B, and when the person with B’s body steps out of the second booth, she is fully psychologically continuous with person A. It seems plausible to claim that A and B have exchanged bodies, that A now has B’s body and vice versa. But if that is the correct description of what has happened, then physical continuity cannot be a sufficient or even a necessary condition of personal identity. Rather, psychological continuity seems to be the criterion of personal identity. Because B-body person is psychologically continuous with A, we are inclined to think that B-body person is identical with A. (Similarly for A-body person and B.) Therefore, to challenge the plausibility of Psychological Reductionism, one needs to challenge our very strong intuitions about Body-switch. One needs to show that our intuitions are either confused or should not be given much weight over competing intuitions in other cases.
Bernard Williams tries the first strategy of showing that our intuition about Body-switch is confused. His argument relies on a variant of Body-switch. This case is just like Body-switch except for the addition of a third person C. C steps into a third booth when A and B step into their respective booths. The resulting A-body and B-body persons are just as they were in the Body-switch case. When the person with C's body emerges, she, like B-body person, is psychologically continuous with A. So there are now two persons who are psychologically continuous with A before she stepped into the booth - the resulting B-body person and the resulting C-body person. Do we still want to say that A and B have exchanged bodies? It seems implausible to suppose that that would be the correct description of this case. After all, C-body person has the same claim to being A as B-body person has. Both are psychologically continuous with A. If there were no B-body person, we would be inclined to say that A and C have exchanged bodies. So we have no more reason to suppose that A is B-body person than we have to suppose that A is C-body person. And A cannot be identical with both B-body person and C-body person, because B-body person is clearly not identical with C-body person. Therefore, we should say that A does not survive in the three-person variant of Body-switch.

It seems, at least initially, that the three-person variant presents no problem for Psychological Reductionism. In this case psychological continuity has taken other

---

12 Williams's example is somewhat different, but the relevant parts of it are the same.
than a one-one form; in other words, psychological branching has occurred. After A-body, B-body, and C-body persons step out of their respective booths, there is, as a result of psychological branching, no unique person who is psychologically continuous with A. In our discussion of Psychological Reductionism, we amended the criterion to deal with such cases of psychological fission. The branching clause of the Psychological Reductionist criterion (clause (iii)) yields that A is not identical with either B-body person or C-body person. Thus, the Psychological Reductionist criterion with the branching clause yields an acceptable conclusion, a conclusion that seems intuitively plausible.

Williams claims, however, that Psychological Reductionism does not appear so plausible when we consider the original Body-switch case not independently but in light of our reactions and plausible claims about the three-person variant. Williams agrees with the Psychological Reductionist conclusion that A does not survive in the three-person variant. But, he argues, it is implausible to suppose that A is not identical with the B-body person in the three-person variant and yet still claim, as the Psychological Reductionist does, that A is identical with the B-body person in the original Body-switch case. If we claim that A does not survive in the three-person variant, then it is implausible to suppose that A survives in the original case. Williams claims that "to speak of identity in the simpler case would be at least quite vacuous" (9). After all, all of the intrinsic facts about A, B-body person, and the relationship between A and B-body person are exactly the same in
both cases. To say that A is identical with B-body person in one case but not in
the other really comes to nothing more than saying that C-body person is not
around in one case but she is in the other. So if the claim of identity is to be more
than ‘vacuous’, we should say that A survives in neither case or that A survives in
both cases. Because we do not want to claim that A survives as the B-body person
in the three-person variant, we should conclude that A does not acquire B’s body
in the original case either. We should simply say that A stands in the same
intrinsic psychological relationship to B-body person in both cases, and, in the
three-person variant, she stands in that relationship to C-body person also. If that
is true, then the initial motivation for Psychological Reductionism is undermined.
We should, if Williams is correct, revise our initial reactions to Body-switch.
Those reactions were confused, due to our considering the case in isolation from
the relevant information provided by the three-person variant.

Whereas I started by saying that the branching clause prevents the
Psychological Reductionist from having any difficulties with the three-person
variant, Williams claims that the branching clause is precisely what creates
difficulties for the Psychological Reductionist. He says that the branching clause is
a ‘quite arbitrary provision’ (21) tacked on merely to save the Psychological
Reductionist criterion from yielding absurd conclusions. To see his reasons for
making this claim, we need to recall that the primary motivation for the
Psychological Reductionist criterion is our intuition about the Body-switch case.
That case suggests to us that psychological continuity is the criterion of personal identity. But the three-person variant shows us that psychological continuity, unlike identity, is not a one-one relation. In the three-person variant, A stands in the relation of psychological continuity to both B-body person and C-body person, but A cannot be identical with both B-body person and C-body person. So the Psychological Reductionist criterion requires the addition of the branching clause, because the Psychological Reductionist does not want to have to say that A is identical with both B-body person and C-body person. But then it appears that the branching clause is 'arbitrary', *ad hoc*, tacked on to avoid potential counter-examples to the Psychological Reductionist criterion. As Williams says, "no principle P will be a philosophically satisfactory criterion of identity for Ts if the only thing that saves P from admitting many-one relations among Ts is a quite arbitrary provision" (21). In other words, any criterion that requires a branching clause or something like a branching clause is unacceptable.

Of course, the three-person variant is only a counter-example to a Psychological Reductionist criterion without a branching clause. A Psychological Reductionist might respond that the branching clause is clearly not arbitrary, because it preserves both the logic of identity and our intuitions about the Body-switch case. Identity is a one-one relation, and we have strong intuitions in the Body-switch case that A has acquired B’s body because A is psychologically continuous with B-body person. Therefore, a natural response to Body-switch is to claim that
psychological continuity, *when it takes a one-one form*, is the criterion of personal identity. This response, as I have said, embodies our initial reaction to the case combined with our prior knowledge of the logic of the identity relation. The logic of the identity relation simply constrains how we can apply the psychological continuity criterion. Viewed in this way, the branching clause is not at all an *ad hoc* or arbitrary reaction to cases such as the three-person variant. Rather, such cases are simply examples of its necessity. After all, our initial reactions are about a case where psychological continuity takes a one-one form. Why should we ever have supposed that those intuitions apply to any cases where psychological continuity takes a one-many form?

So Williams owes us more of a story about why we should regard the branching clause as arbitrary. Williams seems to think that any provision which reflects extrinsic facts rather than only intrinsic facts is arbitrary. “For”, he says, “there is a fairly clear sense in which what is true of B[-body person] when it uniquely bears R [psychological continuity] to A is just the same as what is true of it when it non-uniquely bears R to A; the uniquely...makes...no real difference to B/-body person)” (78; emphasis my own). The sense in which what is true of B-body person in Body-switch is also true of B-body person in the three-person variant is that all of the intrinsic facts about B-body person are the same in both cases. Also, all of the intrinsic facts about A and about the relationship between A and B-body person are the same in both cases. Williams’s claim is that the truth of
the assertion that B-body person is identical with A can only depend on intrinsic facts about A, B-body person, and the relationship between A and B-body person. Therefore, a criterion of identity should depend only on intrinsic facts for the truth conditions of claims about personal identity. The branching clause of the Psychological Reductionist criterion clearly involves reference to extrinsic or relational facts. It tells us that whether A is identical with B-body person depends on whether A also stands in the relation of psychological continuity to some C-body person, where C-body person is clearly not identical with B-body person. Therefore, if Williams is right that the truth conditions of identity claims cannot involve extrinsic facts about the individuals under consideration, then Williams is correct in claiming that Psychological Reductionism is an inadequate criterion of personal identity. Any criterion of personal identity that requires a branching clause would be inadequate, because branching clauses are precisely clauses that say that some extrinsic fact is relevant to claims about the persistence of persons.

But why should we suppose that whether A is identical with B-body person can depend only on intrinsic facts about A, B-body person, and the relationship between A and B-body person? After all, there is one sense, as Williams himself admits, in which what is true of B-body person in the original Body-switch case is not the same as what is true of B-body person in the three-person variant. B-body person has different extrinsic properties in the two cases. In the former case, but not in the latter case, he uniquely bears the relation of psychological continuity to
A. But, Williams urges, such a difference in B-body person’s extrinsic properties is not a ‘real difference to B[-body person]’. Only differences that can make a difference to identity constitute real differences, and, Williams claims, extrinsic properties do not constitute such differences. Intrinsic properties, on the other hand, are real differences. But Williams must have some independent notion of what a real difference is, or he will simply be begging the question against any criterion with a branching clause. After all, without some independent notion of a real difference, the Psychological Reductionist can point out that, given that the logic of identity is one-one, uniqueness would seem to be very relevant to identity questions. For example, in order to determine whether a certain man is my husband, one needs to determine whether I stand in the marriage relation to him and only to him, because the spouse relation is a one-one relation (at least in most states). So we need some independent explanation of why extrinsic facts cannot be relevant to the truth conditions of identity claims.

One reason why extrinsic facts might seem irrelevant to identity questions is because, in all normal cases, persons do not cease to exist unless there is some change in their intrinsic properties. Consider how persons typically cease to exist: they die of diseases, as a result of automobile accidents, they get stabbed or shot, etc. In all of these cases, there is bodily failure. Something happens to a person’s body that causes it to cease to function or to support any sort of mental life, let alone a psychologically continuous mental life. The creation of C-body person in
the three-person variant has no such causal impact on A's physical or mental life. After C-body person steps out of her booth, A's body continues to function and A's mental continues on in B-body person just as it would if C-body person were not around. The causal effects of the creation of C-body person are very different than the causal effects of B-body person getting terminal cancer or of someone putting a fatal bullet wound in A's body. But that is just to say that persons never normally go out of existence as a result of the sorts of events described in the three-person variant. Again, though, the Psychological Reductionist will say that given both our intuition about the original Body-switch case and the logic of identity, we should conclude that persons can cease to exist in other than the ways with which we are familiar, as the result of events which are, in their causal impact, extremely different from diseases or stabbings. As I have emphasized before, mere extrapolation from ordinary cases is not sufficient to defeat a metaphysical criterion of identity. If it were, we would have a quick argument against the possibility of a person getting a new body, because persons never do actually get new bodies.

I do think, however, that there is something compelling in Williams's suggestions. It does seem odd to suppose that whether B-body person is identical with A depends on whether C-body person happens to be around. After all, identity is the relation that everything bears to itself and only to itself. Therefore, in order to determine the identity of B-body person, why should we have to check
any facts about some person such as C-body person who is clearly distinct from B-body person? For imagine that B-body person emerges from her booth, but the door to C-body person’s booth gets jammed. B-body person then works feverishly to get the door open. She wants to find out whether C-body person is psychologically continuous with A so that she will know whether she, B-body person, is identical with A. There does seem to be something implausible in supposing that B-body person must know those facts about C-body person before she can know who she herself is. Of course, as we saw in Chapter 2, we cannot always trust the deliverances of introspection to yield knowledge of personal identity. But if B-body person needs to verify the deliverances of introspection as to whether she is A, it does seem odd to suppose that she can do so by checking facts about C-body person who is clearly distinct from B-body person. Also, can whether or not B-body person is identical with A depend upon facts at some extremely remote distance? For imagine that C stepped into a booth in Siberia while A and B entered booths in Idaho. How can the identity of B-body person in Idaho depend upon the intrinsic characteristics of C-body person far away in Siberia? Considerations such as these do seem to indicate that Williams’s suggestions about real differences and intrinsic facts have a compelling motivation.

But, on a second glance, these considerations may only be compelling if we understand a criterion of identity to be a criterion in some epistemological sense. For consider the cases that I just presented. In the first one, we imagine B-body
person trying to open the C booth so that she can find out who she is. I think that what seems odd about this case is that B-body person must know some facts about C-body person before she can answer the question as to whether she is identical with A. Similarly, in the second case, what strikes us as odd is that B-body person in Idaho must know some facts about a person far off in Siberia before she can know the answer to the question as to whether she is identical with A. We always manage to answer identity questions without any investigation into such remote facts. Commonsense tells us that such facts are irrelevant to our knowledge of personal identity. If Williams is right in claiming that “[t]o enable us to answer such questions is the point of a criterion of identity” (24), then it does look as though the Psychological Reductionist criterion simply fails in certain cases. We can know all of the facts about A, B-body person, and the relationship between A and B-body person and still not be able to determine whether A is identical with B-body person. We can know the entire history of B-body person and still not be able to determine whether B-body person is A. Such results certainly do not respect our intuitions about our judgments of identity and what facts we need to know to know who someone is. Psychological Reductionism seems to leave open the possibility that if we use only the evidence which commonsense tells us is at all relevant, we will make wrong judgments about personal identity.

I do think that certain sorts of epistemological considerations should influence our evaluations of various criteria of personal identity. Recall, for example, my
argument against Cartesian Non-Reductionism. I claimed that the Combined Spectrum argument showed that, if Non-Reductionism were true, it seems that it would be impossible, in principle, for us to ever have knowledge about personal identity. The Non-Reductionist fails to give us an adequate account of how the facts that really determine a person’s identity are related to the facts that we typically use to determine a person’s identity. Also, it seems that we never could have access to those facts which really determine a given person’s identity. So if Non-Reductionism were true, we would be forced to accept a general scepticism about our judgments of identity. We would never have knowledge of personal identity. If we do make correct judgments about personal identity, they are simply lucky guesses, pure happenstance. I claimed that this was a good reason to reject Cartesian Non-Reductionism and to look for a plausible version of Reductionism.

Even though there is an epistemological worry about the Psychological Reductionist criterion, it is not the same sort of worry that we faced concerning the Non-Reductionist criterion. The Psychological Reductionist criterion certainly does not make it impossible, in principle, for us to tell when a given individual has survived. Neither does it make it the case that we can never have access to the facts which determine a person’s identity. We will sometimes have to travel long distances to determine who a given individual is. The Psychological Reductionist criterion simply makes it the case that it may be very difficult to know where to accumulate all of the facts relevant to a particular judgment of identity. But such
cases do not undermine the judgments that we make in normal cases. The Combined Spectrum argument showed that, if Non-Reductionism were true, we could never have knowledge of personal identity. The cases that I presented above as posing a problem for Psychological Reductionism show only that it could be extremely unlikely that we would make correct identity judgments in certain cases of psychological branching. However, we can be reasonably, even extremely confident, in ordinary cases, that psychological branching has not occurred, and, in such cases, we have access to the sorts of facts, facts about psychological continuity, that determine the truth of claims about personal identity. The mere possibility of making an incorrect judgment does not commit us to any general scepticism as did the Non-Reductionist criterion. The fact that a criterion of identity leaves open the possibility that we are not infallible in our judgments of identity is certainly no reason to reject it.

Of course, the Psychological Reductionist must, in certain branching cases, appeal to facts which commonsense denies to be at all relevant to determining whether, for example, B-body person is identical with A. I do not think, however, that this is any reason to reject Psychological Reductionism. After all, here is where commonsense may support some version of Non-Reductionism. For we can imagine cases where it would also seem very odd to require the sorts of evidence required by any form of Physical Reductionism to answer a question about personal identity. For all introspection can reveal, I might be a replica of Diane. In
order to find out, I would need to trace the history of my body. Of course, in all probability I am not a replica, and, therefore, if Physical Reductionism were true, I could have knowledge of my own identity and of the identity of other persons. Most of us, I think, would find it odd to suppose that anything more than introspection was necessary, except in cases of insanity or amnesia, in order to discover our own identity (cf. Johnston “Reasons and Reductionism” 11ff.).

Again, however, we have to be careful not to merely extrapolate from the ordinary run of cases. As I said above, all of the ordinary cases of a person’s going out of existence involve some sort of bodily failure as a result, perhaps, of disease or a bullet wound. Because a person’s mental life always ceases when the body stops functioning, we normally just check if a person’s body is still working when we want to know if that person still exists. Similarly, a person’s mental life normally does not continue on in more than one body; people do not normally undergo fission. As long as a metaphysical criterion allows that what we check in normal cases is generally sufficient in those cases to determine whether a person has survived or whether a person is who she thinks she is, then I do not think that epistemological considerations are sufficient to defeat the criterion. Because I also think that such considerations are at the root of Williams’s worry about the arbitrariness of the branching clause and of the worry about real differences and extrinsic facts, I do not think that Williams’s argument that our intuition about Body-switch is confused is a successful argument. Williams’s argument does not
defeat Psychological Reductionism.

3.3 Another Difficulty for Psychological Reductionism

I said above that in order to defeat Psychological Reductionism one would need to show either that our intuitions about Body-switch are confused or that they should be given little weight in comparison with competing intuitions that we have about other cases. In the previous section, I showed why I think that Williams's attempt at the first strategy failed. Now I want to suggest that the second strategy might be more successful, because I do think that there is a sort of case that raises serious difficulties for Psychological Reductionism. In these cases, our intuitions strongly suggest that physical continuity is not at all irrelevant to questions about personal identity.

Suppose that I am sitting in my office working on April 23, 1992, at noon. Unbeknownst to me, scientists on board the Space Shuttle have developed a method whereby they can clone any person whom they choose. This cloning takes place via a new laser beam that locks on to a person and then reproduces the person's physical and psychological composition. At noon on April 23, they create a replica of me who has a body exactly similar to mine and is fully psychologically continuous with me just before the replication. Psychological branching has occurred, because after the replication there are two persons, the replica on the Space Shuttle and the person in my office, who are both psychologically

91
continuous with me before the replication. Thus, according to the Psychological Reductionist criterion, I die as a result of the replication that takes place on the Space Shuttle. Anyone who enters my office will see not me but a new person who came into existence at noon on April 23.

Consider another similar case. Suppose that scientists have developed a new drug that can cause radical personality changes, and the scientists are able to determine what changes will take place. This drug contains vital information encoded in bio-chemical material, information that will cause the person who takes the drug to become psychologically exactly like the person from whom the bio-chemical material was taken. The scientists take the relevant bio-chemical material from my body. They are able to do this without my being aware of what is happening. The scientists then administer the drug to Madonna, thereby causing her to become psychologically continuous with me before the administration of the drug. Again, psychological branching has occurred. Both the resulting person with my body and the resulting person with Madonna’s body are psychologically continuous with me before Madonna’s taking of the drug. So, according to Psychological Reductionism, I die as a result of Madonna’s ingesting this new drug.

I think that these cases raise two problems for Psychological Reductionism. The first problem I will only mention here, and then discuss it further in the next chapter. It seems very odd to suppose that the sorts of events described in the
Space Shuttle and Madonna cases could cause my death. How could the creation of a replica of me aboard the Space Shuttle kill me? How could Madonna’s ingesting a particular drug kill me? It certainly is a counterintuitive feature of Psychological Reductionism that it yields that I die in these cases. The Psychological Reductionist, however, has a response to this worry about death. He will say that it seems counterintuitive to suppose that I die in these cases only because we typically think that survival matters. For in order to reach the conclusion that the events in these cases could not possibly kill me, one would reason in the following way: It just seems wrong to suppose that it should matter to me whether a replica of me is made on board the Space Shuttle or whether Madonna takes a dose of some drug. But if these events cannot matter to me, and my death is an event that would matter greatly to me, then it cannot be the case that the Space Shuttle or Madonna events can kill me. The Psychological Reductionist responds that we should reject the claim that my death is always an event that would matter to me. As long as there is some future person who is psychologically continuous with me, then I cannot rationally care whether I survive. Psychological continuity, not identity, is what fundamentally matters in my survival. As I have said, however, I will pursue this issue of whether survival matters in the next chapter.

Without addressing the issue of what matters, however, the Psychological Reductionist might claim that my use of the term ‘death’ in these cases is tendentious, because ‘death’ has certain connotations from ordinary cases. As I
said in the previous section, ordinary death involves some sort of bodily failure. So, the Psychological Reductionist will say, of course it seems odd to suppose that the events described in the Madonna and Space Shuttle cases could cause my death, because these events do not cause my body to cease to function. It may be that our concept of death necessarily involves the idea of a person’s body ceasing to function. Therefore, in the absence of such bodily failure, it is odd and, indeed, mistaken, to speak of my death. But that does not imply that I survive as one of the resulting persons in these cases. What we should take it to imply, the Psychological Reductionist can say, is that a person can cease to exist in other ways than by dying. One such way is to be replicated or to undergo some other sort of psychological branching. Because persons in all normal cases only go out of existence by dying, we really should expect to have no definite intuitions about what sorts of events might cause me to cease to exist in some other way than by dying. To say that it is odd to suppose that I die in Madonna and Space Shuttle, then, comes to no more than to say that the creation of a replica is very different from contracting terminal cancer or getting hit by a bus.

Even if we accept this point, however, I think that the Psychological Reductionist will have difficulty with the Space Shuttle and Madonna cases. Let us recall the reason why the Psychological Reductionist, in the face of psychological fission, added the branching clause that states that, in a case of branching, neither resulting person is identical with the original person. In the
three-person variant of the Body-switch case, we said that both B-body person
and C-body person had equal claim to being A because each was fully
psychologically continuous with A; therefore, the best option seemed to be to say
that A did not survive as either B-body or C-body person. But in neither the
Space Shuttle case nor in the Madonna case does it seem to be true that each
resulting person has an equal claim to being me. The fact that the person in my
office has my body seems to settle the matter in her favor. The replica seems to be
just a very good copy of me. And Madonna’s taking of the drug simply makes it
the case that there is now someone around who has Madonna’s body and is
psychologically continuous with me. It seems implausible that I do not survive as
the person with my body in these cases. We do not seem to have sufficient reason
to say that I have ceased to exist either by dying or in some other way.

What can the Psychological Reductionist say about these cases? He might
respond that these cases suggest a modification rather than a complete rejection of
Psychological Reductionism. In a case where my body is destroyed as the replica is
created, it does seem that the replica has a strong claim to being me. In order not
to rule out that possibility, we might admit physical continuity as a ‘tie-breaker’ in
certain cases of psychological branching, but not yet claim that physical continuity
is a necessary condition of personal identity. Thus, the Psychological Reductionist
criterion would say that person x is identical with person y if and only if (i) x is
psychologically continuous with y, and (ii) when psychological branching has
occurred, only if x is physically continuous with y. Such a modified Psychological criterion would yield plausible results in the Space Shuttle and Madonna cases and still allow persons to exchange bodies as in Body-switch or get new bodies as in simple replication.

The problem with this response is that it seems too ad hoc. Why should we suppose that physical continuity can function as a tie-breaker and yet is not a necessary condition of personal identity? Perhaps it is more plausible to accept that physical continuity is a necessary condition of personal identity and deny that persons can get new bodies. Of course, accepting these claims would entail a rejection of our initial intuition in Body-switch. I want to now consider a Mixed Reductionist view so that we can decide whether it has sufficient advantages to force us to view our original intuition in Body-switch as being outweighed by competing intuitions about Madonna and Space Shuttle.

3.4 The Mixed Reductionist Criterion

The Mixed Reductionist criterion says that person x is identical with person y if and only if (i) x is psychologically continuous with y, and (ii) x is physically continuous with y. This criterion avoids any problems involved with psychological branching. For example, neither in the Body-switch case nor in its three-person variant does A survive, because none of the resulting persons in either case is both psychologically and physically continuous with A. The question that we need to
address is whether there is any parallel problem with physical branching. Could there be at some time two distinct persons, B and C, such that both are fully physically continuous with person A at some earlier time?

It does seem, at least at first, quite plausible to suppose that we could have physical branching just as we have psychological branching. As Williams says, “It is possible to imagine a man splitting, amoeba-like, into two simulacra of himself” (23). We imagine a man’s body splitting in half. As it does so, each cell generates a copy of itself so that we end up with two bodies that are indistinguishable from the original body, each of which has half of the cells which the original body had. It does seem right to say that each resulting person is physically continuous with the original person. For imagine a case where half of a man’s body is severely burned. Scientists have developed a process where they can lop off the burned cells and induce the remaining cells to duplicate. Surely a man could survive such a process; therefore, the Mixed criterion should count the resulting body as being fully physically continuous with the original pre-burning body. But then we must count each of the resulting bodies in the splitting case as being fully physically continuous with the original body. Thus, physical splitting does seem to be a real possibility as much as is psychological splitting.

Williams claims, however, that we should not equate physical branching with psychological branching. For, Williams says,

a thorough application of the [physical] criterion would itself reveal the existence of the reduplication situation, and so enable us to answer
(negatively) the original identity question. To enable us to answer such questions is the point of a criterion of identity. Thus, in this case, but not in the others [the cases of psychological branching], the logical possibility of reduplication fails to impugn the status of the criterion of identity...these considerations perhaps suffice for us to say that in a case of fission, such as that of an amoeba, the resultant items are not, *in the strict sense*, spatio-temporally continuous with the original (24; emphasis my own).

Here we can see Williams clearly relying on the epistemological sense of a criterion. If the point of a criterion is to give us a procedure for determining when a person survives, then the Mixed Reductionist criterion probably does fare better than the Psychological Reductionist criterion. We will be more likely to be able to discover when physical splitting has occurred than we will be to discover when psychological branching has occurred. But, to repeat the point that I made when discussing psychological branching, we should not evaluate a criterion of personal identity merely on the basis of how often it makes it easy for us to answer questions of identity. As long as a criterion allows that we can have knowledge of personal identity in the ordinary range of cases, then it cannot be defeated by epistemological worries. So I do not think that Williams has a good reason for making a distinction between the significance of psychological branching and the significance of physical branching.

Also, it is somewhat mysterious what Williams means when he says that "in a case of fission, such as that of an amoeba, the resultant items are not, *in the strict sense*, spatio-temporally continuous with the original". The relationship between the original man A and the resulting person B is intrinsically exactly the same in
both the first burning case and in the splitting case. If A is not, strictly speaking, physically continuous with B in the second case, then we should conclude that he is not, strictly speaking, physically continuous with, and, therefore, not identical with B in the first case. To claim that strict physical continuity depends on extrinsic facts such as whether C is around is simply to build a branching clause right into our conception of physical continuity. The Psychological Reductionist could have done the same thing and claimed that we do not have 'strict' psychological continuity in the three-person variant of the Body-switch case. Nonetheless, it just seems more straightforward to allow that there can be both physical and psychological branching and add branching clauses to both the Psychological and Mixed Reductionist criteria.

Williams might respond by saying that there is a reason why we can make a distinction between strict physical continuity and non-strict physical continuity but we cannot make such a distinction between strict psychological continuity and non-strict psychological continuity. (What follows is not exactly Williams's argument, but rather a modification of some remarks that he makes about identity and exact similarity.) To see why, recall that, in my earlier discussion of physical continuity, I suggested that its use as a criterion of personal identity is motivated by the intuitive idea that a person persists as long as his body persists, i.e. that same body implies same person and vice versa. The concept of physical continuity is intended to capture the intuitive idea of having the same body, that is, it is
supposed to be part of a reductive analysis of the concept of having the same body. So we have strict physical continuity in the cases where we have the persistence of one and the same physical human body. Now it is quite clear that in the physical splitting case, we must say that two new bodies are created out of the original body, because we have no reason to suppose that one of the resulting bodies rather than the other is the original body and it is too odd to claim that we have, after the splitting, one spatio-temporally dislocated body. And, yet, surely the resultant bodies are, in some sense, physically continuous with the original body. So now we have a distinction between strict and non-strict physical continuity.

The concept of psychological continuity, on the other hand, does not, and should not capture the notion of 'same psychology'. I may have exactly the same psychology as my doppelganger on Twin Earth, and yet I am not psychologically continuous with my doppelganger because her psychological states are not causally dependent on mine. Similarly, my self of twenty years hence will not have exactly the same psychology as I now have, because my psychological attributes will change over time. So we do not want to say that we only have strict psychological continuity in the cases where we have the persistence of one and the same psychology. But, if we do not, then we do not have the resources to distinguish strict and non-strict psychological continuity.

I do not think that this sort of response is really going to help Williams out. It seems to me that physical splitting cases simply show that the concept of physical
continuity cannot capture our intuitive notion of same body. For in the case where half of the man's body is burned away and then regenerated, we are inclined to say that he survives with the same body, that there is physical continuity in the strict sense. But then it is difficult to see why we do not have strict physical continuity in the splitting case; after all, one of the resulting bodies in the fission case is exactly the same as the body in the burning case. And yet, clearly, neither body in the fission case can be the same body as the original body. So in the burning case the man's having the same body as the original body just comes to his being uniquely physically continuous with the original man, where physical continuity must be of the same sort that we have in the fission case.

So we do not want to lay any more stress on the notion of 'same body' than we do on the notion of 'same psychology'. The reductive analysis of both concepts simply involves certain continuities in change of either physical matter or psychological attributes when those continuities take a one-one form. I think that, in attempting to use some notion of strict physical continuity, Williams may be defending a form of Physical Non-Reductionism. A Physical Reductionist, as we saw, will deny that we can give a reductive analysis of what it is to have the same body through time. In normal cases it is clear that persons keep the same body, and, Williams wants to say, it is clear that bodies do not persist through physical splitting. If we interpret Williams as a Physical Non-Reductionist, then we can see his earlier arguments as being arguments not simply against Psychological
Reductionism but against any version of Reductionism because all versions require branching clauses.

But the Physical Non-Reductionist faces a real problem with physical splitting cases. Williams agrees with the Mixed Reductionist that neither resulting person in such a case is identical with the original person. The Mixed Reductionist says this because it seems that both persons, on her view, have equal claim to being the original person. The Physical Non-Reductionist, on the other hand, has no reason to suppose that both persons have an equal claim; it may be that one of the persons has the original person’s body. If having the same body is not just to stand in one-one physical continuity, the Physical Non-Reductionist seems unable to tell us why some one of the resulting persons does not have the original body in the splitting case. The Physical Non-Reductionist is going to have some serious epistemological worries, and he is going to have a difficult time explaining why bodies survive certain changes but not others. In comparison, the sorts of epistemological worries that beset Reductionism as a result of branching clauses do not seem particularly troubling. So Williams’s considerations are not persuasive whether they are intended to support some version of Physical Reductionism or to support Physical Non-Reductionism.

The Mixed Reductionism, then, cannot avoid the need for a branching clause. (Notice that he has the choice of adding either a psychological branching clause or a physical branching clause.) But then we need to wonder what motivation there
is for Mixed Reductionism other than its being able to account for our intuitions in the Space Shuttle and Madonna cases. Also, Mixed Reductionism faces the same problem with the Brownson case that the Physical Reductionist faced with that case. Recall that, in that case, Brownson has Robinson’s body, Brown’s brain, and is psychologically continuous with Brown. Our intuition is that Brownson is Brown and not Robinson, and we would think that Brownson is Brown even if there happened to be someone else around who was also psychologically continuous with Brown. Now it is true that in the Madonna and Space Shuttle cases the person in my office stands in the same relation to me as Brownson stands in to Brown, that is, she is psychologically continuous with me as the result of having the same brain as me. (‘Same brain’, of course, can be analyzed in terms of non-branching psychological continuity.) So we might take Madonna, Space Shuttle, and Brownson to be evidence for what I called, in the previous chapter, Narrow Psychological Reductionism. The Narrow Psychological Reductionist claims that person x is identical with person y if and only if (i) x is psychologically continuous with y, (ii), this psychological continuity has the normal cause, continuity of the brain, and (iii) this psychological continuity has not taken other than a one-one form.

When we discussed Brownson above, we saw that appeal to the brain is no more than an appeal to psychological continuity. Why should we suppose that psychological continuity must have its normal cause? The appeal to normality
should make us wonder whether we are relying too heavily on an extrapolation from ordinary cases. Death is tied to bodily failure, and, more specifically, to brain failure. So our intuitions in Brownson and Space Shuttle may be dictated by our sense that if the brain is functioning, the original person cannot be dead and must follow her or his brain. But imagine that I have inoperable brain cancer. I elect to have my neurosurgeon remove my brain, extract the information from it, put the information in a new brain, and then put this new brain in my head. Doesn’t this procedure save my life? Do I really need that particular brain matter in order to survive? If I could have a brain matter replacement operation, then we should reject Narrow Psychological Reductionism. At this point, we might try to think of a way to accommodate intuitions without an appeal to any sort of physical continuity. What differentiates the person in my office from the replica on the Space Shuttle other than the fact that she is physically continuous with me?

### 3.5 A Modified Psychological Reductionism

In this section I want to consider the possibility of offering a purely psychological account of personal identity that will accommodate our intuitions about Madonna and Space Shuttle. To begin, consider the following pair of cases. In the first, I decide that I want to overcome my fear of flying. I therefore decide to attend a series of classes designed to help people overcome a fear of flying. In these classes various sorts of behavior modification techniques are used, and, as a result, after
attending the classes, I am able to fly without undue anxiety. In the second case, I need to undergo brain surgery. While the neurosurgeon is operating, he decides, unbeknownst to me, to do some 'rewiring' that results in my no longer being afraid to fly.

In both of the above cases, I undergo a beneficial psychological change. But there is clearly a difference in the two cases in the way that this psychological change is brought about. In the first case, I decide to bring about the change and deliberately undertake actions that lead to the change. The behavior modification techniques assist me in bringing about these changes, but I interact with this external partial cause of the change. I retain at least some control over the change. In the second case, however, I am completely passive in the bringing about of the change; in fact, I have no knowledge that the change is occurring. The change is in no way a result of my will or intention, it is something in which I have no active role. Let us call this second sort of cause of psychological change an 'external imposition', where an external imposition is such that when it effects a certain change in psychology it destroys a psychological connection. For notice that if the scientist were to do sufficient changing via his process of 'rewiring', we would arrive at a case wherein there is sufficient reduction of connections to make it such that I do not survive in that case. In the first case, however, the change is itself a strong psychological connection. The relationship between my intending to effect the change and my taking action to bring about the change establishes a strong
psychological connection between me and my future self.

Now notice that external impositions are such that they need not interfere with psychological connections, because psychological connections involve not only relations of appropriate causal dependence but also relations of similarity. So if the scientist’s rewiring caused me to maintain and also increased my fear of flying, this external imposition would actually strengthen a connection between me and my future self. Thus, an external imposition could maintain the same degree of psychological similarity that existed before the imposition, but not the same degree of causal connectedness in so far as interference by an external imposition will always weaken causal connections.

Let us return to the Space Shuttle case. The replica in this case is created via a scanning process that results in cloning. This process is such that if it were to effect changes, it would destroy psychological connectedness; therefore, this process constitutes what I am calling an external imposition. It is quite clear, on the other hand, that the person in my office is psychologically continuous with me in such a way that no external imposition is involved. The replica’s being just like me is not a result of any intention or deliberation on my part; I am completely passive in the creation of the replica, while the maintenance of continuity with the person in my office is something over which I have some control. I think that we should continue to speak of continuity with respect to the replica because there are quite a few connections of similarity between me and the replica; however, the causal
connection between similar psychological states would not be in itself enough to establish a psychological connection, whereas the causal connection between similar psychological states is itself a psychological connection between me and the person in my office. Thus, the replica is not as strongly continuous with me as the person in my office is, because the person in my office bears both connections of similarity and appropriate causal connections to me, while the replica bears only connections of similarity to me.

Why should we suppose that lack of any causal connections not involving an external imposition will destroy personal identity, that is, why suppose that I survive as the person in my office rather than as the replica? Is there any motivation for such a modified Psychological Reductionist account of personal identity? Notice that it never seemed to be a plausible option to offer a psychological criterion that claimed that person x is identical with person y if and only if x and y are exactly the same psychologically. We all recognize that persons survive drastic psychological changes throughout the course of a lifetime. These psychological changes do not simply happen to a person; rather, persons initiate changes in themselves and change in response to their environment. This ability to adapt and make reasoned choices about change is an important part of our concept of a person as being not just a subject of experiences but also an agent. So it is plausible to think that non-branching psychological continuity, if it is to be the criterion of personal identity, must involve appropriate causal connections that
preserve the centrality of agency to personal identity. The sort of psychological continuity that involves too many external impositions seems to violate this requirement. So, the Psychological Reductionist will say, there is a purely psychological motivation for claiming that I survive as the person in my office in both the Madonna and Space Shuttle cases.

Nonetheless, there are difficulties with this new view. Imagine a case where I decide to be replicated by a process that simultaneously destroys my brain and body. We can even imagine that I am completely in control of the entire replication process; there are no scientists involved. It seems that the modified Psychological account yields that I survive in this case, because the external cause of continuity is not independent of my will; in fact, it is completely dependent on my intentional decisions and actions. But now suppose that the replication succeeds but my body is not destroyed as I had intended it to be. What does the modified Psychological view say about this case? Surely not that I survive as the replica even though I had intended to do so. But it has no reason to suppose that I survive as the person with my body because mere physical continuity is irrelevant on this view. Could it be that my wanting to survive as the replica rather than the replication's being out of my control could make the difference between my ceasing to exist and my not ceasing to exist? Surely I do not have that much power over my own persistence. At the least, the modified Psychological view has difficulty with a case such as this one.
So far none of the criteria of personal identity that we have considered seems entirely satisfactory. They seem either unmotivated or they yield the conclusion that I do not survive as some future person when we are strongly inclined to think that I must survive as that future person. I suggested above that some of our initial intuitions about cases might be clouded by the connotation that ‘death’ carries of bodily or brain failure and that we should try to use more neutral language. But perhaps ‘survival’ carries certain connotations just as ‘death’ does. In all normal cases, a person’s mental life continues only if she survives. Most of us think that the continuance of our mental lives is valuable and that it is rational to care that it continue; therefore, it seems rational for us to care that we survive. But perhaps it is really the continuance of my mental life, not survival per se that should concern me. If so, then in cases like Space Shuttle where it is clear that my mental life continues, I should be more open to the possibility that I do not survive in those cases; after all, we would still have what matters in these cases. Then the problem for the Modified Psychological account that I presented in the last paragraph would not appear very troublesome. Perhaps I do not survive in the case where I decide to be replicated but my body is not destroyed, but I need not be concerned about my failure to survive. After all, what I wanted was for my mental life to continue in the replica and I have accomplished that goal. Can I really care that my mental life also continues in my body? Let us turn, then, to a discussion of what matters in survival.
4 Should We Care If We Get New Bodies?

4.1 Introduction

Most of us think that it should matter to us whether or not we survive. Thus, when we consider a particular situation such as that described in the Space Shuttle case and decide that the events described could not possibly matter to us, we then conclude that the described events are not such as to be able to affect whether we survive. Perhaps, however, we should conclude instead that we cannot, in certain cases, rationally care whether or not we survive. If we accept such a conclusion, we should shift our focus from questions about whether we are identical with some future persons to questions about how we should regard those future persons, how much and what sort of concern for those persons is appropriate. Is special concern required in some cases even though I do not survive in those cases? I will begin by considering what is involved in special concern for one’s future self.

4.2 Future Selves and Fission Products

Commonsense holds that I have offered sufficient justification for my special concern for my future self when I point out that she is my future self, that she is identical with me. As Sidgwick says,

\[\text{[i]t would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently ‘I’ am concerned with the quality of my existence as}\]

110
an individual in a sense, fundamentally important, in which I am not concerned with the quality of existence of other individuals (498).

What does it come to for me to be ‘concerned with the quality of my existence...in a sense, fundamentally important, in which I am not concerned with the quality of existence of other individuals’? It means, among other things, that I have many private projects, projects which are such that their formulation involves an essential reference to me. For example, I want that I finish my thesis, that I see the Mona Lisa, that I spend time with my friends, that I support my family, etc. For these projects to be realized, I must survive and realize them. My desire that I finish my thesis will not be satisfied unless I survive and finish the thesis. If any of these private projects is justified, and they do seem justified, then I am justified in wanting to survive, in wanting that there is some future self such that she is my future self.

But notice that Sidgwick asserts that these claims about future-directed self-concern depend in some way upon the commonsense belief that ‘the distinction between any one individual and any other is real and fundamental’. Sidgwick himself wonders whether this belief is undermined by the Humean account of personal identity. More recently, philosophers such as Parfit have claimed that the belief is undermined by Reductionist accounts of personal identity. Parfit argues that certain cases show that I cannot, in those cases, rationally care that I do not survive. What matters in my survival is not identity, but, rather, psychological continuity.
To see why Parfit makes this claim, we need to look at cases where certain conditions of my survival hold but some other condition does not. Suppose that I have a degenerative and debilitating disease which has infected my body and half of my brain. My surgeon removes the other half of my brain and puts it in an available functioning debrained body and then destroys the other half of my brain along with my body. If we assume that half of a brain is sufficient to guarantee psychological continuity, any Psychological Reductionist criterion yields that the resulting person, call her A, is identical with me. But now consider a second case. Suppose that the surgeon were to realize, just after he has transplanted half of my brain into A, that the other half of my brain is not diseased. He then transplants it into another available functioning debrained body and then destroys the rest of my body. Call this second resulting person B. Since we no longer have non-branching psychological continuity, neither A nor B is identical with me according to the Psychological Reductionist.

Suppose that the Psychological Reductionist is correct in claiming that I survive in the first case but not in the second case. How should I then regard my fission products A and B? Should it matter to me that neither A nor B is identical to me? Since neither A nor B is identical to me, none of my private projects will be realized if I undergo fission; if either A or B finishes my thesis that is not sufficient to satisfy my desire that I finish my thesis. Thus, prima facie, it would

---

13I will only distinguish Psychological Reductionism from the modified Psychological account discussed earlier in this chapter when the two views yield different conclusions.
seem rational for me to care that I survive rather than undergoing fission, because if I undergo fission, none of my private projects can be realized. Therefore, I seem rational in preferring that only one half of my brain is successfully transplanted.

But notice that my relation to A in the case where only half of my brain is transplanted (the case where A is me) is intrinsically the same as my relation to A in the fission case. Each of them is psychologically continuous with me as a result of having half of my brain. However, in the first case, if A finishes my thesis, my private project will be realized, whereas, in the fission case, if A finishes my thesis, my private project will not be realized. The only reason why my private project is not realized in the fission case is because B is also around. But can I rationally care that B is also around? If not, then I cannot rationally care whether my private projects are realized. I should not care whether my thesis is finished by the A of the first case or by the A (or B) of the fission case. Thus, I should have not private projects, but, rather, quasi-private projects, where a quasi-private project is a project such that its formulation involves an essential reference to someone who is psychologically continuous with me.\(^{14}\) Therefore, I cannot rationally care whether I survive but only that there be someone in the future who is psychologically continuous with me whether or not she is identical with me.

In the previous chapter, I considered Williams’s claim that whether I survive as A can only depend on the intrinsic facts about me, A, and the relationship

\(^{14}\)The notion of a quasi-private project is, in some respects, parallel to Shoemaker’s notion of a quasi-memory.
between me and A. So Williams would argue that, because I do not survive as A in the fission case, then I do not survive as A in the case where only half of my brain is transplanted and the other half is destroyed; therefore, my private project is realized in neither case. However, I rejected Williams’s argument, claiming that it traded on irrelevant epistemological considerations. Now notice, however, that Parfit’s argument about what matters has a structure similar to Williams’s argument about survival. Parfit claims that because the intrinsic facts about me, A, and the relationship between me and A are the same in both cases, and we have what matters in my relationship to A in the case where only half of my brain is transplanted, then we must have what matters in my relationship to A (and to B) in the fission case. Should we suppose that extrinsic facts cannot affect what matters in my relationship to some future person even though they can affect whether I am identical with that person?

I think that it is *prima facie* more plausible to suppose that considerations about extrinsic facts are irrelevant to what matters than it is to suppose that they are irrelevant to questions about our identity. For imagine the fission case with a slight variation. Suppose that the scientist transplants one half of my brain into A and removes the other half of my brain and ships it to Siberia, where it is transplanted into B. A completes my thesis without ever coming to know about B. As I look ahead at my future, it does seem odd to suppose that I should care about whether B exists when it is entirely possible that I know nothing about B.
Here Williams's epistemological worries seem very important. Is it really plausible to suppose that I need to know about B in Siberia before I can decide what sort of and how much concern I should have about A and A's projects? If it is true that extrinsic facts are relevant to my survival but not to what matters, then it must be the case that what matters can come apart from survival.

A Mixed Reductionist can make a similar argument about cases of physical fission. First, we imagine a case where half of my body is burned away. My remaining cells reduplicate and I end up with a body that is just like and is physically continuous with the one that I started with. I survive this process, so the resulting person A can realize my private project that I finish my thesis. But now we imagine a second case in which the other half of my body is not destroyed but itself reduplicates to create a body indistinguishable from and physically continuous with my original body. Now if A finishes my thesis, my private project that I finish my thesis will not be realized because A is not identical with me. The relationship between A and me is intrinsically the same in both cases. So how can it matter to me that B is also around? The Mixed Reductionist will conclude that what matters is psychological and physical continuity, not identity. So I should have not private projects, but mixed-quasi-private projects. In other words, I should want that someone physically and psychologically continuous with me realize my projects whether or not she is identical with me.
The Psychological Reductionist argues that we have what matters in both the burning case and the physical fission case, because in both cases there is psychological continuity. But he also claims that we have what matters in the psychological fission case, and that conclusion depends on the claim that I survive if only half of my brain is transplanted. The Mixed Reductionist rejects this latter claim. So let us look at another case to see if it can shed light on these competing conclusions about what matters.

4.3 Replication and the Branch-line Case

So let us consider a case of replication. Suppose that scientists can create replicas of persons by using the laser scanning device described in the Space Shuttle example. This laser scanner transmits to a waiting body the original person’s psychological states and dispositions. The waiting body was created from entirely new matter but looks quite a bit like the original person’s body. Upon duplication the original person’s body and brain are completely destroyed. Would I survive this process of replication? In other words, would my replica be me? The Psychological Reductionist would answer yes to both questions. It seems that the modified Psychological account would say yes if I willingly consent to be replicated. But now suppose that my brain and body are not destroyed in the

---

15 I am going to consider Mixed Reductionism rather than Narrow Psychological Reductionism as the most plausible view with a physical continuity clause. As I said above, I think that Narrow Psychological Reductionism is an unstable compromise between Mixed Reductionism and some purely psychological account.
replication process. So there are two resulting persons, the replica and a person who is physically continuous with me, who are both psychologically continuous with me, while only one of them is physically continuous with me. Psychological branching (but not physical branching) has occurred. In this case, the Branch-line case, the Psychological Reductionist will, then, say that I do not survive, because there is no unique person who is psychologically continuous with me.

This is the Psychological Reductionist conclusion that I used to motivate Mixed Reductionism. In the Branch-line case I think that most of us have a strong intuition that I survive as the person who is physically continuous with me. How can the creation of a replica possibly cause me to cease to exist? *Prima facie*, it certainly seems that I must survive as the person with my body. The Psychological Reductionist, however, may respond that it only seems that the creation of the replica could not affect whether I survive because we think that I should care whether I survive. When we seem convinced that I survive this process of replication, we are really convinced of the fact that I should not care whether I am replicated. If we claim that identity does not matter, then the conclusion in Branch-line does not seem so counterintuitive. Perhaps I do not survive, but in this case my ceasing to exist is not an event about which I can be rationally concerned.

There is a difficulty for the Psychological Reductionist in so far as our intuition is not just that I survive in Branch-line, but that I survive as the person who is physically continuous with me. Similarly, I think that most of us also have the
intuition that it is rational for me before the replication to care more about the person who is physically continuous with me than about the replica. Suppose that, before the replication, I am asked to decide which person, the one who is physically continuous with me or the replica, I would prefer to have undergo five minutes of extreme pain. Does it seem irrational for me to prefer that the replica undergo the five minutes of pain? Does it seem irrational for me to prefer that the person with my body rather than the replica be given the money to travel to Paris to see the Mona Lisa? I think that most of us would answer no to both questions. Or consider that one of the resultant persons will die several hours after the replication. Am I not justified in caring that the replica rather than the person who is physically continuous with me die in a few hours? Most of us would think me perfectly justified in being more concerned about the person who is physically continuous with me than about the replica. So our views about what matters seem to coincide with the Mixed Reductionist account of personal identity in this case.

Thus, Mixed Reductionism would allow us to retain, at least in this case, the commonsense claim that identity matters.

The friend of the Modified Psychological account might say that, if I choose to be replicated and I then replicate myself with the intention of destroying my body, it is very difficult to see why my concern should be for the person whom I intended to have destroyed. Of course, the Modified Psychological Reductionist will say, I am entirely justified in caring more about the person who is physically continuous
with me if the replication was done against my will or without my knowledge. But if I choose to have my mental life continue in a new body, shouldn’t I be concerned about the replica rather than about the person with my old body? Consider, however, the situation in which I have decided to replicate myself. In one minute, the replica will be ready. I now realize that I do not have the appropriate materials for destroying my body as I had intended to do. I am then presented with the following choice: either the replica or the person with my body will suffer five minutes of excruciating pain. Am I really irrational if I have a strong preference that the replica suffer the pain? It still seems that I am not, that I am justified in caring more about the person who is physically continuous with me than about the replica.

The Psychological Reductionist may respond by saying that it is not at all surprising that we have the intuitions that we have in Branch-line, because in that case, habit leads us to think that the person who is physically continuous with me is also identical with me. After all, the relationship between me and that person is the same as it would be in any ordinary case of survival. So when I look ahead to the future, I am naturally inclined to focus more on the person who is physically continuous with me than on the Replica. In order to try to counterbalance this habitual tendency, we should try to consider the situation from the perspective of the replica rather than from my perspective before the replication. My replica will want to see the Mona Lisa just as much as I now want to see it and she will have
the same reasons for wanting to see it. Also, the replica’s mental life will be exactly like, in all important respects, the mental life of the person who is physically continuous with me. Given these facts, can I really care that the person with my body see the Mona Lisa? Such considerations become even more compelling when we consider the case in which I am replicated and my brain and body destroyed. In this case our initial reaction is that I have been given a new body. But, more importantly for present purposes, it is difficult to see why I should care that the replica is not physically continuous with me. Again, the when we look at the situation from the perspective of the replica, it is difficult to see any important difference between a case of replication and a case where I survive in the normal way. Shouldn’t I be satisfied if my replica finishes my thesis and sees the Mona Lisa? Can the importance of my private projects being realized really involve their being realized by someone who is physically continuous with me? Why should I care whether the person who finishes my thesis has some of the same matter that I have?

There seems to be a very quick answer to these questions. If my survival demands physical continuity, then it can matter to me whether or not there is some future person who is physically continuous with me. If there is no future person who is physically continuous with me, then I will not finish my thesis and I will not see the Mona Lisa. Prior to forming my private projects, I do not know all of the necessary conditions of their realization. I may have an independent belief
that it is trivial whether or not those necessary conditions are realized. For example, prior to forming my project that I finish my thesis, I do not know exactly what is a necessary condition of my doing so. I may find out that, in order to finish my thesis, I have to format my document in a particular way. I may previously have thought that it can make no difference how the document is formatted. However, after learning that such formatting is a necessary condition of completing the thesis, it is rational for me to care that my thesis be so formatted. Once I have projects, I may be committed to caring that the necessary conditions of their realization be satisfied whether or not I know what those necessary conditions are. So if the realization of my private projects demands that there be someone physically continuous with me, then physical continuity is something that I can rationally care about.

There, is, however, a significant difference between the relation of a necessary condition of my persistence to the fulfillment of a private project and the relation of the formatting of my thesis to the fulfillment of my project that I finish my thesis. When I formed my project that I finish my thesis, I had certain reasons for undertaking that project. I had views about how the writing of a thesis could enrich my life and contribute to my self-development, and also about how the completion of a thesis could promote my other project of having a career as a philosopher. I conclude that the formatting of the document is trivial in so far as the formatting has no impact on how the thesis will promote my intellectual
development. Nonetheless, I will conclude that the formatting is important in so far as I need to do whatever is required of me to finish the thesis if I want to insure that I fulfill my other project of having a career as a philosopher. So, in the case of the formatting of the thesis, I have some explanation as to the ways in which the formatting is trivial and as to the ways in which it is significant.

But now consider my project that I see the Mona Lisa. Suppose that, after I form this project, I decide to be replicated and to have my brain and body destroyed. Suppose that I am also told that if I go through with the replication, I will not survive because the resulting person will not be physically continuous with me. Therefore, if the replica sees the Mona Lisa, my private project that I see the Mona Lisa will not be realized because the replica is not identical with me. But all that is missing in this case is physical continuity. In order to justify my private project, I need some explanation as to why physical continuity is significant, just as, in order to justify my project that I complete my thesis, I needed some sort of explanation as to why I should worry about the formatting of the document.

I do think, however, that the case of the formatting of the thesis indicates that we need to be careful in assessing the significance of the body. For in the case of the thesis, it is quite true that the formatting of the document has no independent, intrinsic value. The formatting derives its significance from the role that it plays in helping me to fulfill my project and thereby to fulfill other projects. So we cannot evaluate the importance of physical continuity, or, for that
matter, psychological continuity, independently of having some grasp of what role they play in my projects as a whole and also of what role my private projects qua private projects play in my life. Even if we were to spell out what the completion of a thesis consists in, we might not have an explanation of what is important about my completing my thesis, because we may still need an understanding of how the thesis fits into the structure of my life. Similarly, a reductive analysis of personal identity may not be sufficient to explain why identity matters. So, keeping these considerations in mind, let us look at each of physical continuity, psychological continuity, and private projects to try to see why each might or might not have significance.

Let us begin with physical continuity. Why should I prefer that the replica rather than the person who is physically continuous with me undergo the pain? Why prefer that the person who is physically continuous with me rather than the replica see the Mona Lisa? Do I really care whether someone who is physically continuous with me finish my thesis? If so, how much physical continuity can I rationally care about? Prompted by the cases of physical splitting, we might say that we need at least as much physical continuity as there is between me and one of my fission products. Then I should be equally concerned about my fission products as I would be about my future self, but I should be more concerned about either my fission products or my future self than I should be about my replica. But now suppose that one of my fission products has one more of my

123
original cells than the other has. Can that fact justify me in caring more about the former fission product than about the latter? Surely not. Then suppose that one fission product has two more cells, or three more cells, and so on. Is there some point at which the number of cells has been reached such that now I am justified in caring more about one of the ‘fission’ products than about the other? Again, it seems not. Another option is to say that concern should decrease as physical continuity decreases and that there will be a region of middle cases where it is indeterminate whether concern is warranted. Nonetheless, it will be clear that I should care about the resultant person in the near cases but not in the far cases of the spectrum. The problem with this option is that it is difficult to see why concern should decrease as physical connections decrease. It really is difficult to understand why I should care more about someone who is psychologically continuous with me and has my legs, feet, left arm and pancreas, than about someone who is psychologically continuous with me and has only my feet, one of my legs, and my pancreas. And why care more about someone who has my legs, arms, and skull than about my replica?

These sorts of considerations do seem to support the claim that I should be equally concerned about both the person with my body and the replica. If that is the case, then perhaps it is not implausible to suppose that I do not survive in that case. Even though I do not survive, the Psychological Reductionist can still hold that what matters, psychological continuity, is present in my relationships to
both resulting persons. So my not surviving in this case is not something that
should worry me. In the simple replication case, I should not be concerned about
whether or not I survive. The replica’s relationship to me contains everything that
matters, because the replica is psychologically continuous with me.

Now let us consider the suggestion that psychological continuity is what
fundamentally matters in my survival. Why should I care that there is some future
person such that she is psychologically continuous with me? Of course, a replica
who is fully psychologically continuous with me cannot realize any of my private
projects such as my project that I finish my thesis. However, given the fact that
my replica’s psychological relationship to me is exactly the same as my
psychological relationship to my future self in an ordinary case, my replica will
finish my thesis in exactly the same way as an ordinary future self of mine would
have finished it. In so far as I want that I finish my thesis, I must attach some
significance to its being finished as I would have finished it. But if I am not going
to finish it, why should I not prefer that it be finished by a more talented
philosopher who will present more convincing arguments? Why should I think it
important that it be finished as I would have finished it if I had survived? It might
be that what I want is to be in some way responsible for the completion of the
thesis. As we have noticed several times previously, in order for there to be the
appropriate sort of psychological continuity between me and the replica, it must be
that the replica’s psychological states are not only dependent on mine but are also
in some sense under my deliberate control. More than wanting that my thesis be completed in some particular manner, I want that my thesis be a product of my will and intention. As long as my replica is not psychologically continuous with me as a result of an external imposition, it seems that whatever the replica does is as attributable to my will and intention as is anything that a normal future self of mine might do.

But consider my project that I see the Mona Lisa. Why suppose that I should care that someone whose psychological states are appropriately related to mine see the Mona Lisa, unless that person is me? What is it that I want in wanting that my future self see the Mona Lisa? It just seems wrong to say that I want someone to view it in the way in which I would view it, or that I want someone to view it with the same frame of mind that I would view it. And it seems very implausible to suppose that what I want is to be responsible for a certain person's seeing the Mona Lisa. This project seems to represent a certain sort of concern for a particular person, a sort of concern that is unanalyzable in terms of other concerns. If this sort of project is justified, then it must be identity rather than psychological continuity that matters. But can I justify my having the private project that I see the Mona Lisa instead of my having the quasi-private project that someone psychologically continuous with me see the Mona Lisa?

In order to answer that question, let us look at a case where I have both the private project and the quasi-private project. Suppose that I have a daughter to
whom I am extremely close. I alone have raised her since she was born, and I have educated her myself. She and I spend most of our time together, influencing one another's beliefs, desires, values, intentions, etc. Each of us has a great many projects which essentially involve the other. So there are a great many very strong psychological connections between my daughter and me. Now suppose that my daughter and I have been planning a trip to Paris to see the Mona Lisa. Before we are able to go, I develop a fatal disease. I know that I will not live long enough to accompany my daughter to Paris. I am glad that at least she will see the Mona Lisa but I am, justifiably it would seem, upset and extremely disappointed that I will never see it. My daughter's seeing the painting does not satisfy my desire that I see it, although I may regard her seeing it as the next best thing to my seeing it. If I cannot see it, then I want, at least, that my daughter, the woman in whom I instilled a love of Renaissance art, see it. Notice, as Parfit himself would insist, that the difference in my psychological relation to my daughter and my psychological relation to my future self is only a difference in degree, not a difference in kind. In fact, given the way that I have told the story about my daughter and the nature of our relationship, the difference in degree is not extremely great.

The relationship that I have described is an extremely close one, even abnormally close. Am I still justified in being disappointed that I will not be able to see the Mona Lisa even though my daughter will see it? I think that I am. Of
course, I am very glad that my daughter will be able to see the Mona Lisa, but the fact is that I wanted the experience of seeing it and there is some loss for me if I do not see it. There may be some people who ‘live for their children’, but in my example I still have the private project that I see the Mona Lisa. I seem justified in continuing to have that project even though I have a daughter whom I can influence to have the same projects that I have. If that is the case, then it seems that what I want in wanting my future self to see the Mona Lisa is not captured merely because someone over whom I have a certain psychological causal influence sees the Mona Lisa.

Consider again now the branch-line case, and suppose that I am told that the person who is physically continuous with me will develop a fatal disease. As a result, the person who is physically continuous with me will be unable to see the Mona Lisa, although the replica will be able to travel to Paris to see it. If replication contains everything that matters, then I should have no reason for regret in this case; I should be satisfied by the replica’s seeing the Mona Lisa. In the case of my daughter, I was glad that my daughter would see the Mona Lisa, but I had what seemed to be a reasonable disappointment in my not being able to see it. The relationship between me and my replica is very much like the relationship between me and my daughter, so I should be glad that, if I cannot see the Mona Lisa, then at least my replica will see it. And if the person who is physically continuous with me were to be cured of the disease, I would want that
both she see the Mona Lisa and the replica see the Mona Lisa. But why is it unreasonable in the branch-line case but not in the case of my daughter for me to continue to have the private project that I see the Mona Lisa? Why is my disappointment at my not seeing the Mona Lisa legitimate only in the case of my daughter but not in the branch-line case?

The Psychological Reductionist might respond that the sort of causal influence that I have over my replica is different than that which I have over my daughter. But the difference is only one of a very small degree, not of kind. The difference is one of the strength and number of connections that hold between me and either my replica or my daughter, and that difference is very small. The case of my daughter is unusual in so far as in that case I have an unusually high number of projects that involve another person, and, thus, an unusually low percentage of my projects are private projects. In a normal case, the vast majority of my projects are private projects. (This does not, of course, imply that I am selfish, because a private project may take, e.g., the form that I help person x where x is not identical with me.) Thus, a major component of my psychological relation to my future self is my set of private projects and their persistence or realization (cf. Whiting ‘FFS’). So my private projects are an important part of psychological continuity.

There may appear to be a circularity here in so far as private projects help to constitute psychological continuity which in turn helps to constitute personal identity, because private projects presuppose personal identity. A private project is
a project such that it essentially involves a reference to someone identical with me, but the explanation of what it is for someone to be identical with me involves a reference to psychological continuity which will in turn be characterized by having private projects as components. We can, however, define a new concept of a minimal psychological continuity. Let us say that 'psychological continuity minus' is psychological continuity without any connections involving private projects. Private projects, then, can be characterized in terms of psychological continuity minus. Notice that psychological continuity minus is a very thin sort of psychological continuity (cf. Whiting FFS). Such a relation might be sufficient for personal identity, but barely so. It will be the sort of relation that I stand in to my future self who is in a coma such that her mental life consists only in random memories and imaginings. However, in most cases of survival, the strongest components of psychological continuity will be private projects. The claim that psychological continuity matters, then, presupposes for its plausibility that we are talking about psychological continuity with private projects, not psychological continuity minus. But a quasi-private project must be a project such that it involves an essential reference to someone psychologically continuous minus with me, because quasi-private projects are supposed to replace private projects. Why should I want that someone who stands in such a minimal relationship to me finish my thesis? In normal cases where we have something like psychological continuity minus, cases of certain sorts of coma or mental illness, we are often prompted to
ask whether such a future is even worth having, and most of us think that it is not.

Of course, however, once we adopt quasi-private projects, they will form the core of a strong psychological relation to either our future selves or our replicas or fission products; quasi-private projects will supplement psychological continuity minus. Then again, however, I could adopt all projects regarding my daughter; I could give up all of my private projects and change them to projects such that they can be realized by either me or my daughter. Am I irrational if I do not do so? It seems not. So why suppose that I am irrational if I do not change my private projects to quasi-private projects that could be realized by either me or my replica or my fission product?

Perhaps the reason that we cannot explain the significance of private projects by referring to physical or psychological continuity is because private projects rely on some Non-Reductionist further fact for their justification. It does not seem, however, that the existence of such a further fact really underlies my special concern for my future self. For it is not at all clear why I should care whether someone with, e.g., my soul avoid pain and see the Mona Lisa. It seems very difficult to provide an answer to the question as to why we should be specially concerned about future persons with our souls or Cartesian Egos. The only answer available to the Non-Reductionist is that such a person is identical with me. So future-directed self-concern does not seem to rely in any way on the existence of some further facts.
I think that the case of my daughter shows that I am justified, in the branch-line case and in the simple replication case, in being disappointed if my private projects are not realized. Therefore, if physical continuity is a necessary condition of personal identity, I am justified in caring more about the person who is physically continuous with me than about my replica. However, in the case of my daughter, I may care more that my daughter see the Mona Lisa than that I see it, so I may be more disappointed if she cannot see it than I will be if I cannot see it. But that might be true if my daughter and I had a less close, more normal mother-daughter relationship. My projects involving my daughter are a form of concern for her, and that concern might be so strong that I feel compelled to sacrifice the realization of some of my private projects in order to insure my daughter’s well-being. Similarly, if I decide to be replicated and I put a lot of work into the creation of my replica, I may want to insure the replica’s well-being. Nonetheless, that does not show that I have no reason to regret the loss of my own future even if my replica could take my place. Psychological continuity does seem to matter and to be able to ground concern for persons with whom we are psychologically continuous, but it is not clear that we must conclude that identity does not matter in so far as it is not clear why I should change all of my private projects to quasi-private projects.
4.4 Indeterminacy and What Matters

There is, however, another feature of the Reductionist view which would seem to show that it cannot be identity that matters in survival. Recall the Combined Spectrum, the series of cases such that in each successive case the resulting person was just slightly less connected, both physically and psychologically, to me, the original person. In the middle range of cases, any Reductionist should say that it is indeterminate whether I survive. It is indeterminate whether the resulting person is identical with me. Should I go to a great deal of trouble to insure that the outcome of my operation is such that it is not a case in the middle range but rather in the near end of the spectrum, even if the outcome will be a case very close to the middle range? How can I care, when there is only a slight difference between the two cases? If I cannot care, then it would seem that I cannot rationally prefer one outcome to the other even though I survive in the one outcome but not in the other. So identity cannot be what matters in our survival.

We may wonder why the rationality of our attitudes may not be indeterminate just as our identity is. In the near end of the spectrum I determinately survive, so I can determinately realize my private projects. In the middle range of cases, I do not determinately survive; therefore, in such a case, it is indeterminate whether my private projects are realized if they are realized by the resulting person. If facts about identity are indeterminate, then facts about the rationality of concern which tracks identity, are also indeterminate. Whether or not I adopt such concern is a
matter of psychological fact and will vary from person to person.

Parfit can respond in the following way. Pick the last case in which I determinately survive and the first case in which it is indeterminate whether I survive. If I am specially concerned about the resulting person in the former case, should I not then also be specially concerned about the resulting person in the latter case? The difference in connectedness between the two cases is trivial. How can such a trivial difference make a substantial difference to my attitude towards the resulting person? If it cannot, then I cannot rationally care whether I determinately survive. The mere fact that the resulting person is me cannot make a difference as to whether I should be specially concerned about that person.

It does seem, however, that I do have reason to prefer an outcome in the near end of the spectrum over an outcome in the far end where it is determinate that I do not survive. If it is not the fact that I survive in the former case which should make me prefer that case, then it must be the strength of the connections which should matter to me. As connections weaken, I should have less special concern for the resulting person. Of course, Parfit does want to defend this view. But I think that it is extremely implausible when we consider a variation on our earlier Branch-line case. Suppose that the process of replication is such that it will effect certain changes in me, and it will effect these changes instantaneously. The resulting person who has my brain and body will be like the resulting person in the last case in the near end of the spectrum; the replication process will cause a
significant weakening of connections, but, nonetheless, the person with my brain
and body is still determinately me. The replica, however, will be like the resulting
person in the next to the last case of the near end of the spectrum. So the replica
will be more strongly connected to me now than my future self will be. (Of course,
I am here assuming that the process of replication does not involve any external
impositions. Therefore, replication here does not destroy any of the relevant sorts
of causal connections.) Can I still rationally prefer that the replica rather than my
future self undergo the five minutes of pain?

I think that I can. Why should I care whether the replica is slightly more
connected to me now than is my future self? Mark Johnston suggests that we
consider a parallel case where it is my friend who undergoes the replication process
which alters her ("Reasons and Reductionism" 30ff.). Should I decide to have my
friend or her replica undergo the five minutes of pain? If I truly care about my
friend, then I will care about her even if she changes in certain ways. Even if she
changes dramatically by, say, developing Alzheimer’s disease, I am still justified in
caring more about her than about some similar persons whom I have never met.
So surely I should care more about her than about a replica that is slightly more
connected to her as she is before the replication process. If that is the case, then I
would also seem justified in caring more about my future self than about a slightly
more connected replica. So strength of connections is not the basis of concern. Of
course, connections can ground concern, but they ground concern for particular
persons. Once the concern is in place, it is concern not for the connections but for particular persons.

So I do not think that it is irrational for me to prefer the outcome in which I determinately survive in the Combined Spectrum. But should I go to a great deal of trouble to insure that I determinately survive? I think that that is less clear. Even in normal cases, we do not always think that we should go to great lengths to prevent our own deaths. Imagine that my friend, as a result of the replication process, will develop Alzheimer’s disease. She asks me to let her die and then have the replica take over her life. I will acquiesce, because I do not want my friend to suffer. But I will still regret her loss. So the fact that I may not go to a lot of trouble to insure either that I or a friend survives in no way shows that identity does not matter.

4.5 Fission Products Again

At the end of section 2.3 I suggested that neither psychological nor physical continuity can capture all that matters in private projects, that, in fact, private projects are an important component of any sort of psychological continuity that matters. But there is another component of identity, uniqueness. Perhaps I should be rationally concerned that branching not occur because uniqueness matters. In order to see if this is a plausible claim, let us consider the fission cases in which all that is missing for survival is uniqueness.
Consider the case of fission upon which Parfit based his claim that it is psychological continuity rather than identity that matters. His argument depends upon the claim that I survive when only half of my brain is transplanted but the other half is destroyed, while I do not survive when each half of my brain is transplanted into a new body. I do not find this claim very plausible. After all, if the scientists took away half of my brain and destroyed it, I would still survive with my body and the other half of my brain. Could it matter, then, if they transplanted the other half of my brain? Would I not still survive as the person who is fully physically continuous with me? But the other person is just like the resulting person in Parfit's original fission case. So why suppose that I survive in that case?

So let us concentrate on the case where I split like an amoeba. In this case, I split in half and each half of my body spontaneously reproduces an exact copy of the other half. Each of the resulting persons, A and B, has half of my brain so each resulting person is both physically and psychologically continuous with me. Certainly we do have reason to suppose that I would have survived if half of my body had been burned and then regenerated. Both Psychological and Mixed Reductionists can agree on that conclusion, and on the conclusion that I do not survive in the fission case. There is no reason to suppose that I am one of the resulting persons rather than the other and I cannot be both. How should I regard A and B? Can I rationally care that A and B are not identical with me as long as
they are both physically and psychologically continuous with me?

Parfit claims that in order for me to rationally care that I do not survive in this case, I must care about uniqueness. But how can it matter to me whether there is only one or whether there are two future persons who are physically and psychologically continuous with me? (I am abstracting from practical difficulties involved with sharing, e.g., spouse or career; cf. Wolf 714.) As I said above, it is difficult to see how extrinsic facts can matter in my concern for some future person. *Prima facie,* it is not plausible to think that my concern for A can only be rationally decided after I know whether the other half of my body has been burned or whether, instead, B has come into existence. If my parents find out later that B is around, should they come to regard A in a different light than they did when they thought A was identical with me? It would seem not.

But whether or not A is able to realize my private project that I see the Mona Lisa depends on whether A is identical with me. I have tried to show that it is difficult to explain the importance of my private projects in terms of physical continuity, psychological continuity, or even further facts. And now we see that it is difficult to explain their significance in terms of uniqueness. We might conclude, then, that private projects and concern for particular persons is derivative. Such concern must be derivative from more impersonal concerns if it is to be justified. This is an account that I will consider and reject in the next chapter. An alternative is to deny that we can explain the significance of private projects in
terms of the component elements of identity. The justification of my private projects is independent of or prior to any views about the nature of persons. That does not mean that our private projects are unjustified or irrational. It simply shows that our concern with our identity need not appeal to any analysis of the concept for its justification.\textsuperscript{16}

Such an answer does not seem entirely satisfactory. But I think that it becomes more so when we realize that it does not commit us to the claim that I should be indifferent about my fission products. In my discussion of psychological continuity above, I tried to show that psychological continuity alone does not explain the significance that we attach to private projects. Nonetheless, that does not mean that the realization of a quasi-private project should not matter to me at all. This becomes quite clear when we consider the case of my daughter to whom I am very close. Even though I will be disappointed if I will not be able to see the Mona Lisa, it is still important to me that my daughter see it rather than that some stranger see it, even if the stranger happens to be remarkably similar to my daughter. And that fact is plausibly explained in terms of the relationship in which my daughter stands to me, a relationship which differs only in degree from my relationship to my fission products. Therefore, I should be very concerned that one of my fission products rather than some stranger or even a friend finishes my thesis. The relationship in which I stand to my fission products should ground

\textsuperscript{16}Mark Johnston tries to justify it in terms of our ordinary practise involving self-referential concerns; see “Reasons and Reductionism”.

139
concern for them. Nonetheless, I am not irrational in hoping that I do not split like an amoeba.

Parfit, however, argues that fission is not as bad as ordinary death. In other words, it is not as bad as my dying and leaving behind no one who is physically and psychologically continuous with me. It may be true that it would be better for every person if instead of dying and leaving behind no heirs to their psychology they underwent fission. But, similarly, it may also be true that it would be better for each of us to die with rather than without heirs to our fortunes (or debts). But just because one situation that might obtain after our deaths is better than another (and no one ever doubted that), that in no way shows that if the better situation were to obtain then we should not rationally care about our own deaths. If I really want heirs to my fortune after my death, and if I died now there would be heirs, that does not show that I am irrational in wanting to survive even if by doing so I will not be sure of having heirs. Similarly, even if dying with psychological heirs were better than dying without them, I am not thereby irrational in wanting to live. So even if Parfit regards fission as “not as bad as death”, that in no way supports the claim that fission is “about as good as ordinary survival” (cf. Wolf 715). But fission may be better than ordinary survival if it is the only alternative to my developing Alzheimer’s disease or, worse, a homicidal pathological disorder.

I may be glad that I will have fission products to take over my projects, just as I would be glad if my daughter or student were to do so. But that does not show
that all of my personal concerns should be quasi-personal concerns. Our concerns seem to be independent of and prior to any metaphysical analysis of persons. They need not appeal to any such analysis for their justification. These concerns can be undermined, but they are not undermined by a mere appeal to any Reductionist account of personal identity.

But we may wonder why our projects are not shown, by Reductionism, simply to have no foundation. I think that such a conclusion would be acceptable if it could be shown that private projects are based upon an implausible view of personal identity such as Cartesian Non-Reductionism. As I pointed out above, however, Cartesian Non-Reductionism seems unable to provide any account of future-directed self-concern, and the legitimacy of private projects does not seem to depend upon the truth of Cartesian Non-Reductionism. If our private concerns continue to look justified in spite of the Reductionist analysis of personal identity, before rejecting those concerns, we should attempt to provide another basis for them. The full plausibility of my claim that identity matters depends upon some fuller account of the nature of our practices involving private concerns, but I think that we have sufficient reason to try to develop such an account rather than abandoning our private projects in favor of quasi-private projects.

We may, however, decide that if Reductionism is unable to account for the significance of private projects, then we should reject Reductionism. As I said above, Cartesian Non-Reductionism does worse than Reductionism in justifying
special concern. Perhaps, then, we have reason to accept the Further Fact View, the view that claims only that personal identity is simple and unanalysable. If this view were true, it would certainly explain our inability to analyze the significance of private projects which involve personal identity. The friend of the Further Fact View can agree that psychological continuity matters and that it can ground concern for fission products, because he can say that psychological continuity is involved in some way in personal identity as is evidenced by ordinary cases of survival. Yet, he will deny that personal identity is constituted by facts about continuities; therefore, concern about identity cannot be reduced to concern about continuities and uniqueness. The Further Fact View, however, still leaves personal concern unexplained. Reductionism did this but also gave us an explanation of how psychological continuities are involved in identity. It might, then, be better to accept Reductionism and try to locate the justification of personal concerns in some other place than in the metaphysics of personal identity.

Should we, then, prefer Mixed Reductionism or the modified Psychological account? The advantages of the modified Psychological account are that it accommodates our intuitions about Body-switch and Brownson and it does not seem unmotivated as does Mixed Reductionism’s physical continuity clause. The modified Psychological account, however, does yield unappealing results in Branch-line replication cases. Both views do just as well in explaining special concern for fission products because both incorporate psychological continuity as
an element of identity. However, Mixed Reductionism can do better in the case of my daughter and the Mona Lisa because it can draw a more definite line between self and others by appealing to physical continuity. I do not think that the Mixed Reductionist need worry about the apparent unimportance of physical continuity; as I have said, she can deny that concern about identity reduces to concern about the constitutive elements of identity. Although I prefer Mixed Reductionism on the basis of the Branch-line cases, I think that, given the complexity of our intuitions, it is difficult to decide how much those intuitions about Branch-line are an extrapolation from the ordinary run of cases. What we need is a better rationale for the physical continuity clause; however, I will not attempt to try to find such a rationale here.

Instead, I want to return to the fission cases and the case of my daughter. If psychological continuity is present in both cases as well as in the intrapersonal case, and all of those cases are cases where special concern seems warranted, we might think that psychological continuity can ground special concern. So let us turn to a consideration of the problem of special concern not only for future selves but for other persons.
Chapter 4
The Grounds For Special Concern

5 Introduction

Most of us are supplied at various times with opportunities and resources for benefitting either ourselves or other persons. When such circumstances arise, we typically choose to benefit those persons to whom we are related in special ways. We benefit our own future selves, our friends, and our family members before and to a greater extent than we benefit persons with whom we have had little or no contact. And we do so even if we could have used our resources to benefit a stranger more than we were able to benefit our future selves or our intimates.

How are we to justify such special concern? I think that the Reductionist views that I discussed in the previous chapter, Mixed Reductionism and modified Psychological Reductionism, suggest an answer to this question. According to both views, psychological continuity is a necessary condition of personal identity. My being psychologically continuous with some future self is at least part of what it is for that future self to be my future self; moreover, certain elements of psychological continuity such as private projects are part of what matters in my relationship to my future self. Therefore, if we think that I should be specially concerned about my future self, it seems plausible to suppose that psychological
continuity is at least part of the story as to why such special concern is appropriate. My future self’s being psychologically continuous with me gives me reason to be specially concerned about her.

The cases of fission and of my daughter showed, however, that psychological connections can be present even in interpersonal cases. In fact, such connections can exist in more ordinary interpersonal relationships with friends and family members. So perhaps psychological continuity can provide reasons for concern in not only the intrapersonal but also in the interpersonal case. We could then see the Reductionist view as providing motivation for the commonsense claim that the relationship in which I stand to certain persons is grounds for my having special concern for those persons. My special concern for my future self and my intimates is justified by the fact that they are my future self and my intimates, in other words, by the fact that they bear certain psychological connections to me.

Some philosophers, however, object to the commonsense view. They claim that it must be some feature of the object of concern, some feature independent of me, that justifies my special concern. Otherwise special concern will be too egocentric in that it will depend on the object of concern’s relation to me rather than on some feature of the object of concern which makes her worthy of concern. In what follows, I will develop, in its strongest form, this Impersonalist challenge to the common sense view. I will show that the Impersonalist approach is both empirically inadequate and morally objectionable, and that we should accept the
commonsense position.

6 The Problem of Special Concern

I intend to consider the Impersonalist and the commonsense views as accounts both of the rationality of and the morality of special concern. It seems plausible that it is rational for me to be specially concerned about my future self and my intimates, and also that it is morally justifiable and even required that I have such special concern. We are inclined to think that, all else being equal, I am not irrational in being more concerned about myself and my family than about the governor of Idaho and that I am morally required to care more about myself and my family than about the governor of Idaho. In this chapter I am interested in finding a plausible account of how I can be justified, morally or rationally, in singling out certain persons and having far greater concern for them than I have for all the other apparently equally or perhaps even more deserving and even very similar persons in the world. Even though we need not give the same account of the morality of special concern that we give of its rationality, I think that the commonsense view does and is intended to serve both functions. Although Impersonalism may be intended as an account only of the rationality of special concern, we will see that it has more plausibility as an account of its morality.

I want to begin by considering the commonsense view that special concern can be justified in terms of special relationships. We can begin by noticing that we are
inclined to think that I have offered sufficient justification for my special concern for my future self when I point out that my future self is me. In the intrapersonal case, there is only one person involved, and, as Sidgwick says,

[i]t would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently ‘I’ am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of existence of other individuals(498).

In the interpersonal case, however, the distinction between my mother and me seems as ‘real’ as the distinction between the governor of Idaho and me. So why should I be specially concerned about my mother but not about the governor of Idaho? Such considerations may lead us to think that special concern for others is more difficult to justify.

Nonetheless, special concern for others does seem, in some important respects, to be like concern for one’s future self. As Aristotle says,

one person is most a friend to another if he wishes goods to the other for the other’s sake, even if no one will know about it. But these are features most of all of one’s relation to oneself; and so too are all the other defining features of a friend, since we have said that all the features of friendship extend from oneself to others(1168b2-6).

We can notice that most persons care about themselves for their own sakes and, in the case of the best sorts of friendship, care about their friends for their friends’ own sakes. Given this similarity between the intrapersonal and the interpersonal cases, we have some motivation for accepting and trying to fill out Aristotle’s claim
that the ‘defining features of a friend’ are also features of one’s relation to one’s future self. Therefore, we might try to model the justification of special concern for others on the nature of the justification of special concern for one’s future self.

If so, we need to find some factor present in both the intrapersonal and the interpersonal cases that seems to have justificatory force in explaining special concern. Again, we can look to Aristotle’s claim that a [decent] person “is related to his friend as he is to himself, since the friend is another himself” (1166a30). Perhaps the features that define both my relation to my future self and my relation to my friend can provide reasons for special concern. I bear certain relations to my future self, such as relations of similarity. My future self will have some of the same beliefs, desires, projects, and values that I now have. We do not, however, want to appeal to mere similarity between me and my future self as a grounds for concern. Persons typically change a great deal over time, especially over the course of an entire lifetime, and often initiate positive changes in themselves. It seems implausible to claim that if I improve myself in some way, I then have less reason to be concerned about my future selves.

But we can notice that even if my beliefs change over time, my future beliefs will be causally influenced by my present beliefs. The same is true of my desires, projects, and values. So even if my future self is not very (or at all) similar to my future self, I now will stand in certain relations of causal influence to my future self. In order to accommodate (or at least not rule out) concern for far distant
future selves or for future selves who are the result of self-improvement, I think that we should emphasize causal relations more heavily than relations of similarity as a grounds for intrapersonal special concern.

Recall that in Chapter 2 we defined Psychological Reductionism in terms of psychological continuity rather than in terms of direct psychological connections. This focus on continuity rather than on connectedness is, in other words, a focus on causal influence rather than on similarity. Having the same belief, desire, etc., is a direct psychological connection. I now may be directly connected to myself of a year hence, but will probably not be directly connected, at least not very strongly so, to myself of fifty years hence. If I now am similar to myself of a year hence, and myself of a year hence is similar to myself of two years hence, etc., I now will be psychologically continuous with my far distant future selves. I will not, however, be similar to myself of fifty years hence because similarity is not transitive. Since causality is transitive, I now will have causally influenced myself of fifty years hence.

More importantly, in Chapter 3, in the discussion of external impositions, we saw that an important part of our conception of a person involves the notion of a certain sort of agency. A person can form intentions, make decisions, and change herself in accordance with those intentions and decisions. I will actively work to make my future self different than I now am, and, hopefully, many of the changes that I will effect will be for the better. Thus, it is plausible to suppose that
psychological continuity without external impositions is a necessary condition of personal identity and something that matters in my relation to my future self. We focus on relevant causal relations rather than on relations of similarity, because, in caring about ourselves, we care not about ourselves as static beings with certain particular qualities, but about ourselves as changing, growing beings who can be agents of their own change. (For a defense of the claim that both continuity and connections matter, see Parfit 298-302, 312-315; for a response to Parfit, see Brink 1992 224ff.)

It seems right to claim that the relation that is, in part, what makes it the case that some future self is my future self can provide a grounds for my special concern for my future self, especially when that relation is part of what matters to us in being persons. But the case of my daughter to whom I am very close showed that such causal connections can be present in interpersonal cases. The causal connections that are such an important part of psychological continuity are certainly present in fission cases, because I am fully psychologically continuous with both of my fission products. We saw that it seems plausible to claim that I should be concerned about my fission products even though I am not identical with either fission product. It also seems that I should be specially concerned about my daughter. Given that causal connections are present in both of these interpersonal cases as well as in the intrapersonal case, we have reason to think that these causal connections ground the special concern that is appropriate in all of these cases. It
seems that my relationships to my future selves, fission products, and daughter is what gives me reason to be specially concerned about all of those persons.

Recall, however, that the case of my daughter was an extraordinary case. But, even though we rarely have relationships such as that between my daughter and me, we do causally interact with others, such as our friends. We share ideas through conversation and we share experiences by spending time together (cf. Aristotle 1170b10). We do often have similar beliefs, desires, and goals as our friends do. However, when our friends change, we do not (always) cease to care about them. This may indicate that we regard our causal interaction with our friends as being more important than how similar we are to our friends. So if we try to justify special concern for others on the model of special concern for our future selves, it seems plausible to take causal interaction as at least part of the intrinsic justification of interpersonal special concern. The fact, then, that I causally interact with and influence in important ways both my future self and my friends gives me a reason, in and of itself, for having special concern for both my future self and my friends. I will call this claim the commonsense claim about justification.

Notice that the commonsense claim about justification will support differing degrees of concern depending upon the degree of causal interaction. I should be more concerned about my future selves and fission products than about friends and family members and more concerned about friends and family members than
about neighbors and acquaintances. But the commonsense claim does not entail
that I have no reason to be concerned about persons with whom I have no causal
interaction. I may have other reason to be concerned about persons to whom I
stand in no special relationship. I may be required to forego providing a trivial
benefit to a loved one if I can thereby provide an immense benefit to a complete
stranger. The commonsense claim simply says that causal relations provide
intrinsic reason for special concern, not that they provide the only sorts of reasons
or that the reasons that they provide cannot be outweighed by counterbalancing
reasons. All else being equal, however, I should be specially concerned about
persons with whom I causally interact in proportion to the degree of causal
interaction.

It is also important to see that the sorts of causal connections that provide
reasons for concern are the sorts of connections that exist in the intrapersonal and
interpersonal cases where it seems uncontroversial that such concern is warranted.
In these paradigm cases, there is mutual interaction founded upon the intentions
and goals of both parties to the interaction. In the last chapter, we saw that this
sort of interaction is missing when I interact with those sorts of external agents
that I called ‘external impositions’. The sort of interaction that I have with the
neurosurgeon who rewires my brain is not the appropriate sort of interaction upon
which to base concern. I am not required to have special concern for my
neurosurgeon merely because he rewired my brain. Similarly, external impositions
destroy some of what matters in my relationship to my future self. So if a terrorist captures me and tortures me for years, I will causally interact with him, but that causal interaction does not conform to the causal interaction that is present in the paradigm cases. The terrorist is the interpersonal parallel of an external imposition affecting my relationship to my future self in the intrapersonal case. In both cases, one of the parties to the interaction is passive; the interaction fails to take account of her goals and intentions.

We can, however, imagine a case where I willingly submit myself to a relationship in which I am enslaved, and my intentions and goals all involve my 'master'. Here there is mutual interaction based upon the goals and intentions of both parties. Perhaps I do have an obligation to be concerned about my master in this case. Nonetheless, that does not imply that I am committed to continuing to be guided by the present norms of the relationship. To see why, consider an intrapersonal parallel case in which all of my goals and intentions center around my project to count the blades of grass on the Boston Common. In this case I am required to be specially concerned about my future self, but, because my well-being involves a revision of my projects, I am required to reevaluate my life goals. Similarly, a consideration of the well-being of my master and of myself will show that special concern should take the form of reevaluating the norms of our relationship. Special concern is grounded by special relationships, but, in determining what form special concern should take, we will often need to appeal to
independent values.

If we accept that causal relations justify our special concern for our future selves and our loved ones, we will accept that reasons for special concern are agent-relative, where an agent-relative reason is one which cannot be stated without an essential reference to the person whose reason it is (cf. Nagel 152-3, Parfit 143). My reason to have special concern for my mother is agent-relative because it is based on the relationship in which my mother stands to me. Also, I will not have reason to have special concern for anyone who is not appropriately related to me. Other persons may also have agent-relative reason to be concerned about my mother if they also stand in a special relationship to her.

But recall Aristotle's claim that someone wishes goods to his friend for the friend's own sake. In other words, one should care about one's intimates for their own sakes. If, however, I base my concern for my mother on the fact that my mother is related to me in a certain way, that may seem incompatible with my concern being for my mother for her own sake. It might seem as if I should care about my mother in virtue of some feature of her which is both independent of me and central to her being the person she is. So my reason for special concern for my mother should be agent-neutral, where an agent-neutral reason is one which can be stated without any essential reference to the person whose reason it is. So if, for example, reasons for concern are based on a person's moral worth, they are reasons for anyone to be concerned about any person who has the appropriate
worth. These characteristics of persons will, then, be the grounds for both self-concern and concern for others on the agent-neutral, or, as I shall call it, the Impersonalist approach.

The Impersonalist approach, then, will also explain why we should choose to have special relationships with certain persons. Whatever feature of persons justifies special concern will also provide reason for one to enter into special relationships. So if someone is, for example, morally worthy, I have reason to become friends with her and then to maintain special concern for her. The friend of the commonsense view, on the other hand, will allow that concern can be justified even if I am now in a relationship which I should not have entered, such as my relationship with my 'master' described above. The Impersonalist may say that certain impersonal values must be appealed to, even on the commonsense view, in order to determine what form special concern should take. So it is plausible to appeal to such values in justifying the creation and maintenance of relationships as well.

The Impersonalist view might be further motivated by our noticing that it seems, at least initially, to provide a more informative explanation of special concern than the commonsense view does. The commonsense view seems to be claiming that we are justified in being specially concerned about persons to whom we stand in special relationships because we stand in special relationships to those persons. Such an explanation might seem to be no explanation at all. The
Impersonalist, on the other hand, points to some independent feature of persons such as moral worth, some feature that seems to merit concern, in order to justify special concern. We can understand why we should be concerned about moral virtue, so we can understand why we should care about persons who have such virtue. Given this strong motivation for the Impersonalist approach, it presents a strong challenge to the agent-relative justification of special concern.

I will begin by considering an Impersonalist account of special concern for one’s future self, and I will then extend it into the interpersonal case until we arrive at what appears to be the most plausible Impersonalist account of special concern for others. I will then examine various agent-relative accounts, finally arguing that it is implausible to suppose, as the Impersonalists must, that the relationships in which we stand to others cannot provide intrinsic reasons for special concern.

7 An Impersonalist Account of Special Concern for One’s Future Self

What justifies my special concern for my future self? Let us consider an Impersonalist account suggested by John Perry in “The Importance of Being Identical”. First, notice that each of us has many and varied projects at any moment, where a person’s projects (at a moment) are “any events a person at [that] given moment wants to occur in the future” (Perry 74). For example, I now
have the following projects: that I finish my thesis, that I care for my family, that I see the Mona Lisa, that the Cubs win the World Series, that my friends are successful, that poverty is reduced, etc. My projects, then, include not only those things which I will actively work to bring about, such as the completion of my thesis, but also anything which I simply hope or wish will come about in the future, such as the Cubs winning the World Series.

According to Perry, my present concern for my future self is grounded by my projects. Consider, for example, my concern that I not be in pain in the future. Why do I care that I not suffer great pain in the future? Well, if I am in great pain I will not be able to travel to Paris to see the Mona Lisa and I will not be able to work on my thesis. So I have ‘project-related’ reasons for being concerned about what happens to my future self. Because, as I pointed out above, my future self stands in relations of causal influence and similarity to me, I can expect my future self to share many of my projects. Thus, my future self is the person most likely to finish my thesis, take care of my family, and so on. My reasons for having special concern about my future self derive from what I can reasonably expect my future self to do or try to do in the future. Given the strong relations of similarity and causal influence in which I now stand to my future self, my future self will be the person most likely to further the maximum number of my projects.

As I said above, however, each of us has many and varied projects. Our projects vary not only in their content but in their value as well. Suppose that I
now have both the project that I finish my thesis and the project that I kill your dog. My future self is the person most likely both to complete my thesis and to kill your dog. It seems odd to suppose that I should have more concern for my future self if she is likely to complete both projects rather than being likely to just complete my thesis. In fact, we might think that my concern should be greater if she will probably not be able to kill your dog. So Perry's account is most plausible if we take it that concern is based not simply on my projects but on my valuable projects. My concern for my future self should be grounded on her ability to further many of my valuable projects.

It is important to notice that, according to the present account, the causal relations provide no intrinsic reason or justification for special concern. It denies the commonsense claim about justification. Special concern must be grounded on my projects. I am justified in being concerned about anyone who is likely to finish my valuable projects whether or not that person is in any way related to me. If for some strange reason it happened that the governor of Idaho was more likely than my future self to undertake my projects, I have reason to be more concerned about the governor of Idaho than about my future self. But given that I very strongly influence my future self, and that my future self will be quite similar to me, my future self is far more likely than anyone else to finish a large number of my projects. So my relationship to my future self provides extrinsic reasons for concern. Because these extrinsic reasons are typically very strong, we habitually
develop correspondingly strong concern for our future selves.

But now consider an important distinction in kind among my projects. My desire that I finish my thesis is clearly very different from my desire that the Cubs win the World Series. In order for the first desire to be satisfied, it must be my future self who finishes my thesis. The satisfaction of the latter desire, however, requires nothing of me, not that I help the Cubs win or even that I know that they win the Series. The first sort of project is a private project because it essentially involves reference to the person whose project it is. For any of my private projects to be fulfilled, my future self must be the person to realize them (see Chapter 3 for a discussion of private projects). Given the fact that all of us have many and mostly private projects, doesn’t the relationship in which I stand to my future self give more than extrinsic reasons for concern, because that relationship is a necessary condition of the realization of many of my projects?

Perry responds that private projects are only justified if one has the relevant nonprivate project, where a nonprivate project is one that does not essentially involve reference to the person whose project it is. If I have a given nonprivate project, such as the project that my thesis be finished\textsuperscript{17}, and I also have the belief that if I do not finish it then it will not be finished and I will not have a job with which to support my family, then I am justified in wanting that I finish my thesis. My project that I finish my thesis, then, is only justified if I also have the

\textsuperscript{17}The ‘my’ could be eliminated by giving a fuller description of the nature of my thesis.
nonprivate project that my thesis be finished and the belief that if I do not finish it, then it, and many other of my nonprivate projects, such as the project that my family be supported, will never be finished. So Perry justifies future-directed self concern by reference to my projects. I have a concern that my projects be realized, and because my future self is likely to realize those projects, I am justified in having a special project-related concern for my future self. Of course, if I were to undergo fission or replication, I would have the same project-related reasons to be concerned about my fission products and my replica as I have to be concerned about myself. (Notice that quasi-private projects are only going to be justified by reference to the corresponding nonprivate projects and the belief that if they are not realized by someone psychologically continuous with me then they will not be realized.)

8 Extending the Impersonalist Account into the Interpersonal Case

We can extend Perry’s Impersonalist account of special concern for one’s future self into the interpersonal context. According to Perry’s account, my special concern for my future self is grounded by my valuable projects: I should have special concern because my future self is the person most likely to further the maximum number of my valuable projects. We could say that, analogously, my
special concern in the interpersonal context is grounded by my valuable nonprivate projects. My friend is likely to help me to further my valuable projects and will also attempt to further them on her own. She will share many of my valuable nonprivate projects. My concern is, ultimately, then, for my valuable nonprivate projects, and my concern for any particular person, either my future self or my best friend, is justified by the belief that that person will help to further my valuable projects.

I think that such an account of concern for loved ones is very implausible. Notice that it is not an agent-neutral account, in so far as my reasons for my concern for my loved ones depend upon what projects I happen to have. My reasons involve essential reference to my projects. If someone had different projects, then they would not have reason to have special concern for my loved ones. If someone did not care about my valuable projects, then I would have no reason to be specially concerned about her. But one has to wonder why the set of valuable projects that I currently happen to have should be the basis of my concern for either my future self or my friends and loved ones. What about my friend’s projects? Why do my projects take precedence over hers? If the Impersonalist wants to motivate her view by pointing to its ability to take seriously the claim that we care about our friends for our friends’ own sakes, then it should not entirely ignore my friend’s valuable projects in offering grounds for concern for her.

The present account also leads to some unacceptable conclusions when we look
at certain cases. Suppose that my best friend were diagnosed as having a debilitating and eventually fatal disease. My friend will be too sick to help me further any of my valuable projects. On the present account, I am no longer justified in having any special concern for my friend. Certainly there are plenty of other people who can help me to further my valuable projects, so I should seek out one of those persons. I think that most of us would find such an attitude not only unjustified, but morally objectionable.

So grounding special concern by reference to my current valuable nonprivate projects seems dubious. My friend is a person with projects of her own, so concern for her for her own sake would seem to demand concern for some feature or features of her which are independent of me and my projects. The Impersonalist needs to develop a more agent-neutral account of special concern. In "Impersonal Friends" Jennifer Whiting presents a view "in which the substance or content of a person's character (as distinct from its relationship to one's own) is the ground of concern" (Whiting 11). The justification for concern for my future self will be analogous to the justification for special concern for others. If I consider the character of my future self and my friends to be valuable, then my special concern for each is justified. Thus, I have the same reasons for caring about my friend as I do for caring about my future self, provided that she and I have the appropriate characters.

18 All references to Whiting are to this work unless otherwise indicated.
Notice that this view, which we will call the ‘character-based view’, does handle some of the problems I have raised for the other views that we have rejected. If I care about my friend because she is virtuous, it seems that I care about her for her own sake insofar as she is to be identified with her virtuous character. Her virtue is not dependent on me or on my relationship to her. Also, consider my friend with the debilitating and eventually fatal disease. In such a circumstance, her projects will alter a great deal. Nonetheless, her projects may still be valuable, even if they are not projects that I share. Thus, I have reason to help her in promoting her ends even though they are not directly my ends. My friendship is justified because of the nature of my friend’s character, a character which involves a commitment to valuable projects. I, as a virtuous person with valuable projects, have reason to be specially concerned about any person with a virtuous character and valuable projects.

Recall that I asked, at the beginning of the paper, what justifies me in caring more about my mother than about the governor of Idaho if the governor of Idaho is, in important ways, similar to my mother. Let us suppose that my mother and the governor of Idaho have equally virtuous characters. If concern is grounded by the nature of a person’s character, then it seems that I should be equally concerned about my mother and the governor of Idaho. The Impersonalist could say that special concern is not justified, that I ought to be equally concerned about all equally virtuous persons. If the Impersonalist takes this line, she is not
offering an account of special concern, but, rather, a challenge to the commonsense view that special concern is justified. Although we should take such challenges seriously, I am here concerned only with the Impersonalist view as an alternative account of how special concern can be justified. So we need to see how the character-based view can justify my having more concern for my mother than for the governor of Idaho. In discussing Perry’s account of intrapersonal special concern, we saw that he justified my special concern for my future self by pointing to certain causal and epistemic considerations. Given the relations of similarity and causal influence in which I stand to my future self, she is likely to undertake my projects. Also, I have an intimate knowledge about her, so I can easily influence and direct her actions. Similarly, the proponent of the character-based view can point to causal and epistemic considerations that support my special concern for my mother. I know my mother and am very familiar with her projects, and, thus, I am more capable of helping her further her ends. Similarly, because I interact with my mother on a regular basis, I am in a better causal position to provide her with benefits. Also, I have very limited time and resources. So I must choose a few out of the many equally worthy candidates for my concern. Practically speaking, I should choose those I am in a position to know, such as family members. In the case of friends, I will choose those with whom I enjoy spending time. So the Impersonalist can provide strong extrinsic reasons for special concern. In any normal case, one will be able to provide a justification for
special obligations to one's self, one's friends, and one's family members.

Notice, however, that in offering even a pragmatic or extrinsic justification for my greater concern for my mother than for the governor of Idaho, the character-based view needs to assume that my mother is at least roughly as virtuous as the governor of Idaho. For suppose that it is clear to me that the governor of Idaho is far more virtuous than my mother. How can I then possibly have reason to promote my mother's good rather than the good of the governor of Idaho if I am in a position to do either? Similarly, suppose that it is also clear to me that the governor of Idaho is more virtuous than I am or could possibly be in the future. Should I not, then, on the character-based view, have greater concern for the governor of Idaho then I have for my future or even my present self? 19

This problem becomes even more acute when we consider persons who are not virtuous at all or whose virtues are combined with many vicious or trivial projects. Suppose that my mother is a person of this sort and I come to realize it. Does that fact justify me in ceasing to care about my mother? Should I simply abandon her to her less than fully virtuous ways? Or should I rather try to help my mother develop more valuable projects? These sorts of considerations become even more compelling when we consider the intrapersonal case. Suppose that I recognize that

19Whiting says that her "argument was inspired by Aristotle's conception of the ideal friend as 'another self' " (3; emphasis my own). For Aristotle, of course, the best sort of friendship is the friendship between virtuous persons. So Aristotle's account of character friendship was not meant to cover cases involving less than virtuous persons. In this paper I am concerned with accounts of special concern in a broader sense, with differential concern of any sort. An account which cannot justify my caring more about my mother than about the more virtuous governor of Idaho seems far from being empirically adequate.
I am rather a vicious individual who certainly has more moral failings than virtues. In all probability, I will be unable to change very quickly; it will take many years for me to develop valuable projects. Character concern seems to imply that I am unjustified in caring about myself, so I have no reasons to embark on a course of moral self-improvement. I might as well just let myself and those less than fully virtuous persons whom I know go to pot, as it were.

In developing her character-based view, Whiting even wants to say that if I am not virtuous and neither are my intimates, not only do I not have a reason to promote the projects of myself or my loved ones, but neither do I have reason to promote the projects of virtuous persons such as, perhaps, the governor of Idaho. In fact, she wants to claim that it is not simply that I do not recognize having such a reason but, rather, that I objectively do not have such a reason. Her claim is that only virtuous persons have reason to promote their own ends and the ends of other virtuous persons (Whiting 14). So if I am less than virtuous, I have no reason to have concern for myself, for my intimates, or for anyone else. Therefore, I have no reasons to try to further my projects, my friends’ projects, or the projects of the virtuous. In fact, it looks like I have no reasons to care about anyone and no one has any reasons to care about me. These conclusions seem unacceptable.

Whiting does suggest (see, e.g., 4) that non-virtuous agents do have reasons for special concern ‘in some sense’. She may intend us to love persons selectively even when they are not completely virtuous. Just as I love the virtuous for their fully virtuous characters (Whiting 13), I can love others for the virtues or other admirable qualities which they do possess. I can love persons in spite of their flaws. In so far as Whiting might accept these claims, she seems to suggest a more agent-neutral view such as the one I discuss in what follows.
On the character-based view, the virtuous person has reason to be concerned for and, thus, to promote the virtue of herself and other virtuous persons. So the reasons that the virtuous person has are not fully agent-neutral, they are not reasons for everyone to promote virtue. Specifically, they are not reason for the non-virtuous to promote virtue and valuable projects. The reasons are not relative to one agent, they are relative to the set of virtuous agents. A more fully agent-neutral Impersonalist account would hold that everyone has reasons to promote virtue. It seems plausible to suppose that everyone, be she virtuous or not, has reason to promote virtue and valuable projects. Also, it seems that one has reason to promote virtue in both the virtuous and in the non-virtuous; in fact, one may think it more urgent to promote virtue in the non-virtuous. Now we have reasons for self-improvement if we are less than virtuous, and we also have reason to help to improve those of our loved ones who have serious moral failings.

We now have arrived at a fully agent-neutral Impersonalist account of special concern, a view which I will call the *virtue-promoting view*. Each person has reason to promote virtue in the form of valuable projects. The virtue-promoting view does not justify my concern for a person by citing either that person’s relationship to me or that person’s relationship to my projects. Instead, it appeals to valuable projects, be they mine or my friend’s. So it seems that my friend’s valuable projects provide anyone with a reason to be concerned about her.

Notice that the Impersonalist view could take several forms. Impersonalism
says that I ought to promote virtue whenever I am able to do so. If an
Impersonalist claims that I ought to maximize the virtue that I promote, then she
will have a Utilitarian virtue-promoting view. But the Impersonalist need not be a
Utilitarian. For example, she might accept that there are certain constraints on
how I may treat people as I promote their virtue; Impersonalism could take a
Kantian form. In fact, we may take it that to treat a person as an end is simply to
help to promote her valuable projects. Therefore, the Kantian Impersonalist may
say, I should promote persons' virtue unless, by so doing, I thereby interfere with
someone else's valuable projects. Impersonalism, then, need not take a
consequentialist form. We can interpret the claim that we ought to promote virtue
not as an injunction to attain some goal but as a constraint on how we ought to
treat persons. In arguing against Impersonalism, then, I am arguing against a very
general view, one that could take Utilitarian or Kantian forms.

If I have reason to promote virtue in general, then it appears that I have equal
reason to promote virtue in both my mother and in the governor of Idaho. Also, I
have equal reason to promote virtue in both my future self and in the governor of
Idaho. However, the virtue-promoting view can offer an extrinsic justification of
special concern similar to the one offered by the advocate of the character-based
view. Each of us is in a better position, both epistemically and causally, to
promote virtue in ourselves and in our intimates (cf. Sidgwick 432ff.). I know my
intimates well, so I know what their projects are and which are valuable. I know
their characters and dispositions so I know better than others how to try to
cultivate valuable projects in them. Also, I interact with my intimates regularly so
I am simply in a better position to influence and benefit them. And, of course, I
know myself even better than I know my intimates, and I interact with no one on
a more regular basis than myself. So I am in an extremely good position, both
epistemically and causally, to benefit myself; in fact, I am in a better position to
benefit myself than I am in to benefit my intimates. (See Section 12 for a
discussion of the problems arising from such a conclusion.) According to the
virtue-promoting view, then, the special relationships in which I stand to certain
people does provide extrinsic reasons for concern. From now on I will consider the
virtue-promoting view as the most plausible Impersonalist account of special
concern.

9 Alternatives to the Impersonalist Account

Both the character-based view and the virtue-promoting view are instances of
what Whiting calls a generic strategy or what I have been calling the
‘Impersonalist’ approach, which “is to take some characteristic common (or at
least potentially common) to oneself and others as the ground of concern and to
claim that this justifies taking any reasons for self-concern provided by this ground
to be the same in kind with any reasons for other-directed concern provided by
this ground” (Whiting 8). According to the virtue-promoting view, I have the
same sorts of reasons to care about my future self as I do to care about my friend. I can promote virtue in both.\textsuperscript{21}

Whiting proposes that the alternative to the Impersonalist approach is an \textit{egocentric strategy} that takes the relationship in which one stands to a particular person as providing the grounds for concern for that person: “if someone stands in the right sort of \textit{relationship} to me (whatever that is) then I may have reasons to care for him the same in kind with those I have to care for myself” (Whiting 9). Whiting wants to argue that there are objectionable features of the egocentric strategy that should lead us to adopt the Impersonalist strategy. In her discussion Whiting focuses on the \textit{self-extending} version of the egocentric strategy. The self-extending view takes concern for oneself as given and goes on to justify other sorts of concern as species of self-concern. Special concern for my future self is justified because my future self is just me in the future. So the good of my future self is part of my own overall good. But now we can notice again that the relationship in which I stand to my future self is, in some respects, like the relationship in which I stand to my intimates; relations of causal influence and similarity exist in both cases. If, in the intrapersonal case, this relationship allows me to regard the good of my future self as part of my own good, then, where similar (although weaker) relations exist in the interpersonal case, I should regard

\footnote{Notice that not all views with a generic strategy are agent-neutral. We have already noticed that the character-based view is agent-relative in so far as reasons for concern are not reasons for the non-virtuous.}
the good of my loved ones as also being part of my own good. Thus, concern for others becomes an extension of self-concern (see Whiting 9; for a full discussion of the self-extending view see Brink 1990). My reasons for caring about my best friend are the ‘same in kind’ with my reasons for concern for my future self: the good of each is part of my own overall good.

Whiting holds that the main problem for such a view is in accommodating the thought that we care about our friends for their own sakes. How is that to be reconciled with viewing my friend’s good as a part of my own good? Whiting claims that it requires that just as I view my own good as mine, I should view the good of my loved ones as mine (10). She argues that this is objectionable because it does not “grant independence to the value of another’s good” (10). If the self-extending view is unable to accommodate concern for our friends for their own sakes, then we should reject the self-extending view. However, Whiting would not yet have an argument for the Impersonalist view, because the egocentric self-extending view is not the only alternative to the Impersonalist view.

In fact, the self-extending view is not an intuitively attractive view. It holds that my reason for making a sacrifice for a loved one is the same in kind as my reason for making a sacrifice for my future self; in both cases, the benefit that results from the sacrifice is a benefit to me. Any good to my loved ones is really a good for me as well. But we are often inclined to think that I should sometimes sacrifice my own interests in order to advance the interests of my children or
parents. On the self-extending view, such a sacrifice is literally impossible, because any advance in the interests of my loved ones is an advance in my interests. So we should consider an agent-relative view that preserves our intuitions.

The self-extending view is an agent-relative view, because it says that I have reason to be specially concerned about a person if that person's good is part of my overall good. The self-extending view is not, however, the only agent-relative view. Consider the agent-relative view Broad called *Self-referential Altruism* (279).²² Broad claimed that Self-referential Altruism is the commonsense position. It holds that each of us has specially urgent obligations to benefit certain individuals and groups which stand in certain special relations to himself, e.g. his parents, his children, his fellow-countrymen, etc. And it holds that these special relationships are the ultimate and sufficient ground for these specially urgent claims on one's beneficence (280).

Among other obligations, I have an obligation to be specially concerned about those who stand in special relationships to me. My special concern for my mother is justified by the fact that she and I stand in a very special relationship to one another, the relationship of mother and daughter. Like the self-extending view, Self-referential Altruism is self-referential, or, agent-relative, because it says that I owe special obligations and concern to persons who stand in certain relationships to me. Unlike the self-extending view, Self-referential Altruism does not say that the good of those others is part of my own good and so by furthering their good I am actually furthering my own good. Rather, it's altruism requires that I must

²²Whiting herself mentions this view, but only in passing (5).
sometimes sacrifice my own good for the good of my mother or for the good of my best friend. I do not view their good or their projects as my own. Thus, I do acknowledge that the good of my intimates is independent of my own.

The important thing to notice about Self-referential Altruism is that it does not engage in what Whiting calls “an unnecessary and potentially objectionable sort of colonization” (10). It does not require me, in order to justify my special concern for others, to appropriate their good as part of my own. So it seems that the Self-referential Altruist will have no problems in accommodating the claim that we should care about our friends for their own sakes. I also think that Self-referential Altruism has a fair amount of initial plausibility. As Broad says, it does appear to be the view of common sense, as I said at the beginning of this chapter. The attractive feature of the view is that it takes special relationships such as the friendship relation to provide, in and of themselves, reasons for concern. It endorses the commonsense claim about justification. We want, I think, to say that further concern is both justified and required once initial concern is in place no matter how the initial concern came about. The fact that I have been friends with Emma for over three years in itself justifies my having greater concern for her than for someone that I just met yesterday.

Whiting wants to be able to agree that “the friendship relation itself...is taken to provide reasons for concern additional to those (if any) existing prior to its establishment” (7). On the virtue-promoting view, however, the friendship relation
cannot itself provide a reason for concern. It will only be able to provide extrinsic reasons for concern. As Whiting points out, I am justified in caring more about my friend Emma whom I have known for years than about an equally virtuous individual whom I have known for only a week because I will know Emma better. I will be better able to promote her ends because I will have a better idea what her ends are. Also, I will have spent more time with Emma and will thus be better able to appreciate her virtues (Whiting 22-23). In the following two sections, I will show that this Impersonalist account of my special concern for Emma and the parallel account of concern for my future self is unable to yield plausible results about duties of friendship and the justifiability of certain private projects. Moreover, we will see that the Impersonalist cannot provide any reasons consistent with her view for taking 'the friendship relation itself to provide reasons for concern'.

10 The Rejection of the Virtue-Promoting View in the Intrapersonal Context

The virtue-promoting view tells us to promote valuable projects wherever we are able to do so. The reasons why we are under a special obligation to promote the valuable projects of our intimates and our future selves are extrinsic. We are in better epistemological and causal positions to promote virtue in ourselves and in
our intimates. I now want to consider what sort of picture of future-directed self-concern and friendship emerges when we consider the implications of the virtue-promoting view.

Let us begin with the intrapersonal case of special concern for one's future self. Why, on the fully neutral, virtue-promoting view, should I be specially concerned about my future self? I have certain projects now that I recognize as valuable and I know that my future self is the person most likely to carry out the maximum number of those projects. Also, given the causal relation in which I stand to my future self, I am in an excellent position to see that my future self adopts other valuable projects and strives to realize them. Ultimately, then, my concern for my future self is grounded on a concern for the development, pursuit, and accomplishment of valuable projects.

This sort of explanation of concern for my future self seems to leave entirely unjustified certain forms of future-directed self-concern that seem entirely rational. Many of my private projects are not grounded on nonprivate projects, and yet they do seem to be justified projects. The justification of some projects seems to depend on something other than the promotion of value, on some essentially personal component. Consider, for example, my project that I see the Mona Lisa. One may wonder, unless one is a hedonist, how valuable this particular project is. Whether or not it is a valuable project, others will see the Mona Lisa whether or not I ever see it. However, that does not seem to be sufficient to undermine my
concern for my future self that she see the Mona Lisa. Also, I may be in a position such that I can more easily get others to see the Mona Lisa than I can get myself to see it. Again, that does not seem to undermine my concern that I see the Mona Lisa and that I exert effort to insure that I see it. My project that I see the Mona Lisa is not grounded on any valuable nonprivate project. This project is essentially private and yet I seem justified in being concerned that it be fulfilled.

If special concern for my future self is, at least in part, constituted by concern for the fulfillment of private projects, then the grounds for concern cannot be simply the promotion of value or virtue. These private projects have a non-derivative, non-value-dependent personal component. Thus, if these projects are justified, and they certainly seem to be, then special concern must be, at least in part, justified by something other than the promotion of value.

What can the Impersonalist say about private projects that are not value-dependent? Here I think we want to be careful to distinguish the Impersonalist’s ability to explain the empirical psychological fact that we have non-derivative personal concern in the form of private projects from her ability to offer an appropriate justification of such concern. As we have seen, the Impersonalist can claim that, given my typically strong extrinsic reasons for being specially concerned about my future self, I will develop habitually strong concern for my future self (section 3). Eventually, I will develop special concern for my future self that is non-derivative, or not based on nonprivate projects.
Psychologically, my special concern will become dissociated from the projects that initially grounded it, just as a miser’s love of money for its own sake can develop from an initial concern for what money can buy. So the Impersonalist may claim that her account is empirically adequate.

The difficulty for the Impersonalist is that her account is not normatively adequate in that she must claim that even if, psychologically, my special concern is non-derivative, its justification is still derivative from my valuable nonprivate projects. My habitually strong non-derivative self-concern is justified only because the presence of such concern is typically necessary for the promotion of value.

Concern for my future self is necessary for the promotion of my valuable non-private projects, and if it typically leads to non-derivative self-concern, then non-derivative self-concern is justified only by reference to my non-private projects. But my point above is that my concern that I see the Mona Lisa seems justified without any reference to the promotion of valuable non-private projects. In fact, for many of my private projects, it is difficult to see what the relevant nonprivate project could be. I will return, at the end of this section, to the question of whether this Impersonalist account is even psychologically adequate.

However, Whiting claims that non-derivative personal concern can be given more than a mere instrumental or extrinsic justification. She says that “[a] certain degree of self-concern (even if only instrumental) is pragmatically necessary for

28I owe this suggestion and the example to David Brink.
agency” (21). She claims that each agent must see himself as a particular agent with his own particular history and ends. As Bernard Williams says, “his actions and his decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified” (“A Critique of Utilitarianism”, hereafter CU; 116-117). It is impossible for a person to view himself as merely a tool for the promotion of value without becoming dissociated from the projects that make him a particular agent. In fact, if he does not identify with certain non-derivative private projects, he will cease to be an effective agent for the promotion of value. Without private projects it is hard to see what value there could possibly be to promote (Whiting “Friends and Future Selves”, hereafter FFS; 579, Williams CU 110).

I think that Whiting is pointing to important considerations. However, it is not clear that the Impersonalist can accept these claims in a wide enough range of cases. Perhaps the Impersonalist can say that if I am to be an effective agent of any sort, I must have some projects that are not value-dependent and are essentially personal. But consider again my project that I see the Mona Lisa. As we noted above, this project does seem to be non-value-dependent and essentially personal. Is this project necessary for any sort of agency?

Suppose that I could use the money that I have saved either to travel to Paris to see the Mona Lisa or to provide a great benefit to a complete stranger. The benefit to the stranger would promote her virtue considerably. I think that the
Impersonalist must say that I should give my money to the stranger. My project that I see the Mona Lisa is not a project such that I could not give it up and still remain what we would properly call an agent. It does not seem to be in the same class with my project that I not murder anyone (or even with my project that I finish my thesis). All of us have many projects like my project that I see the Mona Lisa, and these projects seem justified even if their forfeiture could lead to the promotion of another's virtue. The Impersonalist appeal to the necessity of certain projects as a condition for agency is not going to justify a wide enough range of projects.

At this point, the Impersonalist could simply say that a lot of projects which we had previously thought justified are not. Among such projects is my project that I see the Mona Lisa. A concern for such projects does not effectively promote value and should be abandoned. If the Impersonalist makes this response, she has ceased to offer a justification of special concern. My having special concern for myself consists, at least in part, in my wanting to provide certain benefits to myself more than I want to provide equal or even larger benefits to persons whom I do not know. My special concern for myself involves my wanting to see the Mona Lisa more than I want to promote the virtue of a stranger. Non-value-dependent personal projects partly constitute special concern. If such projects are unjustified, then special concern is unjustified. The Impersonalist project looks wrong in the intrapersonal case, because it cannot support a concern for one's own projects.
which is not proportional to the value promoted by such projects. But such projects are part of special concern.

An egocentric or agent-relative view offers a much more plausible looking account of the phenomena that I have been discussing. I want my future self to avoid pain because I care about my future self. That concern is justified because of the very special relationship in which my future self stands to me - the relation of identity. So my personal projects do not derive their justification from an underlying concern for valuable projects in general. Rather, the personal component of my projects seems nonderivative. (In an earlier paper, Whiting seems to agree with this view. See FFS 576ff.)

In fact, we can recall again that I stand in certain relations of similarity and causal influence to my future self. According to both Mixed and Psychological Reductionism, these relations are necessary conditions of personal identity across time. The sorts of causal connections which exist between me and my future self are connections between desires and intentions and the fulfillment of those desires and intentions. My nonderivative concern for my future self is composed of such psychological connections as my desire that I see the Mona Lisa, my intention to avoid pain, my plan to hear operas, etc. So nonderivative concern is a part, and a very important part, of the special relationship in which I stand to my future self. I have been claiming that these private projects provide intrinsic reason for special concern, so it seems right to suppose that the relationship of which they are
components intrinsically grounds a more general special concern for my future self (cf. Whiting FFS 564-566). So it is wrong to claim that my special concern can be derived from an underlying concern for valuable projects. A general concern for projects is psychologically derivative from a more basic personal concern, although I will probably come to see that my private projects can be given an independent justification. And if this is the case, then we can question even the empirical adequacy of the Impersonalist view. My nonderivative concern for private projects gives rise, psychologically, to more general value-dependent concern, rather than vice versa as the Impersonalist must claim (cf. Whiting FFS 578).

Of course, the Impersonalist would be right if she claimed that private projects are not essential components of personal identity. As we saw in the previous chapter, I could persist without having the sort of concern that is composed of private projects, even though most normal cases of persistence involve such projects. Psychological continuity minus may be sufficient for survival. But again, if I am to stand in a relationship that matters to my future self, I must have such private projects. Those projects are constitutive of the sort of psychological continuity that matters; once they are in place, they provide a grounds for concern for my future self. Also, once they are in place, they are part of the psychological continuity that makes my future self me; now I could lose certain elements of psychological continuity minus and still persist. It is true of any given connection that I could survive if it were destroyed, but all connections taken together
constitute my relation to my future self. We could ask why I should have such projects in the first place, in other words why I should stand in the relationship of psychological continuity rather than psychological continuity minus to my future self. Here is where we need to point to independent considerations, considerations that arise from comparing a normal case of persistence to a case of, e.g., an irreversible coma where we have psychological continuity minus. (Of course, if one's mental life were to cease completely, one would not survive falling into such a coma. I am assuming here that one maintains some sort of minimal mental life, perhaps through random memories and imaginings.) The friend of the commonsense view can say that one relation to one's future self is more valuable than another, and still hold that once one is in such a relation to one's future self, it is the relation itself, not the value, that provides a grounds for maintaining concern. Given that most of us are spontaneously psychologically continuous in the full sense with our future selves, such worries never really arise.

11 The Rejection of the Virtue-Promoting View in the Interpersonal Context

So the problem with the virtue-promoting view in the intrapersonal context is that it does not provide an account of future-directed self-concern that explains how actual concern is developed. It simply does not ring true. But we can remember
that part of what appeared to be strong motivation for the Impersonalist approach was the plausibility of the claim that we should care about our friends and other intimates for their own sakes. So we might think that the Impersonalist view is more likely to ring true in the interpersonal context. More particularly, we might think that, given its neutral approach, the Impersonalist view might work better as an account of our special moral obligations than it does as an account of the rationality of special concern. *Prima facie* an agent-relative approach seems more properly suited to an account of rationality than of the morality of special concern. Let us, then, look at some implications of the Impersonalist account of special concern for other persons.

I will begin by looking at what is, perhaps, one of the closest interpersonal relationships, that of parent and child. Let us consider first the parent’s concern for her child. The advocate of the virtue-promoting view must claim that the parent’s concern for her child is justified by reference to the child’s virtue. There are very strong extrinsic reasons for the parent to be specially concerned about her child. Because a parent is so close to her child, she has great influence over her child and can guide and mold her child to a great extent. Thus, she can try to instill in her child the desire and willingness to further valuable projects. In fact, the Impersonalist can explain the extraordinary strength of a parent’s attachment to her child in terms of the fact that she is in an excellent position, both causally and epistemically, to further her child’s virtue from the very day of the child’s birth.
This account of the justification of parental love and concern looks very odd as did the Impersonalist account of special concern for our future selves. Of course, it would be natural and admirable for the mother to want her child to be virtuous and to pursue valuable ends. But the Impersonalist is saying that a mother’s special strong concern for her child is or should be based upon a concern for the child’s virtue or valuable projects. In other words, a parent’s concern for her child as a particular person is derived from or secondary to a more basic concern for valuable projects. However, analogous to the intrapersonal case, the parent’s concern for her child’s well-being, which includes her desire that the child be fed and warm, and her intention to provide pleasure for the child, is partly constitutive of the parent-child relationship. At least part of the reason why she wants her child to be warm and fed is because this particular child is her child. These sorts of desires and intentions are strong psychological connections and also nonderivative forms of concern for the child as the particular child she is. The parent’s concern for the child’s virtue seems to be derived from the more basic nonderivative concern that is a very part of the special relationship in which she stands to her child. So it is getting it backwards to say that the parent cares about the child because she can promote virtue in the child. Her concern for her child’s virtue is based upon and derives from the relationship in which she stands to her child.

The Impersonalist view, then, does not ring true in either the intrapersonal or the interpersonal context. In the interpersonal context, however, the Impersonalist
view also leads to some morally objectionable conclusions. Again, these are clearest in the case of the parent-child relationship. For it is not simply her own child that the parent could greatly influence and affect. There are many orphans simply waiting to be nurtured by parents. Nonetheless, surely it is the case that once a mother has her own child, she is justified in having far greater concern for her own child than for all the others out there. Otherwise, one has to wonder what would be wrong, according to the virtue-promoting view, in a parent exchanging her own baby for one whom she could more easily benefit or make virtuous. Parental feelings would usually prevent such occurrences, but it seems that we want parents to have such feelings. The Impersonalist must say that such feelings are unjustified if they prevent a parent from promoting virtue as effectively as she is able. However, the lack of such feelings seems, in itself, a failing on the part of the parent.

There are also cases where it seems that a parent’s inability to promote virtue in her child actually provides more reason for the parent to be specially concerned. Many children are born with degenerative illnesses or severe mental retardation. These factors certainly make it more difficult for a parent to promote the virtue of her child. However, it certainly would not be acceptable, let alone justified, for a parent to take one of these factors as a reason for not having special concern for her child. We have a similar case when elderly parents develop Alzheimer’s disease. What reason do I have, on the virtue-promoting account, to continue to
have special concern for my mother when she develops the disease, making it impossible for me to promote virtue in her? The answer, it seems, must be, none, because the fact that she is my mother cannot, in and of itself, provide me with reasons for special concern.

Let us now turn to the case of friendship, a case that, initially, may seem more congenial to the virtue-promoting account when we recall another motivation for the Impersonalist view. Impersonalism can give an account of how we come to be justified in entering friendships. As we have noted several times, each of us has limited resources. We have only so much time so we must choose a few people out of the many available persons who could be our friends. So we need some way of choosing whom we will befriend. Now it does seem plausible to suppose that, when I am presented with a choice, I should enter into that relationship which offers the best opportunities for me to promote virtue. We recognize virtue in persons, or we recognize a chance to promote virtue in a person, so we become friends with that person and become even better able to promote her virtue. Of course, we are only presented with choices in the case of friendships, not in the case of family relationships. I cannot choose whom I will have as my mother or my brother as I can choose whom I will have as my best friend. So the Impersonalist view may be more adequate for friendships than it is for family relationships. However, even here Impersonalism gives the wrong account.

Let us suppose that I chose to befriend Emma because I thought that a
friendship with her would enable me to promote many valuable ends that I could not promote as effectively alone or with any other available person. Once Emma and I are friends, I develop many projects which involve Emma. Should my concern for Emma still be justified by reference to our ability to promote virtue? Let us consider my project that I spend time with my friend Emma. How, according to the virtue-promoting view, can I justify my project? I must say that by spending time with Emma I will be better able to promote valuable ends. Also, spending time with someone other than Emma is unlikely to lead to the realization of as many valuable projects or opportunities for the pursuit of valuable projects. So my concern that I spend time with my best friend must be grounded on the fact that Emma and I will pursue certain projects in the future. Now this sort of justification of my entering the friendship with Emma may be morally acceptable (although, in most cases, it is far from being empirically adequate), but it is not a morally acceptable justification of my maintaining the friendship with Emma. In fact, such a justification makes a travesty out of friendship. It certainly does not represent what we think the best sort of friendship is like. For, on the virtue-promoting view, I should continue to see my friendship with Emma as simply an expedient way to promote valuable ends that I could also pursue with others. So if I meet someone better equipped to my purposes, I should ignore Emma and spend time with the new person. But I care about Emma’s projects because I care about Emma, and I make projects involving her because I care.
about her. She and I may share ends that we both consider valuable, but we care about each other independently of caring about our shared ends. Once Emma and I stand in the friendship relation to one another, I should not view her as one of many persons with whom I could promote virtue. The fact that I stand in such a relationship to her means that I must view her differently than I view others. I now have special obligations to her. If continuing to be friends with Emma means that some of my valuable projects may suffer, then, at least up to a certain point, I am required to sacrifice the pursuit of some of my projects.

The Impersonalist could say that special concern is not justified in most cases. Most of us have special concern when it is not warranted. Again, I am concerned with the adequacy of the Impersonalist view as an account of special concern. A view that would yield that the vast majority of friendships are unjustified is not such an account. So can the Impersonalist justify the maintenance of friendships? The advocate of the virtue-promoting view might respond by saying that she is not committed to saying that I am sometimes justified in simply ceasing to be friends with Emma. She could try to defend her claim in one of two ways. First, she might try to claim that the sorts of circumstances that arise in close friendships are necessary for the realization of valuable projects. For example, if Emma and I are close and committed to one another we will be more likely to achieve our valuable goals. We will be less likely to misunderstand or distrust one another. This sort of response is inadequate. It is at least possible that an
opportunity will arise when such circumstances are irrelevant to the pursuit of valuable projects or even hinder such projects. Emma and I may become so committed to one another that we become less committed to our valuable ends.

There is another response that the Impersonalist might make. She may claim that the circumstances of closeness and intimacy are themselves valuable or are necessary conditions of valuable goods such as the experience of intimacy. If closeness and intimacy are intrinsically valuable goods, then, the Impersonalist can say, the friendship relation does provide intrinsic reasons for concern because closeness and intimacy are constitutive elements of the friendship relation. Once Emma and I are close and intimate friends, the presence of valuable elements in our relationship entails that Emma and I have special duties of concern to one another. These duties are grounded by the nature of the relationship in which we stand to one another, a relationship with intrinsically valuable components. Only by having special concern for each other will Emma and I have the sort of relationship that has intrinsic value.

I think that the Impersonalist is right in claiming that closeness and intimacy are valuable and ought to be promoted. If I have an opportunity to further any close and intimate relationship between two persons, all else being equal, I ought to try and do so. However, I do not think that we should ground special concern on the value of closeness and intimacy. If we do so, then what reason do I have for maintaining and furthering my special relationships with my intimates that I do
not have for furthering and helping to maintain any special relationship between any two persons? We want to be able to say that I am justified in having special concern for my intimates and, thus, that I have special duties to promote valuable components of my relationships to my intimates. If special concern is justified in terms of the value of the relationship, then my duty is to promote special concern and special relationships wherever I am able to do so. And there will be cases where I can further others' relationships better than I can further my own relationships with my intimates. So while I want to agree that certain components of special relationships are valuable and that these components should be promoted, I want to argue that special concern should not be justified by reference to the value of those components, otherwise special concern will too often be trumped by our more general duties to promote concern as such.

12 Self-referential Altruism and Self-Concern

I have suggested Self-referential Altruism as a plausible agent-relative alternative to the Impersonalist view. Self-referential Altruism holds that each of us has special obligations resulting from the special relationships in which each of us stands to 'his parents, his children, his fellow-countrymen, etc.' My main purpose in this chapter has been to show that an agent-neutral approach to special concern is implausible, rather than to argue directly for a particular agent-relative view. Nonetheless, I want to consider Self-referential Altruism in a bit more detail so
that we can see what the advantages of the view are. I also want to point to some difficulties for the view which I will not attempt to deal with here.

To begin, we need to be clear on exactly what Self-referential Altruism says and what it does not say. Self-referential Altruism is, at least initially, intended to be an account of special obligations, not of all obligations. It says that I have, all else being equal, an obligation to benefit my mother rather than a stranger. Nonetheless, I may, in certain circumstances, have an obligation to help the stranger. If the benefit that I can give my mother is trivial compared to the immense benefit that I can give the stranger, then I ought to benefit the stranger. I should save a stranger’s life before I give my mother five dollars. Self-referential Altruism is not a denial of the claim that we have duties to even complete strangers; rather, it is the claim that, all else being equal, we ought to benefit those persons to whom we stand in special relationships.

The Self-referential Altruist may claim that our duties to persons other than our intimates are grounded by something other than the relationship in which we stand to those persons. But we can see Broad claiming that the relationship between fellow-countrymen is a special relationship giving rise to special duties, probably duties such as loyalty. Presumably, one’s duties to one’s children and parents will be stronger than one’s duties to one’s countrymen, because one’s relationship to one’s parents and children is stronger than one’s relationship to one’s countrymen. So Self-referential Altruism can provide an account of special
duties that extend beyond intimates and future selves to not only neighbors and colleagues but also to fellow-countrymen. And the case of fellow-countrymen suggests that the independent basis of our obligations to all persons may be the relationships in which we stand to those persons. I do stand in certain relationships to all persons; we are all morally equal, rational beings. Perhaps this relationship grounds some duties, albeit duties more minimal than those owed to intimates and countrymen. This sort of view sounds plausible and would yield plausible conclusions about the relative weight of our various duties. Special relationships of psychological continuity will ground special concern for future selves, intimates, and fellow-countrymen, while the relationships in which all rational beings stand to one another will ground a more general sort of concern for all persons.

My special duties to myself are my strongest duties because of the very strong relationship in which I stand to my future self. This can seem an attractive feature of Self-referential Altruism, because it will help keep my special duties to intimates and countrymen from dominating my life. I do not have to and am required not to completely abandon my own self-development for the sake of my children or parents. But, what about cases where I can provide an equal benefit to or prevent an equal burden on either myself or my small child? If I have stronger reason to be concerned about myself than I have to be concerned about my child, then I should take the benefit or impose the burden on my child. Surely, though, a parent ought to give the benefit to her child or she should impose the burden on herself. At the
least, she is not irrational or morally wrong if she does so, but Self-referential
Altruism yields that it is irrational and immoral of her to help her child instead of herself in this case.

How can the Self-referential Altruist deal with this difficulty? One strategy is for her to point out that, given the dependence of a child on his parents, a parent is typically able to provide great benefits to the child, benefits that outweigh the burden imposed on the parent. In normal cases, then, a parent's duty will be to her child. Therefore, a parent has a moral duty to develop an extremely strong disposition to benefit the child, a disposition that will lead her to benefit her child even when her duty requires her to benefit herself instead. Parents are required to develop dispositions that will lead them, in extraordinary cases, to do the wrong thing, so they will not be blameworthy for doing the wrong thing in those cases. But it is still odd that a parent does something wrong in any case when she makes a sacrifice so that her child would not have to make that sacrifice.\(^{24}\)

The only thing that I can say here is that the Impersonalist has the same difficulty. Given that my relationship to my future self is stronger than my relation to any other person, my extrinsic reasons for benefitting myself will always be stronger than my reasons for benefitting others. I can usually be more confident in

---

\(^{24}\)This problem for the Self-referential Altruist recalls the example in the previous chapter where either I can see the Mona Lisa or my daughter can see the Mona Lisa but we cannot both see it. The question there was whether I should be disappointed if my private project is not realized and also whether I should prefer that my private project or that my quasi-private project involving my daughter be realized. I think that a fuller account of the role of private projects, particularly in relation to our projects involving our intimates will help to shed light on the problem concerning differential levels of concern. Unfortunately, however, I must leave these issues for another time.
determining how to promote my own valuable projects than I can be in
determining how to promote anyone else's, even my child's. Of course, the
Impersonalist can point out that my reasons for having greater concern for myself
are only extrinsic reasons, that, on her view, my intrinsic reasons for concern for
my child are just as strong as my intrinsic reasons for concern for myself. But, on
the Impersonalist view, my intrinsic reasons for concern for the governor of Idaho
are just as strong as my intrinsic reasons for concern for either myself or my child.
I have the same intrinsic reasons for promoting virtue in one person as I have for
promoting virtue in any other given person. For the Impersonalist special concern
is a function of extrinsic reasons arising from the special causal and epistemic
positions that I have with respect to my intimates and my future self. So the fact
that my extrinsic reasons for concern for myself are stronger than my extrinsic
reasons for concern for my child is as much a problem for the Impersonalist as is
the fact that my intrinsic reasons for concern for myself are stronger than my
intrinsic reasons for concern for my child a problem for the Self-referential Altruist.

Recall that, in motivating the Impersonalist view, I said that the Impersonalism
may look attractive in so far as it offers an explanation of why I am justified and
required to have special concern for myself and my intimates. My duty is to
promote virtue, and, the Impersonalist points out, it just so happens that I am
able to promote virtue best if I concentrate my attentions on myself and my
intimates. The Self-referential Altruist, on the other hand, seems to say just that I
have reason to be specially concerned about persons to whom I stand in special relationships because I stand in special relationships to those persons. Why suppose that special relationships can ground concern?

I think that, once we look at the nature of those special relationships, it becomes clear why they can function as the grounds of concern. The causal connections that help to constitute these relationships involve, in the intrapersonal case, private projects. My desire that I finish my thesis and my future self’s working on the thesis in order to finish it make a strong psychological connection between me and my future self, one of the connections that makes it the case that that future self is my future self. These projects, in a very literal sense, define and shape my life. My commitment to them is a part of what makes me the person who I am. Similarly, my connections to my intimates is a part of what makes them my intimates, the connections shape and define their role in my life. These connections, in so far as they are central to my identity and represent my commitments, seem to demand and warrant a very special concern.
Chapter 5
Persons, Compensation, and Utilitarianism

13 Introduction

According to Utilitarianism, we ought to maximize benefits and minimize burdens and it is not morally significant, in itself, who gets the benefits and burdens. This implies that we ought to balance benefits and burdens between persons. Just as a person should, all else being equal, impose a burden on herself at one time in order to receive a larger benefit at a different time, so should we, all else being equal, impose a burden on one person in order to provide a larger benefit to a different person. But, some critics of Utilitarianism have pointed out, if a person bears a burden and receives a larger benefit, she is compensated. However, if we impose a burden on one person and someone else receives the benefit, the first person has not been compensated. Because Utilitarianism treats the interpersonal case like the intrapersonal case, Rawls claims that it “does not take seriously the distinction between persons” (27). Call this objection to Utilitarianism the ‘separateness of persons’ objection.

Derek Parfit argues that if we accept that Reductionism commits us to the claim that psychological continuity, not identity, is what fundamentally matters in
our survival, then we should not find the separateness of persons objection compelling. According to Parfit, if identity is not what matters, then it cannot be a ‘deep’ truth that each of us is a separate person. Therefore, compensation is not morally important and we can balance benefits and burdens between persons as well as within lives. So Parfit is claiming that a Reductionist view of personal identity supports Utilitarianism by undermining the separateness of persons objection.

In the last chapter, I argued that we can view Reductionism, Mixed or Psychological, as motivation for a particular agent-relative account of special concern, an alternative to an agent-neutral approach. Utilitarians, of course, are among those who offer an agent-neutral account. Even though I argued in chapter 3 that Reductionists are not committed to the claim that identity is not what matters, I will consider, in this chapter, whether Parfit is right that such a claim would offer support for Utilitarianism. I will show that Parfit’s arguments not only fail to support Utilitarianism, but, in fact, suggest a moral theory in which relationships in which persons stand to each other, in and of themselves, determine certain duties and obligations. (In what follows I will focus on Psychological rather than Mixed Reductionism because Parfit accepts a version of Psychological Reductionism.)
14 Utilitarianism and the Separateness of Persons

14.1 Utilitarianism

Utilitarianism is a *teleological* ethical theory. Teleological theories have two parts: (i) a theory of value or the good, and (ii) a principle which states what the right is in terms of what has value, or, more specifically, which states that the right is whatever maximizes value (cf. Rawls 24). Utilitarianism is distinguished from other types of teleological theories by its theory of value, according to which what has value is the happiness of human beings, or any other sentient creatures. (Although Utilitarian theories differ over the issue of what constitutes happiness, this issue is not important for my purposes; therefore, I will use ‘benefits’ to refer to whatever has positive value and ‘burdens’ to refer to whatever has negative value.) I will be concerned only with *act-utilitarianism*, which evaluates individual actions in terms of the amount of benefits which they bring about. According to act-utilitarianism, a right action is that action which of the alternatives available to the agent brings about the greatest net amount of benefits. The agent is morally required to maximize benefits and, in so doing, he should not be concerned about who gets the benefits and burdens. In other words, Utilitarianism is *person-neutral*: it claims that it is not morally important, in itself, who gets the benefits and the burdens. Only the net amount of benefits produced is morally
important.

Why might one be motivated to accept a person-neutral moral theory such as Utilitarianism? To begin, let us put aside the moral question and consider some questions of intrapersonal rationality. Suppose that I can accept a certain benefit later only if I forego a smaller benefit now. Assume that I can be certain of getting the larger benefit if I sacrifice the present benefit, and that I do not forego any other benefits by sacrificing the present one. It is plausible to claim that, rationally, I should forego the present benefit and wait for the larger benefit. The mere fact that one benefit is a present benefit and the other a future benefit should not matter to me - both are benefits to me. Mere temporal placement of benefits does not seem to be rationally significant; therefore, when someone is the only person concerned, it seems plausible to suppose that he should maximize benefits within his life, or ‘aim at his good on the whole’ (Sidgwick 381). I have no reason not to get as many benefits in my life as I can possibly get. Of course, I may decide that the way to achieve my overall good is to distribute benefits evenly throughout my life, but, at least initially, it seems that I should only do so if that is the way to achieve the most benefits possible.

But, Sidgwick says,

just as this notion [i.e. “the ‘Good on the Whole’ of a single individual”] is constructed by comparison and integration of the different ‘goods’ that succeed one another in the series of our conscious states, so we have formed the notion of Universal Good by comparison and integration of the goods of all individual human - or sentient - existences. And here again, just as in the former case, by considering
the relation of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any one individual is of no more importance, from the point of view...of the Universe, than the good of any other... And it is evident to me that as a rational being I am bound to aim at good generally...not merely at a particular part of it(382).

In the purely intrapersonal case where I am the only person affected, all else being equal, it should not matter when I receive a benefit: whether I receive a benefit at one time or at some other time, I still receive a benefit. From the perspective of rationality, what is important is that I maximize benefits within my life. But morality seems to demand that I consider not only my own life but all lives ‘from the point of view of the Universe’. Then, just as temporal placement of benefits in the intrapersonal case is, in itself, unimportant, it seems that, analogously, it is unimportant in the interpersonal case who receives a benefit. It should not matter whether I receive a benefit or you receive a benefit, as long as someone receives a benefit, because no person’s good is more important than the good of any other. Our moral theory, then, should be person-neutral. We should be concerned only with maximizing benefits among all persons as Utilitarianism does. Thus, Utilitarianism can be motivated by viewing the morality of distribution of goods among persons as analogous to the rationality of distribution of goods within a single life.
14.2 The Separateness of Persons and Compensation

Person-neutrality, however, is the very feature of Utilitarianism that leads to the objection that it does not take seriously the 'separateness of persons'. What is the 'separateness of persons'? Dennis McKerlie says that "[d]ifferent people live different lives. Each life consists of experiences that are not shared with the other lives. These facts are sometimes referred to as the 'separateness of persons'" (205). As Sidgwick claims, "[i]t would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental" (498). The claims that 'different people live different lives' and that the distinction between lives is 'real' are obvious truths. Surely, Utilitarianism does not deny these metaphysical truisms. The critics' charge is that Utilitarianism does not treat these truisms as 'fundamental'. Utilitarianism does not respond appropriately to the separateness of persons.

In what way does Utilitarianism respond inappropriately to the separateness of persons? Robert Nozick argues:

There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Nothing more. What happens is that something is done to him for the sake of others ... To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. He does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him (33).

Suppose that I can provide one person with a large benefit only if I impose a
smaller burden on some other person. Utilitarianism says that, all else being
equal, I am morally required to impose that burden because, by doing so, I
produce a net benefit. But, as Nozick points out, the person who bears the burden
does not receive an ‘overbalancing good’ - he does not receive a net benefit, a
benefit to some other person will not always make his life go better, because he is
a separate person. So it might be said that, given that each of us is a separate
person, we cannot force someone to bear a burden merely because, by doing so, we
could provide a larger benefit to a different person. In other words, a larger benefit
to one person will not always ‘morally outweigh’ a burden on a different person.
This claim is what Parfit calls the ‘Objection to Balancing’ (337; cf. Nozick 33,
Rawls 28). The Objection to Balancing is simply that we cannot treat the
interpersonal case in the same way that we treat the purely intrapersonal case;
even though it may be permissible to balance benefits and burdens within a life, it
is not permissible to balance benefits and burdens between lives. As Nagel says,
“[t]o sacrifice one individual life for another, or one individual’s happiness for
another’s is very different from sacrificing one gratification for another within a
single life” (138). The explanation for the disanalogy is that, whereas there is
intrapersonal compensation, “there can be no interpersonal compensation for
sacrifice” (Nagel 142).

If we are to compensate a person for a burden that we have imposed on him, we
must provide him with a benefit that counterbalances or makes amends for the
burden. We have compensation when the person who bears the burden is identical with the person who receives the benefit, i.e. there is intrapersonal compensation. In the interpersonal case, however, where one person bears the burden and some other person receives a benefit, the first person is not compensated for his burden. Now we do often attribute moral significance to compensation. For example, if you are speeding recklessly and you hit me with your car as a result, thereby causing me to be hospitalized for six months, we think that you have a moral obligation to compensate me for the hardship which you have forced upon me. You must in some way make amends for the harm that you have wrongly inflicted upon me. Suppose, however, that you were rushing to the hospital because you have been poisoned and, if you do not slow down to avoid hitting me, you will reach the hospital in time to receive an antidote that will prevent you from having to be hospitalized for eight months. In this case, the benefit to you is greater than the burden on me (total aggregate hospitalization is reduced by two months if you hit me). Nonetheless, it still seems that you ought to compensate me for my six months of hospitalization, because the benefit to you did not compensate me for my burden.

These sorts of considerations support the following principle:

---

25It is true that if you are to compensate me, you must provide me with a benefit; however, in certain cases, you can provide me with a benefit by providing someone else with a benefit. For example, I have a desire that my loved ones be benefitted; therefore, if you benefit someone I love, you thereby provide me with what I will call an 'indirect benefit'. To compensate me, then, you must provide me with either a direct or indirect benefit that makes up for the burden that you have imposed on me.
Morality of Sacrifice: There are burdens such that if you impose one of those burdens on me, then you ought to compensate me for the burden.

Utilitarianism, however, says that each person should maximize net benefits. Thus, suppose that, after you hit me with your car, you can either pay for my hospital expenses or use your money to provide a larger benefit to some other person. According to Utilitarianism, you ought to give the money not to me but to the other person. But you will not compensate me by giving that other person the money; therefore, if Morality of Sacrifice is true, Utilitarianism is wrong.

It is important to notice that I have stated Morality of Sacrifice in such a way that it allows cases in which I can impose a burden on you and yet not be required to compensate you. For example, suppose that you and I are standing by a lake when I spot a child drowning in the lake. In order to save the child, I must push you out of my way so that I can dive into the water. You fall and scratch your knee. I have imposed a very small burden on you in order to benefit greatly the child - it is not at all clear that I have any obligation to compensate you for your scratched knee. Nozick (33) and Rawls (4) sometimes speak as if we can never balance benefits and burdens between persons, but it seems that we sometimes can if the benefit is very great and the burden very small.

We can connect these considerations about compensation with the claims about private projects in previous chapters. Each person has many private projects and, in fact, most of our projects are private projects. My project that I finish my thesis is a private project. Now, benefits to some other person will not always help
me to finish my thesis. So if someone imposes a burden on me that disrupts my work on my thesis, I must receive a counterbalancing benefit, a benefit that will help me to further my thesis or some other of my private projects. If I do not receive such a benefit, my project may not be realized. Given the fact that private projects are central to my life and actually are the core of my relationship to my future self, most interferences with my private projects will demand compensation, because benefits to someone else may not help me to further my projects.

So it is a significant fact that you and I are separate persons because someone cannot always compensate me by providing you with a benefit. I can impose a burden on myself in order to receive a larger benefit, because the benefit makes up for my burden - there is intrapersonal compensation. But, because there is no interpersonal compensation, it is not always permissible to impose a burden on one person merely to provide a larger benefit to some other person. Because Utilitarianism treats the interpersonal case like the purely intrapersonal case, it is said that it does not accord the proper moral significance to the boundaries between lives.
15 Reductionism and the Separateness of Persons

Parfit offers two arguments that are supposed to show that the above objections to Utilitarianism are weakened if we accept a Reductionist view of personal identity. His second argument, which I will discuss in section 4, is directed against the claims about compensation which are supposed to explain the significance of the separateness of persons claim. His first argument, which I will discuss in this section, is directed against the separateness of persons claim itself. Parfit argues that Psychological Reductionism implies that the separateness of persons is not a 'deep' fact. Therefore, he claims, the boundaries between lives cannot be as significant as we had thought.

15.1 Parfit’s Argument: The ‘Depth’ of Personal Identity

The critics of Utilitarianism insist that, in distributing benefits and burdens among persons, we ought to treat the boundaries between lives as morally significant as Utilitarianism does not. Parfit argues that on Psychological Reductionism the boundaries between lives are not as significant as we had thought, because, on the Psychological view, identity is not as ‘deep’ as we had thought it to be. Recall Sidgwick’s claim that the distinction between one person and any other person is ‘real and fundamental’. Parfit claims that acceptance of
Reductionism should lead us to find that distinction less ‘real and fundamental’.

Why is identity less ‘deep’ on the Psychological view than it is on Non-Reductionism? Parfit offers two reasons for his claim about depth. Consider the fission case where I split like an amoeba. Each of the resulting persons in this case, A and B, is both psychologically and physically continuous with me.

Nonetheless, neither A nor B is identical with me, because psychological continuity has taken a branching form. But I would have survived as either A or B if the other had not existed. So my not being identical with A simply comes to the fact that B happens to exist. B’s existing, however, has no intrinsic effect on me, A, or the relationship between me and A. But, then, how can we view the distinction between me and A as ‘real and fundamental’? The fact that A and I are distinct persons simply comes to the extrinsic fact that B happens to exist, and it is difficult to see how that extrinsic fact can matter. As we saw in Chapter 3, Parfit concludes that it is not identity that matters, because identity depends upon extrinsic facts. What matters is that I am psychologically continuous with A whether or not I am identical with her. Thus, I should change my private projects to quasi-private projects.26

The fission case is, of course, a very unusual case. Nonetheless, the fission case shows that psychological continuity is what matters, and psychological continuity

26If we are Mixed Reductionists who accept this argument, we would conclude that I should have the mixed-quasi-private project that someone psychologically and physically continuous with me finish my thesis, unless we accept Parfit’s further argument for the claim that physical continuity cannot be part of what matters (282-287).
is present not only in relationships to future selves and fission products, but also in ordinary interpersonal relationships. Even though the only persons with whom you are fully psychologically continuous are fission products, you can bear some psychological connections to friends and family members: intimates often causally influence one another's beliefs and desires, share goals and projects, etc. Also, connections usually weaken over time within a life: beliefs and values change, memories grow dim, goals and projects are altered or abandoned, etc. So, on the Psychological view, the connections which hold between persons are not different in kind, and sometimes not even different in strength, from the connections which determine personal identity across time, although there will be a larger number of stronger connections in the intrapersonal case. 27

Parfit claims that these facts about Reductionism and psychological continuity support our giving less significance to the boundaries between lives. The difference between lives is simply a matter of degree, not a deep difference in kind. What matters in my relationship to my future self is also present in my relationship to other persons. The mere fact that some future person is me is not, in itself, important. What is important is that I am psychologically continuous with that person; I should replace my private projects with quasi-private projects. Because elements of psychological continuity are present in interpersonal contexts, we

27 If one is a Mixed Reductionist, one will hold that certain intrapersonal connections, physical connections, are only present in very special interpersonal cases, such as fission cases and cases of mothers and their children.
should “regard the subdivisions within lives as, in certain ways, like the divisions between lives” (Parfit 334). Reductionists should be open to viewing intrapersonal contexts as analogous to interpersonal contexts, as Sidgwick suggested in motivating Utilitarianism. The separateness of persons should no longer seem like a fundamental fact, because the separateness of persons is only a matter of degree.

15.2 Response to Parfit: Reductionism and Full Person-Neutrality

Parfit claims that “[w]e ought to be Reductionists. If this is a change of view, it supports several changes in our beliefs about...morality” (347). I agree with Parfit that acceptance of Psychological Reductionism should lead us to make some revisions in our moral views, but, unlike Parfit, I do not think that Psychological Reductionism lends plausibility to a fully person-neutral moral theory such as Utilitarianism.

First, recall that Parfit’s reason for claiming that identity is less ‘deep’ on the Psychological view than we had thought it to be derives from a consideration of the fission case. If you undergo fission, the distinction between you and A comes to nothing more than the fact that B happens also to exist. Therefore, Parfit concludes, the distinction between persons is not a deep distinction. But what has Parfit really shown? I agree that it does not seem to be a ‘deep’ fact in the fission case that you are not identical with A. We should conclude that the boundary
between your life and A’s life is not morally significant. A moral theory, then, can justifiably claim that it cannot matter, in itself, whether you or A receives a certain benefit. Utilitarianism, however, says that it is not morally important, in itself, whether one person, or any other person receives a certain benefit, that is, Utilitarianism is fully person-neutral. Does the claim that the difference between you and A is not a ‘deep’ fact support the claim that the difference between any two persons is not a ‘deep’ fact? I think not. You bear many strong connections to A, but, at the most, you bear some very small number of weak connections to the governor of Idaho. The relation between you and the governor of Idaho is not like the relation between you and A in the fission case. So we can continue to view the boundary between, for example, your life and the life of the governor of Idaho as morally significant. The Psychological view, so far, supports only a very small revision in our moral views, i.e. a revision concerning how we treat fission cases. Parfit’s argument depends upon extrapolating from the claim about we should treat relations between persons and their fission products to a claim about how we should treat all interpersonal relations. Given the special nature of fission relations, this is an unwarranted extrapolation.

Recall, however, that psychological continuity can be a ‘matter of degree’, i.e. you can bear some psychological connections to persons other than your fission products. For example, intimates often have similar beliefs, goals, values, etc., and they causally influence one another’s beliefs, etc. You bear a great many
psychological connections to your best friend or to your spouse. You bear fewer and weaker connections to colleagues, acquaintances, etc. Given that on the Psychological view continuity is what matters, not identity, we should cease to see the distinction between a person and his intimates as being as 'deep' as we had thought it to be. Again, I agree with Parfit on this point. But your relation to your spouse or best friend is still very different from your relation to the governor of Idaho. In fact, your relation to your spouse or best friend is significantly different from your relationships to the vast majority of persons. The Psychological Reductionist view decreases the number of morally significant boundaries, but it does not show that none of the distinctions between lives is 'deep'; it does not support the full person-neutrality of Utilitarianism.

So the claim that psychological continuity is what matters should lead us to draw new moral boundaries rather than leading us to reject all moral boundaries. When we accepted that identity matters, we saw a sharp distinction between each individual life. Now if we accept that psychological continuity matters we should accept that distinctions between appropriately related persons are not sharp but are matters of degree. Just as I am causally connected to my future self, so am I causally connected, albeit to a lesser degree, to my mother and, to an even lesser degree, to my colleagues. So the fact that my mother and I are separate persons is not a very important fact, certainly not as significant as the fact that the governor of Idaho and I are separate persons. Nonetheless, there is still a distinction
between me and my mother and an even greater distinction between me and the governor of Idaho. Our moral theory, then, should reflect these gradations in significance of the distinction between persons. It might be reflected by, for example, allowing more interpersonal balancing between intimates than between strangers, by allowing that a benefit to my mother will outweigh a burden on me in cases in which the same benefit to the governor of Idaho will not outweigh the same burden on me. And, of course, I myself will be required to make certain sacrifices for my mother that I am not required to make for the governor of Idaho.

This sort of moral theory is not person-neutral; rather, on this view, the relationships in which persons stand to one another have intrinsic significance.

We can make a similar point if we consider private and quasi-private projects. Parfit argued that if identity is not what matters, then I should abandon my private projects. But if psychological continuity does matter, I am justified in replacing my private projects with quasi-private projects. I am justified in continuing to care that at least someone psychologically continuous with me finish my thesis rather than that my thesis be finished by a complete stranger. So I can hope that at least my fission product finishes my thesis. My project will only be fully realized if my future self or my fission product finishes my thesis, but, because my friend is continuous with me to some degree, my project will be partially realized if my friend finishes my thesis. (Perhaps my friend must bear some relevant connections to me such as sharing an interest in philosophy for this
to be the case.) But my project will not even be partially realized if the governor of Idaho finishes my thesis. The distinction between me and some other person is more or less significant depending on the degree to which that other person can realize my quasi-private projects.28

The claims about the separateness of persons are, however, intended to support certain claims about compensation. If Parfit can undermine the claims about compensation, he will have lessened some of the import of the separateness of persons claim. He will have forced us to wonder why even degrees of separateness should matter. So let us turn to Parfit’s argument against Morality of Sacrifice.

16 Psychological Reductionism and Compensation

Parfit argues that Psychological Reductionism supports both (i) a change in the scope of distributive principles such as Morality of Sacrifice, and (ii) a change in the weight of such distributive principles. According to Parfit, the Psychological view should lead us to widen the scope of Morality of Sacrifice. Since the Utilitarian ignores Morality of Sacrifice, widening the scope of the principle offers no support for Utilitarianism. But, Parfit claims, we should also give the principle

---

28If we accepted that physical continuity is part of what matters, the distinction between me and another person would be more or less significant depending on the degree to which that other person can realize my mixed-quasi-private projects. My fission products and children will be able to realize these projects to a greater degree than any other person will be able to.
no weight; once we do that, it makes no difference what the principle's scope is. I will focus, then, on Parfit's claims about weight; however, I will begin with his claims about the scope of Morality of Sacrifice.

16.1 The Scope of Morality of Sacrifice

To begin, recall the case that I described in section 14.2 above: you hit me with your car and cause me to be hospitalized for six months. In order to compensate me for that burden, you must provide me with a counterbalancing benefit; for example, you ought to pay for my hospital expenses and give me some extra money to make amends for my pain. Suppose that it takes twenty years of court battles before I am able to get any money from you - I was a child at the time of the accident and I am an adult when you finally pay me. Parfit claims that in cases such as this we might think that you cannot compensate my childhood self by giving benefits to my adult self. After all, connections weaken over twenty years - beliefs, values, etc., change a great deal. Thus, because, e.g., my adult self is similarly weakly connected to other persons, Parfit says that we should "treat weakly connected parts of one life as, in some respects, or to some degree, like different lives" (337). Because benefits to one person do not always compensate another person for his burden, we should think that benefits to the adult may not compensate the child. In such a case, it would not even be permissible to balance freely benefits and burdens within a life. Parfit concludes that Psychological
Reductionism supports our giving distributive principles concerning compensation "more scope, so that they apply even within a single life" (335).

Notice that Parfit is not denying that you can compensate me by giving me the money twenty years after the accident. He is denying that you can compensate my childhood self by providing a benefit to my adult self. At this point, Parfit is assuming that my childhood self is an entity distinct from my adult self and that both my childhood self and my adult self are entities different from me. Parfit's shift in focus raises two questions, a metaphysical question and a normative question. The metaphysical question is, are there such entities? Some philosophers do not think that there are such entities as my childhood self and my adult self; rather, they claim that we have only one temporally extended entity, the person, who was a child twenty years ago and who is now an adult. In chapter 2, when I discussed the definition of 'continuity', I said that I would remain neutral as to whether we should understand talk of persons at times as talk about literal temporal parts of persons. I think that the more interesting issue for our purposes is whether or not, even if there are temporal parts, we should be concerned with temporal parts rather than with persons. Therefore, I am not going to discuss this metaphysical issue. I will grant to Parfit his assumption that my childhood self is an entity distinct from my adult self, but I think that it is important to notice that Parfit and I are both making this assumption and that it

---

29 See, e.g., Judith Thomson, "Parthood and Identity Across Time" 201-221.
is a controversial assumption.\textsuperscript{30}

I will, then, address only the normative issues that arise from Parfit’s shift in focus. My objection to Parfit is that he has not given us sufficient reason to think that Morality of Sacrifice should be changed to

\textit{Morality of Sacrifice II}: There are burdens such that if you impose one of those burdens on a temporal part of me, e.g., my childhood self, then you ought to compensate that temporal part of me.

Parfit simply points to the fact that parts of a life can be weakly connected, as weakly connected as different persons. But we need to consider whether the weakening of connections within a life is such that we should think that parts of a person rather than persons are the morally relevant units with respect to compensation. Suppose that in the case where you are driving recklessly and hit me with your car, you cause me to suffer brain damage. Brain damage will, of course, weaken the psychological connections between my pre-accident self and my post-accident self - I will lose memories, some of my abilities will be impaired, personality changes will occur, etc. Also, as a response to these changes, I will have to change my life plans, thereby weakening the connections within my life even further. It seems that you ought to compensate me for the brain damage that you have caused. But, according to Parfit, you should compensate that part of me which bears the burden. But the weakening of connections happens to me, not to any part of me - I should be compensated for your having lessened my life.

\textsuperscript{30}Parfit himself does realize this at several points. See, for example, p.135.
prospects. In this case, it seems that the very weakening of connections is what makes it such that you ought to compensate me; therefore, the weakening of connections within a life cannot be, in itself, a reason for thinking that Morality of Sacrifice should be changed to Morality of Sacrifice II.

Parfit's shift in focus from persons to smaller parts of persons raises the issue of why we thought Morality of Sacrifice was important in the first place. Why does compensation of persons seem morally important? Why ought you to compensate me if you impose certain very great burdens on me? I think that it is clear from the brain damage case that something about a person's life prospects and plans are at issue. Persons have plans involving their lives as a whole, and in those plans they allow for certain changes to occur in their personalities, etc. A person's life plans should not be constantly jeopardized by burdens imposed on him by others; thus, if certain very large burdens are imposed on a person, he ought to be compensated. It just is not at all clear that similar concerns will remain if we shift focus to smaller parts of a person. However, I will not pursue this issue further here, because it raises the issue of whether a person is justified in having plans concerning his whole life - Parfit argues that, on the Psychological view, this claim is undermined to some extent. But I do not want to look at that argument here.

I have only suggested a possible response to Parfit's claim about the change in scope. Parfit's important claim, in so far as his 'defense' of Utilitarianism goes, is that distributive principles such as Morality of Sacrifice should be given no weight.
So let us turn to Parfit’s argument for the claim that Morality of Sacrifice should be given no weight.

16.2 The Weight of Morality of Sacrifice

Parfit argues that compensation is not morally significant and, that, therefore, the Utilitarian is free to balance benefits and burdens between lives as well as within lives. His argument is as follows:

Compensation presupposes personal identity. On the Reductionist view, we believe that the fact of personal identity over time is less deep, or involves less. We may therefore claim that this fact has less moral importance. Since this fact is presupposed by compensation, we may claim that the fact of compensation is itself morally less important (338).

I agree with Parfit that ‘compensation presupposes personal identity’, that if you are to compensate me for a burden imposed on me, then you must provide me with some counterbalancing benefit. Of course, as I pointed out in section 14.2, sometimes you can indirectly benefit me by benefitting one of my intimates; however, I must receive some benefit, direct or indirect, if I am to be compensated. Does this fact, together with the fact that identity is less ‘deep’ on the Psychological Reductionist view, support Parfit’s claim that compensation is not as morally important as we had thought it to be?

Parfit sometimes writes as though he thinks that the mere fact that we have rejected Non-Reductionism and accepted Reductionism justifies our changing our
moral views as well. For example, he says that "[a] change in view often makes it plausible to give to a moral principle a different weight. If this cannot be plausible in the present case, this needs to be shown. I believe that it could not be shown" (344). But I think that it is clear that the burden of proof lies with Parfit. He needs to show how a lack of 'depth' undermines our attribution of moral significance to compensation. The Psychological Reductionist view simply tells us under what conditions person x is identical with person y. It tells us when the person who bears the burden is identical with the person who receives the counterbalancing benefit. The mere fact that we have changed our view about these identity conditions does not justify us in revising our moral claims. Parfit owes us a story about how the importance of compensation presupposes that identity is a 'deep' fact. After all, we might just think, upon accepting the Psychological view, that since compensation presupposes personal identity, and identity is not a 'deep' fact as we had thought it to be, then compensation does not presuppose any 'deep' fact.

Recall that one of Parfit's reasons for claiming that identity is not a 'deep' fact on Psychological Reductionism involves the fission case. In the fission case you are psychologically continuous with both of the fission products A and B; therefore, you are not identical with either A or B. Now suppose that, before your fission I hit you with my car and put you in the hospital for several months. Your fission occurs before I pay you any money for hospital expenses, time lost, etc. Parfit
claims that in the fission case I cannot compensate you by providing those benefits to A, because A is not you - you receive no counterbalancing benefit if I give the money to A. But then, according to Parfit, it is not plausible to suppose that benefits to A can compensate you for your burden, even when B does not exist. How could the presence of B make a difference as to whether you are compensated by benefits to A? If benefits to A in the fission case do not compensate you, then neither do benefits to A in the case where A is identical with you. Therefore, Parfit concludes, "[w]e can defensibly hold that a benefit at one time cannot provide compensation for a burden at another time, even when both come within the same life" (343).

We need to alter Parfit's argument a bit. If you care about A in the fission case and desire that A be benefitted, then I can provide you with an indirect benefit when I directly benefit your fission product. So I could compensate you by providing A with benefits. I cannot, however, directly or automatically compensate you by benefitting A. Then Parfit should conclude that, similarly, I cannot automatically compensate you by benefitting A in the case where A is identical with you; automatic intrapersonal compensation is not possible on the Psychological view. Of course, we may think that if there is not automatic intrapersonal compensation, then, at the very least, compensation is not as significant as we had thought, if not impossible, on Psychological Reductionism.
I do not think that this conclusion supports full person-neutrality. It seems right to say that, on Psychological Reductionism, because psychological continuity, not identity, is what matters, it should not matter whether you or your fission product receives a certain benefit; therefore, we need to focus on some notion which, unlike compensation, does not ‘presuppose personal identity’. So let us define a new term, ‘quasi-compensation’.³¹ Let us say that I can quasi-compensate you for a burden that I impose on you if I provide a larger benefit to someone psychologically continuous with you. For example, I can quasi-compensate you by providing benefits to A when A is you or to A or B in the fission case. As I pointed out in section 15.2, Psychological Reductionism should lead us to view the boundaries between psychologically continuous persons as less ‘fundamental’ than the boundaries between complete strangers. However, the distinction between you and the governor of Idaho remains significant. Shifting emphasis from compensation to quasi-compensation reflects the change in view about morally important distinctions. So we should change Morality of Sacrifice to:

_Morality of Sacrifice III:_ There are certain burdens such that if you impose one of those burdens on me, then you ought to quasi-compensate me.

According to Morality of Sacrifice III, we can balance benefits and burdens between psychologically continuous persons, but still not between strangers.

³¹I borrow this term from Parfit (see “Comments”), but I define it somewhat differently and use it for a different purpose. Cf. Shoemaker on quasi-memory in “Persons and Their Pasts”.
But we have to remember that psychological continuity can be a matter of degree, so we should think that a person can be quasi-compensated to varying degrees. For example, you will be fully quasi-compensated by benefits to someone fully psychologically continuous with you such as your fission product or yourself, while you will be less, or partially quasi-compensated by benefits to a friend or a family member. Quasi-compensation, then, reflects what the morally relevant boundaries are, that is, the boundaries drawn by psychological connections between persons. So I think that Parfit is right to think that we should not worry about compensation, but why should we not worry about quasi-compensation? If we do, we cannot be fully person-neutral as is the Utilitarian. Rather, in balancing benefits and burdens we have to pay attention to the relations between persons.

So let us consider what sort of moral theory is supported by Psychological Reductionism. Consider again the case where I hit you with my car. My obligation is to quasi-compensate you for your burden. But, suppose that you die as a result of the accident. I can still quasi-compensate you by providing counterbalancing benefits to your children or your spouse, although I will thereby quasi-compensate you to a lesser degree than if I had been able to provide counterbalancing benefits to you. Notice that my quasi-compensating you through your family does not entail that the benefits that I provide to your family are benefits to you, not even that they are indirect benefits to you. By shifting focus from compensation to quasi-compensation, we are saying that what is morally significant is not that you
receive a benefit that outweighs your burden but that someone psychologically continuous with you to some degree receive such a benefit whether or not you even know about the benefit. Of course, by providing a benefit to your spouse to whom you are very closely connected I quasi-compensate you to a greater degree than if I benefitted your colleague to whom you are less closely connected.

The importance of quasi-compensation also supports the commonsense view about special obligations that I described in Chapter 4. Given that I am quasi-compensated by benefits to my mother, I am obligated to bear burdens whether it is she or I who will receive the benefit. Of course, given that I will be fully quasi-compensated only by benefits to myself, my strongest obligations will be to myself (see Chapter 4 for a discussion of the difficulties created by this conclusion). So I have no complaint if someone gives my mother a certain benefit rather than giving me a somewhat smaller benefit; in either case I am quasi-compensated to about the same degree. But I am not obligated to view a benefit to a stranger in anything like the way that I view the same benefit to myself. So my special obligations to my intimates arise from the special relationships in which they stand to me. They are continuous with me to varying degrees, so benefits to them quasi-compensate me to varying degrees.

The notion of quasi-compensation is very natural once we accept the significance of quasi-private projects. As long as it was justifiable for me to care that I finish my thesis, your interference with that project demanded that you
provide me with a counterbalancing benefit, something that will help me to further if not that private project, some other of my private projects. But, if psychological continuity is what matters, I should have the quasi-private project that someone psychologically continuous with me finish my thesis. Then, if you interfere with that project, you should provide a counterbalancing benefit to someone psychologically continuous with me. (Again, one might think that the benefit should go to someone who is connected to me in some way relevant to that particular project. But I think that I should view my intimates’ quasi-private projects in a way similar to the way in which I view my own. Therefore, as long as my intimate can further her quasi-private project, I am quasi-compensated.)

In describing the moral theory that is supported by Psychological Reductionism, I have said that one can be quasi-compensated to various degrees depending upon the degree to which one is continuous with the person who receives the counterbalancing benefit. Degree of continuity between you and some other person is determined by the strength and number of connections that compose the relationship between you and that person. As we have seen, in certain cases degree of continuity may depend upon the nature of the connections. We might wonder, however, why we should not concentrate simply on continuity rather than on the connections that constitute the relations of continuity. After all, as I argued in Chapter 3, it seems plausible to suppose that continuity alone is what matters, and in section 16.1, I argued that full intrapersonal compensation need
not be problematic as connections weaken across time within a life. I think that we have to remember that what matters can decrease, even in the intrapersonal case, if the causal connections are not of the appropriate sort, for example, if they are the result of an external imposition. In Chapter 4 we noticed that there is an interpersonal parallel of external impositions. In cases where causal interaction fails to be the result of and in accordance with the intentions and plans of both parties to the interaction, then we do not have grounds for special concern. These sorts of cases show that it is important to look at the nature of the connections. For purposes of quasi-compensation, then, degree of continuity is perhaps determined more by the nature of the connections than merely by their number and strength. The mere fact that someone is continuous with me is not sufficient to determine how I should view benefits to that person. I must know whether the connections involved in the continuity are, for example, primarily connections of similarity or whether appropriate kinds of causal connections are also present.

Parfit could reply to my objections that there is a Utilitarian justification for considering in what relationship I stand to those affected by my actions, that there “are good Utilitarian reasons for agents to possess and act from a differential concern for their own welfare and the welfare of those close to them” (Brink 1989 289). As Sidgwick points out (432ff.), each person benefitting himself and his intimates will tend to be the best way for him to contribute to the overall greatest happiness. For example, he himself will get greater pleasure from benefitting a
loved one than from benefitting a complete stranger. I have greater knowledge of how best to benefit my loved ones, and I am in a better position to do so because I interact with them on a regular basis. So, on such a view, my connectedness to my intimates justifies my giving their interests greater weight when I decide what action I ought to perform. Similarly, when I can provide a benefit to one person only by imposing a burden on some other person, I must consider the relationship of the two persons involved: it will be easier for the first person to bear the burden if he cares about the person who is going to receive the benefit.

This reply is inadequate. Psychological Reductionism implies that the nature of certain relations, in themselves, justifies a weakening of the significance of certain boundaries. In other words, it offers intrinsic reasons for treating quasi-compensation as the morally important notion. The current Utilitarian reply, however, offers only extrinsic reasons, reasons that do not appeal to the nature of the relations in themselves. The friend of Utilitarianism might say, however, that he can allow that the relations provide intrinsic reasons for treating quasi-compensation as significant. He can agree with the Impersonalist (see section 11) that certain components of relationships such as the friendship relation have intrinsic value and, thus, in and of themselves, provide reasons for regarding certain boundaries between persons as more morally important than others. Again, however, the Utilitarian will have the same sort of problem with which the Impersonalist was faced. Components of friendship will become more benefits to be
weighed in the balance. If more good in the form of intrinsically good relationships can be promoted by providing counterbalancing benefits to a stranger, then the Utilitarian must say that the stranger ought to receive the benefits even if I am not thereby quasi-compensated. So even if the Utilitarian can agree that the relations provide intrinsic reason for treating quasi-compensation as morally important, Utilitarianism will still too often allow the value of those relationships to outweigh the significance of any particular case of quasi-compensation.

17 Identity and Compensation

Suppose that we were to deny that identity is not what matters. I argued in Chapter 3 that we can continue to view our private projects as justified even though no reductive analysis of identity can fully account for their significance. If we accept my conclusion, then we can continue to view compensation as morally important. Recall that in section 4.1 of this chapter I said that we do not think a person’s life plans should be constantly jeopardized by burdens imposed on that person by others. Given the central role that private projects have in my life, I am owed compensation for any unwarranted interference with my pursuit of those projects. My private projects involve me and their significance does not reduce to the independent significance of psychological and/or physical continuity. Therefore, I must receive a counterbalancing benefit when certain burdens interfere with my private projects.
Of course, fission cases show that the next best thing to my finishing my thesis is someone psychologically continuous with me finishing my thesis. Given the importance of psychological continuity, we should accept the importance of quasi-compensation. Nonetheless, if you quasi-compensate me by providing a counterbalancing benefit to my mother, you do not compensate me. By providing the counterbalancing benefit to me, you would have both quasi-compensated and compensated me. The quasi-compensation in the former case differs only in degree from the quasi-compensation in the latter case. But the quasi-compensation in the former case differs in kind from the compensation in the latter case. If you unable to compensate me for a burden that you imposed on me, and you then provide a counterbalancing benefit to my mother, it is not that you do what you ought to do to a lesser degree; rather, you do the next best thing to what you ought to do. You should have compensated me; if you cannot, then you should at least quasi-compensate me.

If we continue to view compensation as an important notion, then we should view the intrapersonal case as different in kind not in degree from the interpersonal case. We will not have a continuum of similar cases, but, rather, an analogy between the intrapersonal and the interpersonal cases given the presence of psychological connections in both. But, given the importance of compensation, my duties to myself will be different in degree and in kind from my duties to my intimates. This conclusion would seem to heighten the difficulty faced when I have
a choice to benefit myself or my child; now my duty to myself is not only stronger
but of a different order. Again, this is a somewhat unattractive conclusion that I
do not know how to avoid.
Works cited


Williams, Bernard. Problems of the Self. New York: Cambridge University