

Embodied Minds:
A Critical Response to McMahan
on Personal Identity

Oritsegbubemi Anthony Oyowe

**This dissertation is submitted in partial fulfilment of the requirements for the degree of
Master of Arts in Philosophy, in the School of Philosophy and Ethics.
Faculty of Humanities, Development and Social Sciences,
University of KwaZulu-Natal, Pietermaritzburg
November 2009**

Declaration of Originality

The research for this dissertation was carried out through the School of Philosophy and Ethics, University of KwaZulu-Natal, Pietermaritzburg from February to November 2009.

The author hereby declares that the content of this dissertation, unless specifically indicated to the contrary, is his work, and that the thesis has not been submitted simultaneously, or in any form for any degree or diploma to any other University. Where use has been made of the work of others it is duly acknowledged in the text.

Date.....

Sign.....

CONTENT

Declaration of Originality	i
Content	ii
Introduction	1
Chapter One	
1. The Problem of the Isolated Subject	
1.1. Introduction.....	7
1.2. A Dividing Theory.....	7
1.3. Pre-persons, Persons and Post-persons.....	8
1.4. An Awkward Implication.....	10
1.5. In defence of the Psychological View.....	12
1.6. A Further Complication: The Isolated Subject.....	15
1.7. conclusion.....	20
Chapter Two	
2. Disarming the Self and the Future	
2.1. Introduction.....	21
2.2. Thought Experiment One: Body Swap.....	21
2.3. Thought Experiment Two: Step by Step Progression.....	23
2.4. The differences in description of William’s thought experiments.....	24
2.5. What do the thought experiments really establish?.....	27
2.6. Conclusion.....	31
Chapter Three	
3. Surviving without a Brain	
3.1. Introduction.....	32
3.2. Three Types of Continuities.....	32
3.3. Why Physical continuity is unimportant	35
3.4. McMahan’s Rejoinder.....	37
3.5. Which Intuition is Stronger?.....	41
3.6. Conclusion.....	45
Chapter Four	
4. Courting the Enemy: McMahan on the Unity of the Mind	
4.1. Introduction.....	46
4.2. Why the Standard View is Flawed.....	46
4.3. An Alternative Account.....	49
4.4. A Tale of Two Minds?.....	50
4.5. Some Notes on McMahan’s Solution.....	53
4.6. Conclusion.....	58
Conclusion	59
Bibliography	60

Introduction

At some point, many of us will have to contend with questions regarding what sort of beings we essentially are as well as the necessary and sufficient condition(s) of our survival. Suppose after experiencing an ordeal, which caused severe injury to your brain, someone wakes up in your place without all of your memories, beliefs, desires as well as your distinctive personality traits. Is this someone you? Your answer to the question is in many ways telling. It will reveal not only your belief about what you essentially are but also the necessary and sufficient conditions of your continued existence over time. In other words, your answer to the question “what are the conditions for my continued existence” depends on your answer to the question “what am I?” As might be expected, philosophers have speculated widely about your essence as well as the necessary and sufficient conditions of your survival. A significant number of them believe that you are essentially a simple, immaterial soul and that your continued existence is a matter of the persistence of this soul. As a result, you survive the ordeal, which presumably does not destroy your soul.¹

Some claim that the resulting person following the devastating psychological ordeal is you chiefly because they believe that you are essentially a human organism and that you continue to exist as long as this biological organism exists. Notable proponents of the biological criterion of identity include Olson (1997) and De Grazia (2003, 2005). Others deny that you survive the ordeal since they take identity to be a matter of psychological continuity through time. The disruption in your psychology is just too radical to preserve what matters for your survival. Some notable defenders of the psychological criterion of identity are Locke² (1694), Parfit (1984) and Shoemaker (1970). Still there are others who believe that identity cannot be analysed in terms of physical or psychological continuity. This sort of view is defended in Madell (1981) and Swinburne (1984). They believe that you will survive the ordeal but not because there will be physical or psychological continuity between you and the resulting person, but because identity is a further fact—the simple fact that the resulting person is *you*.

Other philosophers not content with these metaphysical issues about identity have made the truth-seeking trip all the way attempting to answer specific practical questions. They wonder about why it is morally wrong to kill beings such as ourselves as well as the moral

¹ This sort of view is found in Plato and more recently Descartes. McMahan (2002) argues extensively against the view that we are souls whether of the Platonic or Cartesian type.

² Although Locke appealed only to memory connections, I have included him as one of the proponents of the psychological criterion mainly because I believe his view comes as close as anything to the psychological criterion of identity.

justification for or against abortion and euthanasia. Why are you justified in having special prudential concern for some future entity and not others? As Shoemaker observes, the significance of these normative questions is based on the insight that our answers to them depend on our metaphysical views on what we essentially are and what preserves our identity over time (2009:11). So, we return once again to the metaphysical questions by attempting to answer the practical ones.

These and other related issues are on the borderline between metaphysics (which deals with the fundamental nature of reality) and ethics (which deals with what ought to be the case and why). Although personal identity and ethics are two independent fields of study, philosophical discussions on the relation between them provide the appropriate context for dealing with such questions. This is so because such questions, which are increasingly being termed *marginal*, as they arise at the beginning and ending of our existence, bring together metaphysical issues in personal identity and ethical issues that affect our moral choices and the ways we ought to live our lives.

In his book, *Ethics of Killing: Problems at the Margins of Life* (2002), McMahan makes an important contribution to the ongoing debate on the relation between personal identity and ethics. His entire discussion is set in the context of standard accounts of personal identity and their relation to applied ethical issues. After presenting a thorough investigation into the various accounts of personal identity and arguing in detail against each one, McMahan puts forward a new account, the view that we are essentially embodied minds and that our identity over time involves the physical and minimal functional continuity of the same brain (2002: 68). On this view, I am identical to some future person if and only if that person has enough of those regions of my brain capable of generating consciousness. He also develops an account of what grounds egoistic concern for the future alongside his account of identity, having accepted Parfit's (1984) arguments that identity and egoistic concern can come apart. In what follows, I briefly spell out the core aspects of the embodied mind view. My aim is to provide some background to the issues tackled in the chapters of this work.

As a preliminary, it is important to emphasize that the embodied mind view, as McMahan presents it, divides into two theories—the Embodied Mind Account of Identity and the Embodied Mind Account of Egoistic Concern. This is a result of Parfit's showing that identity and what matters in egoistic concern for the future may come apart. It is exemplified in his

My Division story in which what matters (i.e. psychological continuity) is present in one's relation to two distinct persons, following the successful transplant of one's hemispheres to two decerebrate organisms (Parfit 1984: 254–255). Clearly, none of the resulting persons is identical to oneself as identity is a transitive relation and there are two of them³.

Very broadly, the embodied mind view stresses the continuity of the brain as necessary and sufficient for identity and egoistic concern. In order to clarify this, McMahan distinguishes three forms of continuity of the brain. First, he claims that there is physical continuity of the brain when the same constituent matter of the brain continues to exist or when the constituent matter is gradually and incrementally replaced just in the same way the cells of the other organs of the body are. Second, he claims that there is functional continuity just when the basic psychological capacities are continually had by the brain (McMahan 2002: 68). These include the capacities for consciousness and thought. The embodied mind view stresses that the continuity of the brain in either of these ways is crucial for survival and as a basis for egoistic concern. It maintains that physical and functional continuity (of those areas of the brain in which consciousness is sustained) is both necessary and sufficient for identity and is the basis for rational egoistic concern for the future. Elsewhere he adds that the degree of rational egoistic concern for the future may vary with the degree of psychological unity⁴ in a person's life. (McMahan 2002:78).

All of these relegate the third type of continuity of the brain—organizational continuity—to some level of unimportance. He claims that there is organisational continuity if the various tissues of the brain that underlie connections among mental states are preserved. Psychological theorists also appeal to connections among the various mental states but that sort of continuity is different from McMahan's since he requires that the various states are sustained in the same physical realizer—the brain. In any case, he doesn't judge them to be crucial for survival. With regards to survival, McMahan is clear that we survive as long as our brains continue to exist and retain the *capacity* for consciousness. It is important, he thinks, to emphasize the *capacity* for consciousness rather than *continuous* consciousness “so that the renewed appearance of conscious states following a period of unconsciousness is always the reappearance of the same consciousness, or the same mind” (2002: 67).

³ I shall provide details later as regards this thought experiment.

⁴ He claims that psychological unity “is a function of the proportion of mental life sustained..., the richness of that mental life, and the degree of internal reference among the various earlier and later mental states” (2002: 75). But he clearly does not imply that connections among mental states are necessary for identity and egoistic concern (:79).

Another important feature of McMahan's embodied mind view that is worth noting is that it takes the notion of same consciousness to be equivalent with the notion of same mind. For the most part, McMahan uses them interchangeably. To see the importance of the notion of a mind on this account, it is useful to include here a common distinction in the literature between two classificatory concepts: *phase sortals* and *substance sortals*. The former involves those concepts under which an entity may be classified only for some time in its history. That entity may cease to fit into that category yet continue to exist: it is only a phase in its history. For instance, being a child is a phase in the history of a person, who may continue to exist after ceasing to be a child. The latter includes those concepts under which an entity may be classified for its entire history: that entity ceases to exist once it ceases to be that thing. For instance, a dog ceases to exist once it ceases to be an animal. On McMahan's view of personal identity, mind is a *substance sortal*. We begin and cease to exist as minds.

As McMahan notes, "because they necessarily apply to individuals throughout their entire histories, substance sortals specify necessary conditions for the identities of those individuals" (2002: 6). Accordingly, the necessary condition for identity over time is the continuity of the same mind or consciousness. He further requires that the continuity of the same mind is grounded in the physical continuity of the brain (that is, those regions of the brain in which consciousness is generated and sustained) so that on this account a mind is individuated by the same brain.

Without a doubt, McMahan's work has taken centre stage for its seemingly thorough arguments against standard accounts of personal identity. On the surface, his proposed embodied mind view seems attractive and provides the philosophical underpinnings for justifying early abortion and other issues of practical concern. As such, it is an ideal point of focus for philosophical analysis. As I earlier noted, he proposes his embodied mind view as a better alternative to other accounts of personal identity, which he examines and rejects.

The research follows the methodological approach⁵ employed in McMahan's book as it is a critical response to his embodied mind theory. More specifically, the approach involves assessing a purely metaphysical account of personal identity—the embodied mind account,

⁵ In the literature, opinions are divided on what the right sort of method should be. Here, the methodological assumption is that personal identity is prior to ethics, which McMahan's discussion of the relevant issues favour. Shoemaker (2009) presents a detailed discussion of the conflicting methods.

which McMahan then applied to relevant other-practical concerns like abortion and advanced directives. In this methodological approach, “metaphysics seems to be prior to ethics, so that the relation runs in one direction, from personal identity to ethics” (Shoemaker 2009: 278). Although McMahan’s other aim is to assess the performance of the embodied mind view on a wide range of applied ethical issues that arise at the margins of life, that undertaking is beyond the immediate scope of this research. The main thrust of the research is both to respond to the criticisms McMahan puts forward against the standard psychological view of personal identity and to assess the plausibility of his preferred embodied mind view. The research shall highlight the significance of the embodied mind account, while showing that that view is faced with a number of metaphysical problems.

The chapters in this research project have been systematically arranged. Each one begins with a brief summary of the specific issue it deals with. These issues pertain to McMahan’s criticisms of the psychological view as well as his defence of his embodied mind view. In the first chapter, I closely examine his objections against the psychological view of identity and argue that against the success of his case. In the next chapter, I deviate a little to discuss issues that do not directly feature in McMahan’s book but are fundamental to the view he ultimately defends. In particular, this concerns McMahan’s reference to the intuition that one could rationally fear future pain in the absence of psychological connections from day to day—of the sort psychological theorists appeal to. The intuition seems to recommend, against the psychological view, that physical continuity is crucial to our survival. He refers to Williams’ defence of that view in *The Self and the Future* (1970) and he believes that it offers some support to his embodied mind view. I argue that the thought experiment which ostensibly supports that intuition is unconvincing. In so doing, I intend to undermine any form of support the embodied mind view may glean from it.

The third chapter takes the issue of the importance of physical continuity to our survival even further. It responds to McMahan’s own positive and negative cases in support of it arguing that he fails to establish his case in support of the importance of physical continuity and against the importance of psychological continuity to our survival. In the final chapter, I attend to an interesting issue regarding the unity of the mind. This chapter deals with certain phenomena that challenge the notion of a mind as we intuitively understand it. As this poses a unique challenge to the embodied mind view, the chapter basically assesses McMahan’s solution to the problem. It suggests that his solution to the problem is fraught with ambiguities

and must appeal to connections among the various psychological states, which he previously denies any real importance in these matters. As such, I claim that it is not plausible.

As a final point, some of the issues discussed in the chapters will overlap. One will notice that the various aspects of the different views of identity considered would resurface a significant number of times in the each of the chapters. This is largely due to the fact that each chapter in some sense can stand alone as they each deal with a specific theme. As a result, it becomes necessary to reiterate certain points. However, the research avoids unnecessary repetitions as much as possible. One aspect that keeps recurring, though, concerns thought experiments which are often used to elicit our strongly held beliefs regarding our survival. Unlike other philosophers who remain suspicious of that methodological approach⁶, McMahan is quite happy in his use of them to bring to the awareness of his readers what he thinks corresponds to our strongly held beliefs about survival and egoistic concern for the future. The research project accepts the methodology and discusses a number of these thought experiments—I ultimately argue in favour of those experiments that suggest, against McMahan's own view, that physical continuity is not crucial to survival.

⁶ In the literature on personal identity, Mark Johnston (1987), White (1989) and Wiggins (1980) are among philosophers who are suspicious of the use of thought experiments as a reliable methodological tool.

1. The Problem of the Isolated Subject

1.1. Introduction

The psychological view of identity has been the target of a number of criticisms and some important recent ones have been advanced by McMahan in his *Ethics of Killing: Problems at the Margins of Life* (2002). His main idea is that psychological continuity and/or connectedness comes to nothing in the face of the intuition that one could rationally fear future pain in the absence of psychological connections between oneself now and one's future self.⁷ The charge culminates in what he sees as the problem of the isolated subject. My aim in this chapter is to closely examine his assessment of the psychological view and argue that that view may have more appeal than he is willing to concede. In reaching that conclusion, I shall examine McMahan's other claims that the psychological view is flawed as it implies the absurdity that our existence is preceded and succeeded by the existence of pre-persons and post-persons respectively. I claim that the force of this charge turns on an erroneous reading of Parfit's notion of psychological continuity and that its triviality follows from a reductionism about persons. In the end, I shall be a little more ambitious in claiming that the isolated subject charge fails: I plan to analyze and reject two ways of understanding the isolated subject. I conclude that this aspect of McMahan's case provides no reason for adopting the embodied mind view.

1.2. A Dividing Theory

The Psychological view is the last of the identity theories to come under McMahan's (2002) thoroughgoing analysis having rejected the views that we are essentially souls and organisms respectively. On that view, we are essentially psychological beings and the criterion of our identity over time is psychological continuity and/or connectedness. Parfit takes the relation of continuity to mean "the holding of overlapping chains of *strong* connectedness" (Parfit 1984: 206). Given that what counts as strong connectedness is a vague matter, Parfit tells us that psychological connections are *enough* to constitute identity if there are, "over any day, *at least half* the number of direct connections that hold, over everyday, in the lives of nearly every actual person" (Parfit 1984: 206). Psychological continuity is necessary for identity. But Parfit's *My Division*⁸ is held to show that it is not sufficient for identity as it could take a

⁷ Williams' arguments in "Self and Future" (1970) offer strong support for that intuition. I believe there's good reason to think that his description of the thought experiment, which ostensibly supports that intuition, is put forward in question-begging terms. I advance that case in a subsequent chapter.

⁸ I consider this thought experiment in some detail in a subsequent chapter.

branching form—i.e. two people could be psychologically continuous with one person if one's hemispheres were successfully transplanted into two decerebrate organisms. If I divide in this way, neither of the resulting persons is identical to me as identity is a transitive relation and there are two of them. But there is good reason to expect that I would care in an egoistic way about what happens to the resulting persons since all that is normally present in ordinary survival is present in my relation to the resulting persons.

Parfit's hypothetical division shows that identity and what matters in egoistic concern⁹ can come apart as what matters (viz. psychological continuity) may be present even in the absence of identity. As a result, McMahan suggests that the psychological view divides into two accounts—of identity and of egoistic concern. The former holds that psychological continuity in non-branching form is the basis of identity while the latter is the view that psychological continuity and/or connectedness irrespective of the form it takes is what grounds egoistic concern (2002: 42). It is useful to discuss separately the problems he envisages for each account beginning with the theory of identity.

1.3. Pre-persons, Persons and Post-persons

As a prelude to the pre- and post- person objection, we note that the criterion of our identity depends on what we essentially are. The psychological view assumes that we are essentially persons—beings with highly developed capacities for self-consciousness as this is necessary for important psychological connections to hold from day to day. In Parfit's words, "a being must be self-conscious, aware of its identity and its continued existence over time" to be a person (1984: 202). The view that we are essentially persons has interesting implications for when we begin and cease to exist. It implies that we began to exist sometime in early childhood, when mental capacities have been highly developed to ensure that enough psychological connections hold from day to day in line with Parfit's stipulation. McMahan points out that the view implies that we, who are now persons, were not once infants and so it must provide some account of the conscious entity that existed during the later stages of pregnancy and just after the birth of one's organism:

There was, in short, a conscious being present both immediately before and after the birth of one's organism—a being whose mental life was in some ways continuous with one's own... But the assumption that this

⁹ The term is taken to mean a special *kind* of concern for the future (McMahan 2002:41 & Schechtman 1996:52)

being was actually oneself is, of course, mistaken if the Psychological Account is right... Who, or what, was this conscious being? (2002: 46).

Similar considerations apply to end of life cases. In the case of Alzheimer's disease, in which one's mental life gradually deteriorates, psychological connections from day to day will eventually fall below Parfit's threshold. The psychological view of identity implies that one would have ceased to exist in the later stages of Alzheimer's, though there will be conscious activity and weak psychological connections from day to day. McMahan writes, "[w]hen psychological continuity comes to an end, the person ceases to exist, according to the Psychological Account of Identity" (2002: 47). Again, he thinks the question of who or what is the conscious entity with weak psychological connections, that exists in association with one's organism, upon the death of oneself, is one of interest. His proposal is that the psychological view is committed to the existence of a pre-person that precedes the existence of the person as well as a post-person that begins to exist upon the death of the person.

It seems that psychological theorists must claim that the conscious life in association with one's organism before and after the existence of the person does not belong to oneself but to two other entities. This seems one too many. The crux of the charge then is that that view populates the world with more new entities—pre and post persons. McMahan thinks that this is an extravagant supposition and one that offends against parsimony—the principle that “a theory is less plausible the more new entities it is required to postulate” (2002: 46, 47).

Before discussing McMahan's objection against the psychological account of egoistic concern, it is important to make some general comments about his criticisms of that view. The claim that psychological theorists must posit the existence of pre- and post- persons is chiefly motivated by McMahan's reading of Parfit's claim that the identity of persons over time consists in the “the holding of overlapping chains of *strong* psychological connectedness” (1984: 206). For McMahan, conscious entities that lack *strong* psychological connections from day to day are not persons in Parfit's terms. They must therefore be pre- or post- persons depending on whether they occur at the beginning or ending of a person's life. However, it is worth noting that McMahan is less than charitable in his interpretation of Parfit. He reads Parfit as offering a necessary condition of identity over time when in fact he is only suggesting a sufficient condition.

On a more careful reading, it can be seen that Parfit claims that when direct psychological connections between an earlier and a later person are weak, it is indeterminate whether one continues to exist. Indeed in such borderline cases as occur at the beginning and ending of life, it is a vague matter whether one continues to exist. And for a theory that admits of vagueness, this is far from being controversial. But when the connections are enough—i.e. when “over any day, *at least half* the number of direct connections that hold, over everyday, in the lives of nearly every actual person” holds between an earlier and a later person (1984: 206)—then we can plausibly recognise the identity of persons over time. Clearly, Parfit is pointing to a paradigm case rather than a limiting one.

Consequently, when there are enough connections in Parfit’s sense, then we have a case which clearly meets the condition of identity over time. But it seems that what is necessary for identity is that psychological continuity—understood here as some (unspecified) degree of psychological connectedness—holds between two persons. McMahan seems to confuse sufficient with necessary conditions. Once we see the distinction we also see that the pre- and post- person charge fails: although the condition for Parfit’s paradigm case is not met in the relation between the pre- person and the person, there are weak psychological connections between them that may or may not be enough. Therefore, what can be asserted with confidence is not that the pre-person is a conscious subject whose existence precedes the existence of the person (as McMahan claims); but that it is indeterminate whether the pre-person is identical to the person (similar considerations apply to the post-person as well). I shall return to the issue of indeterminacy when I attempt a further defence of the psychological view in a later section.

1.4. An Awkward Implication

The problem McMahan identifies for the psychological account of egoistic concern remains to be considered. While the theory of identity, which forms one divide of the psychological view, is held to imply the existence of pre and post persons distinguished from oneself, the theory of egoistic concern yields another awkward result. It is that one has reason to care egoistically about what happens to conscious entities that are not oneself. In particular, the theory implies that one can be egoistically concerned about what happens to the post-person that exists in association with one’s organism during the later stages of Alzheimer’s. Similarly, if we can contemplate a backward-looking egoistic concern, one may have reason to care in an egoistic way about what happened to some pre-person, even though the

psychological account of identity implies that that is not oneself. This is possible since identity and what matters can come apart—in this case, psychological continuity, which constitutes the relation of identity, ceases to hold while connectedness, which is one of the relations that ground egoistic concern may continue to hold albeit to a lesser degree. Parfit himself claims, “it can be rational to care when one of the grounds for caring will hold to a lesser degree” (1984: 313). In the case of Alzheimer’s, there is reason to care egoistically as long as there are weak psychological connections from day to day, though there is no basis for identity—i.e. psychological continuity defined in terms of strong psychological connectedness.¹⁰

I think it is quite startling that McMahan finds this outcome awkward (2002: 48). The reason he finds it awkward is that it implies that I can have egoistic concern for some entity, a post-person say, which is not me. Although he admits that it is open to proponents of the psychological view to argue that identity is not what matters in actual cases as well as in the hypothetical division scenario, he maintains his preference for a theory that minimizes the divergence between identity and egoistic concern on the grounds that it is not accidental that “we have hitherto assumed that personal identity is what matters and thus the scope of rational egoistic concern coincides with persona identity” (: 52). As he rightly acknowledged, we are not rationally compelled to take that proposal seriously if we are reductionists about persons. This is because there is “no deep fact about identity... that compels us to accept an account that minimizes the divergence between identity and egoistic concern” (: 51).

Even so, I would like to emphasise what I think is a source of major worry. In particular, it bothers me that while he accepts that identity is not what matters as shown in *My Division*, he is wary of seeing the full weight of that claim for the psychological view. If we accept the view that identity is not what matters, we should also accept a divergence between our theory of identity and theory of egoistic concern. So why does McMahan find it awkward that one could care egoistically about an inheritor of one’s psychology even though that is not oneself? Granted, one would have ceased to exist by then according to the psychological view but this is consistent with the view that identity and what matters in egoistic concern can be prised apart. It seems to me that his unwillingness to accept this implication reflects his doubts over

¹⁰ In order to eliminate this divergence between identity and egoistic concern, McMahan proposes an elaborate revision of the psychological theory of identity so that it is akin to the theory of egoistic concern. In it, he replaces Parfit’s psychological continuity (i.e. strong psychological connections up to Parfit’s required threshold) with the notion of broad psychological continuity. The latter involves psychological connectedness to any degree of strength (2002: 49 -51).

the identity does not matter view when applied to actual cases in which psychological continuity is non-branching. Or perhaps, he does not accept that identity is not what matters despite what he says. Thus, he proposes some kind of congruence between our account of identity and what matters. However while he is wary of the results the identity does not matter view yields for the psychological view, he sees no problems when he employs the same view to his advantage.

For instance, it is crucial in his formulation of what he calls the Time-Relative Interest Account¹¹ (TRIA) (2002: 194), according to which the badness of death and/or the wrongness of killing is proportional to the time-relative interest of the victim. TRIA is based on the core idea that identity is not what matters. In accounting for the badness, TRIA focuses on the effect death has on the victim's time-relative interests and not on the value of his life as a whole. In his view, TRIA rivals the standard account—the Life Comparative Account—which evaluates the badness of death in terms of the effect it has on the value of the victim's life as a whole (: 105). It is crucial to note that in considering the life of an individual as a whole, LCA presupposes that identity is what matters. On the other hand, TRIA seem to presuppose that identity is not what matters by focussing on the individual's interest as a particular time.

My aim here is not to query the plausibility of these accounts but to suggest that while McMahan is happy to use the identity does not matter view to his advantage, he denies its implication for the psychological view. I believe that once we have fully endorsed the fission story, we should be committed to its implications. That he finds an implication of the identity does not matter view for the psychological account awkward is inconsistent to say the least.

1.5. In defence of the Psychological View

The preceding objections ultimately target a Parfitian view of persons and personal identity. Parfit labels that view a reductionist view, according to which facts about persons and personal identity consist in the holding of more particular facts concerning brains, bodies, and series of interrelated mental and physical events (Parfit 1984: 210-211). What follows if we are reductionists about persons? An important result though not an inevitable one is that in certain cases our identity would be indeterminate. That is, there could be certain cases in

¹¹ The Time-Relative Interest Account replaces what he terms the Life Comparative Account (of the badness of death and the wrongness of killing), according to which identity is what matters.

which there is no fact of the matter whether a person still exists or not. McMahan himself is quite at ease with this result and, in fact, claims that “[t]he theory’s vagueness about when psychological connectedness ceases to hold, and thus indeterminacy about when we begin or cease to exist, should be regarded as a virtue rather than a failing” (2002: 44). The reductionist likens vagueness about persons and personal identity to similar instances of vagueness exhibited by concepts like that of club. Parfit sets out the analogy in this way:

Consider, for example, clubs. Suppose that a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and the same rules. We ask: ‘Have these people reconvened the *very same* club? Or have they merely started up *another* club, which is exactly similar?’ (1984: 212)¹²

Although there is a sense in which one can trace the origin of the *new* club to the *old* one, our answer to the question will be arbitrary. Parfit’s own stance is that there would be no true or false answer to the question¹³, but we can know all the facts. The reason is that we are reductionists about clubs, and take them to be fundamentally the existence of members, behaving in certain ways according to certain rules. If we know these facts, we know all there is to know. When applied to persons, we see that it is indeterminate whether the conscious entity that exists at the late stages of fetal development and soon after the birth of one’s organism is oneself—in just the same way it is indeterminate whether an earlier club, which declined in membership, meetings and reconvenes many years later, is the same club. In such borderline cases, Parfit’s proposal is that though there is indeterminacy we can decide to provide an answer (: 214). We *could* decide to call the conscious entity in the latter stages of Alzheimer’s a post-person or the same person as the entity in the early stages of the disease.

The distinction between persons, pre and post persons does not reflect a metaphysical fact but simply our use of the term “person”, which could have been different. It lies in the different ways we group together more basic facts concerning brains and interrelated mental events. So, in borderline cases, it is indeterminate whether some conscious entity is a person as what is sufficient for personhood is strong connectedness. But the person’s existence can be tracked back to that conscious entity, who perhaps is a pre-person as its psychological connections from day to day are significantly fewer than what is normally sufficient for personhood.

¹² Emphasis as in the original.

¹³ Here, Parfit assumes that there are no rules or legal facts that support a ‘yes’ or ‘no’ answer to the question (1984: 214), thus implying that some kind of stipulation is warranted.

If we are reductionists about persons, we can contemplate these possibilities and there is no deep reason to prefer an account that denies the existence of pre- and post- persons over one that does not. In line with Parfit, “this is not a decision between different views... we are merely choosing one of two different descriptions of the very same course of events” (1984: 214). So, while it can be indeterminate whether a pre-person or post-person is the same as oneself, we know the important facts namely, the degree of physical and/or psychological connectedness.

Although McMahan claims to be a reductionist and admits that identity could be indeterminate in certain cases, he has trouble accepting the reductionist reply to the pre and post person charge. He is specifically dissatisfied with the result that persons are analogous to clubs or nations and thus no more ontologically basic than these entities. He claims that it is “difficult to accept this radical reductionism as applied to ourselves” as though the mere statement of a claim closes an argument. Parfit himself admits this difficulty and so attempts to show why what we “naturally” believe about ourselves is mistaken. Even so, it is not clear that the difficulty by itself is enough to undermine the reductionist response.

McMahan’s suspicion of the reductionist response to the pre- and post- person charge rests on his reading it as implying that we are no more ontologically basic than clubs. But neither Parfit nor reductionism about persons implies that we are no more ontologically basic than clubs. Parfit at no point commits himself to the view that persons and clubs belong to the same ontological category; his reductionist view only implies that our existence consists in more particular facts – brains, bodies and a series of interrelated mental events. Indeed, clubs and persons belong to different ontological categories and we begin to see this once we recognise that these are different kinds of things with different persistence conditions¹⁴. Moreover, while clubs are a product of our activities, persons are not.

The analogy between clubs and persons does not imply that persons are literally clubs. On a more charitable interpretation, it is Parfit’s way of saying that the concepts exhibit vagueness in importantly similar ways and that we *could* deal with issues of indeterminacy with regards to persons in the same way we do clubs. In the final analysis, this is what reductionism about

¹⁴ One possible way of differentiating between clubs and persons is to think of them as occupying different levels in an ontological hierarchy. While both will not be part of the most fundamental elements of our ontology, persons will be more fundamental than clubs in terms of that ordering.

persons comes to. So, if there is an issue about the ontological status of persons, it should be that persons are no more ontologically basic than the facts which constitute them. But this should not be a serious problem for McMahan who himself defends what seems to be a reductionist account of identity. So, it is important that a reductionist reply does not deny that there is a sense in which one can speak of a pre- or post- person, but this is a virtue of that account which McMahan is not fully prepared to accept. In order for us to take the pre- and post- person charge seriously, he needs to show why it constitutes an anomaly for the psychological view – but he clearly hasn't done this. I do not see why the reductionist should be troubled by his worries about the ontological status of persons or the existence of pre and post persons. However, McMahan anticipates a further problem for that view.

1.6. A Further Complication: The Isolated Subject

Any attempt to rescue the psychological view faces a further problem¹⁵. Since it takes identity to be a matter of continuity and/or connections from day to day, the absence of such connections between a person now and a conscious entity in the final stages of fetal development or in the very late stages of Alzheimer's appears to be potentially damaging to the view. While the account of identity implies that that conscious subject is not oneself, the account of egoistic concern implies that one has no reason to care egoistically for it. In view of that, the psychological account inexplicably isolates this conscious subject as there are no psychological connections from day to day between oneself now and the conscious subject at the beginning or ending of life, when mental life begins to develop or gradually deteriorates as the case may be.

The significance of the problem is often expressed by appeal to the intuition that one could rationally fear future pain in the absence of psychological connections between oneself now and one's future demented self. Bernard Williams defends that intuition by imagining a scenario in which "someone in whose power I am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to tomorrow in great apprehension" (1970: 167). He also supposes that the proposed torture will be preceded by an operation in which one's psychology is replaced by the duplicated psychology of an actual person.

¹⁵ McMahan himself admits that his revisions of the psychological view of egoistic concern such that the connections are *real*—i.e. realised in *this* physical brain, will deal well with the other difficulties he has raised for the view. In the case of the isolated subject, however, he thinks that the revised version cannot capture our intuition about what matters as there are no connections whatsoever that can ground egoistic concern between oneself at the early stages of Alzheimer's and a conscious subject at the later stages.

The force of the thought experiment turns on what one's reaction would be when faced with the scenario. Williams' own response is that 'fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen – torture, which one expects to happen to oneself, and to be preceded by certain mental derangements as well' (1970: 168). Clearly, in the scenario described there is no psychological continuity between the earlier self and the deprogrammed self. But if fear (of torture to *oneself*) is the proper reaction, then there seem to be a basis for identity and egoistic concern in the absence of psychological continuity and/or connectedness. I am not persuaded by Williams' thought experiment, as it is described question-beggingly; I pursue that objection in a subsequent chapter. For now, I assume that the intuition is widely held and that there is a subject which the psychological view ostensibly isolates. My aim ultimately is to suggest that we should not be persuaded by McMahan's charge; more specifically, I deny that one can survive as an isolated subject.

McMahan's discussion of the problem of the isolated subject centres on general considerations on progressive dementia—in particular Alzheimer's, with its characteristic gradual wearing down of a person's mental life. His main idea is that as the physical basis of a person's mental life disintegrates, the basic psychological capacities of that person's brain are lost. He claims that this gradual and destructive process reaches an advanced stage in which there are absolutely no psychological connections between oneself at the early stages of the disease and the conscious subject at the very late stages. The psychological view needs these connections to account for identity and egoistic concern; since they are lacking that view implies that the conscious subject at the late stages of Alzheimer's is not oneself and that at the early stages of the disease one has no reason to care egoistically about what happens to it. Suppose that pain is a feature of the disease at the very late stages, one has no reason to be concerned about the pain of the isolated subject. McMahan believes that there is a fairly robust intuition that what matters is present throughout the progress of the disease—something the psychological view apparently fails to capture.

We can attempt to make sense of the isolated subject. The term is potentially misleading. It might suggest that there is an entity over and above the conscious state(s) and who in some way is the "owner" of conscious state(s) and unifies the various experiences and mental states. If we were non-reductionists, we would entertain that idea—the view that what unifies the various mental states and experiences is that they are had by this subject who is the owner

of these mental states. But this is not what McMahan implies. When a reductionist speaks of persons as subjects, she only means that as our way of talking without implying that there is an entity over and above these various mental states. Consequently, talk about a conscious subject consists in the talk about more particular facts concerning brains, bodies and mental states. In my view, there are two ways we can attempt to make sense of McMahan's isolated subject. I will discuss each in turn.

In setting out his embodied mind view, one of McMahan's important insights is that we are essentially minds—that is, beings with the capacity for consciousness. He believes that our identity over time is a matter of continuous retention of this basic capacity by the brain. It resonates with Unger's (1990) view, which makes a distinction between core and distinctive psychology. The former involves a brain's *capacities* such as capacities for thought, conscious experience etc. while the latter involves the *contents* of one's mental life: memories, intentions, desires, beliefs etc. Unger also argues that our identity consists in the "causal furtherance" of one's "core psychology" (1992a: 132-134)¹⁶. He thinks distinctive psychology is of no significance. It seems then that one way of characterising McMahan's isolated subject is as a being with the capacity for consciousness—a being that retains only the bare minimum of one's core psychology.

I suspect that most people believe that as regards our survival we go with our distinctive psychology. It seems pretty clear that the continuous retention of core psychology by the brain cannot guarantee survival or provide grounds for rational egoistic concern. As such, it is hard to see why the charge that the psychological view isolates a being with one's core psychology is of any significance.

The following case should make the point clearer: I have been diagnosed with a brain defect similar to that suffered by a patient with progressive dementia with the exception that in this case all of my distinctive psychology will be wiped in a short while. But that is enough time for me to have all of my distinctive psychology recorded and stored by means of a brain-state-transfer device that leaves my original body with just my core psychology¹⁷. The device will

¹⁶ Note that the important difference between McMahan's and Unger's view is that while the former insists on the continuity of a particular physical realizer—the brain, the latter requires no particular realizer of consciousness.

¹⁷ The similarity to Unger's (1992b) amnesiac moron (characterized as having one's core psychology) should be obvious. The intuition that survival involves the continuity of our distinctive rather than core psychology is supported by a thought experiment in Beck (2001).

then create out of new matter a body and a brain which now has my distinctive psychology. At the end of the process two bodies emerge: the original Body *O* has my core psychology but retains the brain defect while the replica Body *R* has my distinctive psychology without the defect. In McMahan's terms, Body *O* typifies an isolated subject in the sense that it is not psychologically continuous with oneself at the beginning of the process.

Suppose that prior to the operation I were to choose self-interestedly who between Body *O* and Body *R* should be tortured and who should be rewarded, who should I choose? McMahan's view does not only imply that I survive as Body *O* (which further implies that undergoing the process carries no great weight), it also implies that I choose torture for Body *R*. I think that most people like myself believe that I survive as Body *R* and that the process was a last-ditch attempt to ensure my survival. I think that most people will choose torture for Body *O* and reward for Body *R*. In the case described, one of the resulting bodies has my original core psychology while the other has my original distinctive psychology. But it is highly doubtful that at the beginning of the process anyone will intuitively respond that they will go with their core psychology. In view of that it is hard to see how one can survive as an isolated subject—understood as a being with one's core psychology. Since what matters in survival or concern for the future is not the causal furtherance of my core but the continuity of my distinctive psychology, I should have ceased to exist by the time psychological connections completely cease to hold from day to day. Consequently, it seems that the charge that the psychological view isolates that subject is unconvincing.

The first proposed way of understanding the isolated subject fails to further McMahan's charge against the psychological view. If our intuition is that we go wherever our distinctive psychology goes, then it seems that the belief that we cease to exist when the distinctive features of our mental life are eroded carries more weight. But perhaps the significance of McMahan's objection can be best appreciated if we understand the isolated subject differently. In this second sense, a being may have various mental states and experiences that are discrete, isolated and unconnected to each other: there are no causal connections or any degree of internal reference between earlier and later mental states. In many ways, this being would be similar to McMahan's imagined sentient creature "whose mental life is a stream of consciousness without psychological connections" (2002: 75). Its psychological life is not carried forward in time. At the very late stages of Alzheimer's the experiences it has are disconnected from each other as well as from all of the contents of one's mental life up until

these late stages. It is precisely in this sense that the subject at the late stages of the disease is held to be isolated from the individual at the early stages.

Although I believe that it is the first characterisation that truly captures McMahan's thinking about the isolated subject, I shall suggest that understanding the isolated subject in this second sense also fails to convince. As a starting point, it is highly unlikely that it is a feature of Alzheimer's disease, which provides the basis for McMahan's discussion of the isolated subject charge, that there are such momentary psychological states that are completely disconnected from each other. It seems rather incredible to suppose that at some point all psychological connections cease to hold and then at a later time other mental states are generated and that these new mental states would be disconnected from future ones as they would also cease to hold at some point. One must recall that the idea that these mental states are no longer continually had is a corollary of the claim that the disease wears down the very tissues that generate these conscious states. To suppose that certain mental states completely disconnected from previous ones will at some future point be generated in those same regions is to claim that by some miracle the brain's lost capacities are restored.

A significant number of psychological connections hold in the life of a normal adult. But this gradually declines in the life of an Alzheimer's patient due to the gradual wearing away of the brain's capacities. When the disease reaches its advanced stages, it becomes plausible to recognise that not only are there no longer psychological connections among mental states but also that the brain is incapable of generating or sustaining such states. These mental states like pain, beliefs, and desires are ones that are usually carried forward in time. It is not clear how after all of these connections suddenly cease to hold because the disease has destroyed the tissues in which they are sustained that they will resurface completely disconnected from earlier mental states. It is more plausible to hold that while some of the mental states cease to hold at some point other ones continue to hold albeit weakly until all of the brain's capacities are completely destroyed.

On the surface, it appears that McMahan doesn't think that what matters is preserved when there are weak psychological connections between the patient at the beginning of the disease and the conscious entity at the very late stages. He seems to be suggesting that there have to be *strong* connections among the various mental states in order for the patient at the early stages of the disease to have a basis for egoistic concern for what happens to the conscious

being at the very late stages. But as I suggested earlier on, the requirement for enough or *strong* connections in Parfit's sense and as obtains in the life of nearly every actual person only provides a sufficient condition. It is consistent with Parfit's view that there is a basis for egoistic concern when psychological connections hold below what is held to be sufficient.

1.7 Conclusion

The objections McMahan launches against the psychological view are not convincing. On the one hand, the pre- and post- person charge is based on a wrong reading of Parfit's notion of psychological continuity. But more importantly, the significance of the charge falls apart if we are true to our reductionist commitments. On the other hand, as regards the problem of the isolated subject, I have proposed two ways we can make sense of it: either way McMahan's arguments fails to convince. Since he takes it to be an important motivation for the embodied mind view, I claim that that view is not yet motivated.

2. Disarming the Self and the Future

2.1. Introduction

The most remarkable feature of Williams' *The Self and the Future* (1970) is its portrayal of two thought-experiments, the second of which yields a conclusion that is at variance with the first, his revision of Locke's body-swap experiment. He points out the conundrum that arises. The first and second experiments endorse a psychological and physical criterion respectively, yet they are in essence the same¹⁸. His solution is that the neatness of the first description conceals its flaws. Even so, his preference for the second description has been analyzed and rejected by Beck (1998) and Noonan (1989, 2003). I shall argue with them that the second thought experiment, which supports the physical criterion, is flawed as its description of the scenario is question-begging. My plan is to set out both experiments and then show that when the description of the second experiment is made such that the predicted changes are applied in the opposite direction (that is, involving not psychological but physical changes), it supports the psychological view¹⁹. I take this to imply that the structure of Williams' second thought experiment is such that it invariably supports a presupposed outcome. That this is not true of the first description should reinforce my claim.

The experiments in question have been extensively reviewed in the literature on personal identity, with a variety of views emerging. In reaching my conclusion, I shall be citing where necessary philosophers that have made significant contributions to the debate. My aim is to show that McMahan's embodied mind view cannot draw the support it needs from Williams' second experiment.

2.2. Thought Experiment One: Body Swap

In the first thought experiment, we imagine a sophisticated machine that transfers psychologies. Upon entering into it, persons *A* and *B* emerge changed such that *A*'s memories and other distinctive features of her psychology are now united with *B*'s body and vice versa. The proposal would look as follows:

¹⁸ Although Williams claims that the thought experiments are the same, the question of whether they are relevantly similar is itself controversial. I shall return to it later.

¹⁹ Williams' description of the second thought experiment lends support to the physical criterion; I intend to follow his strategy with the exception that I will be assuming that one can survive enormous physical changes; this assumption will be revealed to be true. I contend that the problem with the second thought experiment lies in the way the scenario is described.

Before entry	Machine	After exit
Person A	Takes up <i>B</i> 's body but retains <i>A</i> 's memories	<i>B</i> -body-person
Person B	Takes up <i>A</i> 's body but retains <i>B</i> 's memories	<i>A</i> -body-person

There are four possible descriptions of this outcome each of which would indicate our beliefs about the necessary and sufficient conditions of identity over time.

- a. Person *A* is *A*-body-person.
- b. Person *A* is *B*-body-person.
- c. Person *A* is both *A*-body-person and *B*-body-person.
- d. Person *A* is neither *A*-body-person nor *B*-body-person.

The third option is rejected as an outright impossibility as it violates the principle of the transitivity²⁰ of identity; it implies that *A*-body-person and *B*-body-person are identical, which is clearly false. The fourth option is the claim that person *A* has ceased to exist. On the one hand, this is a denial that a person can survive change. But that view is widely at odds with our shared understanding of the world, reflected in our language. On the other hand, one would have to be a non-reductionist about personal identity in order to opt for d). In this case, the view is that person *A* is neither of the resulting persons simply because identity is not reducible to psychological or physical continuity. However, if we assume²¹ for the moment that identity is reducible to psychological or physical facts, then option d) is temporarily out of the question. Thus, from the perspective of a reductionist options a) and b) are the only live options—though there is enormous support for the view that person *A* is *B*-body person.

The scenario, as Williams describes it, is even more intriguing. We are told that once the process is completed, one of the resultant persons would be rewarded with \$100,000 while the other would be tortured. Persons *A* and *B* are expected to choose on selfish grounds which treatment should be given to the persons that emerge from the machine. Suppose that *A* decides that the pleasant reward should be given to *B*-body-person, we can expect that after the experiment, *B*-body-person would say: “I got exactly what I asked for” since s/he has *A*'s memories and remembers making this choice. Likewise, if *B* opted for the pleasant reward to be given to *A*-body-person, we should expect that after the experiment *A*-body-person should be horrified because s/he remembers asking to be given the pleasant treatment (since s/he has *B*'s memories) but has now being handed the unpleasant one—torture.

²⁰ The principle holds that if *x* is R-related to *y* and *y* is R-related to *z*, then *x* must be R-related to *z* (see Noonan (2003). Here, it implies that *A*-body person and *B*-body person are identical, which is outright false.

²¹ This momentary assumption has to be made here because, I think, the debate between non-reductionism and reductionism on personal identity falls outside the immediate concern of this chapter and, in fact, is not part of Williams's concern here. Even so, the view itself is hard to defend (cf. Parfit 1984).

In both cases, the implication is that persons *A* and *B* have changed bodies; the outcome of the experiment can be best described as body swap; that is, *A* is *B-body person*; a description that refutes any physical criterion. In particular, it actively damages McMahan's view as the thought experiment seems to show that concern about what happens to me in the future is not necessarily a concern about physical continuity but rather psychological continuity with some future person. It seems then that psychological continuity is the criterion of identity over time.

2.3. Thought Experiment Two: Step by Step Progression

McMahan still maintains that physical continuity is crucial to identity. He hopes to find some support from Williams' second thought experiment. In this case, we imagine that 'someone in whose power I [say, person *A*] am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to tomorrow in great apprehension' (: 167). Suppose next that each of the following changes would be produced in *A* prior to the torture:

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

After being informed of these envisaged changes in his psychology, what would be *A*'s (my) reaction to the proposed torture? Williams believes that in each of these instances of change, 'fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen – torture, which one expects to happen to oneself, and to be preceded by certain mental derangements as well' (: 168). He is convinced that none of the steps above represent a change that would justify a change in attitude; the identity of the person being experimented upon is unaffected by each and every change induced. Moreover, he is reasoning on the grounds that one can only fear one's own torture. In this case, then, if *A*'s fear of future torture is justified, the inference is that *A* thinks it will be him who suffers the proposed torture immediately after the sequence of psychological changes have been completed.

We can now tentatively state Williams' conclusions. First, the principle that concern for future physical pain is not excluded by changes in one's psychology – that is to say, concern for future torture is actually concern for this particular body. This principle supports the physical criterion of personal identity. Indeed, he claims that it is possible to conceive that I have being changed psychologically such that I think I am someone else, yet this would not have the effect of reducing my horror when told authoritatively that I would be tortured eventually. I think it is easy to see why fear would be the proper reaction especially in stages (i) – (vi) since it is possible that one's fear of torture can extend beyond the changes described in these stages. Intuitively, most people would admit that these induced psychological changes would not preclude the fear of future torture. What is of interest, however, is what takes place at stage (vi) for it is precisely the scenario described in the first experiment. This brings us to his second conclusion—that on this showing, we are confronted with a puzzle in which two similar thought experiments yield contradicting outcomes.

The conclusions are related; if indeed experiment two warrants its conclusion, then it follows that a rather disturbing conundrum arises. Before discussing what we should make of the puzzle, I will investigate the issue of the relevant differences between the experiments.

2.4. The differences in description of William's thought experiments

I think it is helpful to begin with the two major differences Williams acknowledges. The first is that 'in the second thought experiment, the torture is throughout presented as going to happen to me: "you", the man in charge says persistently' (: 169). In other words, after every proposed instance of psychological change induced in me, the *man in charge* repeatedly insists that the resulting person would be me. He takes it for granted that I would survive these psychological changes. He is not neutral about the outcome. On the supposition that the resulting person would be me, my fear of future torture in spite of each and every predicted psychological change would be rational. This feature is missing in the description of the first thought experiment. But Williams doesn't see why this should be a problem. He insists that through every step of the predictions made by the man in charge, I seem to be able to follow him successfully. He goes on to say that if we reflect on the fact that I was able to follow him successfully, we discover an underlying principle 'that my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time...' (: 169). This principle justifies my fear of future torture notwithstanding the vast changes in my psychology.

There are two related questions worth highlighting. First, does my following the predictions successfully guarantee that the resulting person would be me? Second, is my fear of torture justified? Perhaps I was tricked into the whole scheme such that the description of the process instantiated my persistent fear. When the ‘man in charge’ said it would be me, it is possible that I was unduly influenced by him. But maybe I really had no reason to be afraid. The issue here is that the ‘man in charge’ is not neutral with regards to the outcome of the experiment and Williams thinks he shouldn’t be. But it seems uncontroversial that bias in how a thought experiment is described downplays the significance of the conclusion it seeks to establish. As long as the second thought experiment is described in the way Williams does, it establishes nothing by way of a principle. We see this once we see that a second form of bias towards a certain criterion of identity on the part of a participant in the experiment can dictate its outcome. Suppose that this person was John Locke and is already biased towards the psychological criterion. One would expect that every suggestion that the resulting person would be him after each and all of the proposed changes would be strongly opposed. We can imagine Locke saying, ‘I resist any temptation to be horrified. I am convinced that the resulting person would not be me since I would have changed so significantly’.

We know in advance what Locke’s response would be and if the experiment was actually done on him, it would not reveal anything that we did not previously know. Likewise, if it was done on Williams we would know his response in advance and once the experimental process is compromised in that way, it reveals nothing new and interesting. These suggestions would certainly not go down well with Williams and, in fact, is the beginning of a very strong case against him. They suggest to us that not being neutral in the description of an experiment that is meant to support some principle may undermine the validity of the argument for that principle. The point here is that if re-describing the experimental process in a neutral way would change the original response, then we have reason to be suspicious of Williams’ lack of neutrality in describing the second experiment. I shall further substantiate this claim and explain why the description of the second thought experiment is actually misleading.

A second difference that Williams highlights has to do with the alleged omission of the other person *B* who featured prominently in the presentation of the first thought experiment. We must note that person *B* is merely introduced to take up ‘the somewhat incidental role of being the provenance of the impressions I [person *A*] end up with’ (: 171). According to Williams, it is not clear why this should matter. In his words, ‘[m]y selfish concern is to be told what is

going to happen to me, and now I know: torture preceded by changes in character, brain operations, changes in impressions of the past.’ (: 171). He goes on to say what happens to another may affect me in other ways but would be irrelevant to my fear of future torture predicted for *me*. Even so, does not this leave out the very person who, as we see in the first thought experiment, is the actual reason for my fear? Williams again objects by claiming that if my fear of future pain is rational in step one to five (as described in the second thought experiment), then I should expect to fear this predicted torture even in step six when person *B* is actually introduced because the difference between step five and six is too insignificant to affect my expectation of pain. This appears to be a reasonable argument.

However, it brings out another feature of the second thought experiment which marks a crucial difference between the experiments. It is that it conceals the enormity of the changes between some of the stages. It appears that there is a huge difference between the first and sixth case in Williams’ second thought experiment. This difference is not insignificant. One may easily buy into the minor changes that take place between two successive steps while the major differences between two steps that are further apart are illegitimately concealed. The strategy and problem here is similar to that of *The Problem of the Heap*. In that case as well, one is led to absurd conclusions when one allows that removing a grain of sand does not significantly affect the composition of the heap. So, the changes in step two (certain changes in character traits) and four (total psychological replacement) are totally different in kind but Williams seem to suggest otherwise. Perhaps, a more accurate description of what takes place in step two would elicit a different sort of response than the anticipated one.

Noonan²² has put forward an objection against Williams’ description of the second experiment on the basis of the foregoing idea. In particular, he points out the nature of the change in stage (iv) and (v) and then argues that a line can be drawn between them—that is, proponents of the psychological view against whom the experiment is formulated can deny that *A* survives the change induced in (v). His strategy is to restate what takes place in (v) as the inducement of memories and other psychological traits by a ‘special causal process’, which is lacking in step (iv), and, as such, what takes place in (v) involves a case in which one of *B*’s hemisphere is transferred into *A* (Noonan 2003: 185–187). So, *A* divides in just the

²² The inclusion of Noonan’s argument here is only to show one way of objecting to the success of Williams’ second experiment in the literature. I agree with him that that thought experiment fails without necessarily endorsing the very same argument he launches against it. For arguments against the success of Noonan’s case, see Beck (1998).

same way Parfit's protagonist divides in the case of 'My Division' (Parfit 1984: 254–255). As a result, Williams' objection against drawing the line between steps (iv) and (v), on the grounds that A survives step (v) precisely because B-body person remains the same, fails.

Despite these differences in description, it is clear that at least in one respect both thought experiments are held to be similar. What takes place in step six in the sequence of changes described in experiment two is very much the same with what is described in experiment one. I do not think this is contentious. The issue – and Williams admits it although he doesn't think it makes any real difference – is that both experiments are described in different ways. In my view, the first difference in description that Williams highlights, and that I have briefly discussed, largely influenced the conclusion his rendering of experiment two sought to establish. I shall now consider it in some detail.

2.5. What do the thought experiments really establish?

The issue we have been dealing with is relevant and worth considering. This is because what we say about the differences would affect our judgment about what each experiment seeks to establish. If we say the differences are trivial, then we may also say both experiments are equally plausible or implausible as the case may be. In which case, we would be anything but reductionists about personal identity. If we say the differences in description make a huge difference in the outcome of the experiments, then we would most likely be partial towards a reductionist account of personal identity. There are four possible responses we may make with regards to what the thought experiments establish. We could say:

- a) Both are convincing, thus thought experiment as a method is shown to be useless
- b) The thought experiments are convincing; this shows non-reductionism to be true
- c) One or the other of the thought experiments is not convincing, thus they establish:
 - a. Physical account of personal identity: first thought experiment is wrong
 - b. Psychological account of personal identity: second thought experiment is wrong
- d) Both thought experiments are problematic; they do not establish anything whatsoever

In my view, option a) is not true. This is because, as I shall argue, the experiments are not both convincing, which is a precondition for that option—we are *led* mistakenly to the conclusion of the second thought experiment. However, that option enjoys some support in the literature. Notable proponents of the view include Stephen White (1989) and Mark Johnston (1987). Their view is that the puzzle generated by Williams' twin experiment show that “the entire method of using thought-experiments to support theories of personal identity

is mistaken” (Beck 1998: 217). The response to which I appeal against this view is that given by Beck—he argues that the force of that view depends on whether one accepts the strong assumption that thought experiments are supposed to elicit from us “philosophically strong correct intuitions”. He is persuasive in his arguments against that assumption and in his proposal that we favour instead the view that thought experiments only show “relative strength of commitment to underlying principles which they reveal” (1998: 218–220). It is worth reiterating the point that the success of this view depends on how successful one is in showing that one or the other experiment is flawed.

Further, on a correct analysis of both experiments, I argue that option b) is untenable and is unnecessary. For it to be true it once again requires both experiments to be convincing; but they are not since, as I shall show, the description of the second experiment is misleading and does not warrant its conclusion. Even so, it is an option that enjoys a considerable amount of support from philosophers who endorse what has come to be known as the Simple View, according to which identity cannot be analysed in terms of more basic facts like physical or psychological continuity. Geoffrey Madell (1981) and Richard Swinburne (1984) both defend that view. They contend that the experiments in question attempt to analyze identity in the way they deny and, as Beck puts it, ‘together they show how identity can be unaffected by a total psychological change and a total physical change’ (1998: 217). This proves the Simple View to be true only if both experiments are indeed convincing.

Further, it is not clear that both experiments are problematic. Those who endorse option d) usually furnish arguments against both experiments, which is precisely what Noonan (2003) attempts to do. I have briefly discussed his objection to the second experiment in the previous section. I now examine his other objection against the first experiment. Noonan believes that the alleged plausibility of the first description is due to Williams misrepresenting what the post-experiment persons are likely to say. His main idea is that what they are likely to say is bound to reflect their views on identity. Noonan makes a distinction between what the post-experiment persons recall in virtue of the memories they now have, and what their own choices might be—he argues that they are entitled to dismiss their apparent memories as illusory on the assumption that they take identity to be a matter of physical continuity (2003: 181-185). Thus Williams gets it wrong precisely because he allows his own assumptions to influence the participants’ assessments of what transpires.

As it turns out, Noonan's objection is itself flawed. Beck rightly points out that "Williams' suggestions as to how A and B will react after the experiment are based on the intuition that individuals will take themselves to be the persons they feel like" (Beck 1998: 215). His denial of Noonan's claim that Williams influenced the participant's post-experiment views is given added significance by his contention that belief in the physical criterion doesn't discredit the outcome of the first thought experiment (1998: 215). It seems then that one of the experiments is misleading, while the other is not and in the absence of any convincing argument that shows the other to be problematic²³ it should be taken that option d) falls through.

My claim is that the two thought experiments establish option c). To be more specific, I claim c) b. – that the description which supports the second thought experiment is misleading. I intend to show this by arguing that the presupposition of the *man in charge* that the resulting person after each and every instance of change would be me – a major difference between both experiments – is what led to the conclusion reached in the second thought experiment. That conclusion is not warranted by the description because it is already borne out by that description.

I am going to now show this by applying Williams' description of the second thought experiment to a case that involves not psychological but physical changes. Every other thing would remain as they are in Williams' account of the second thought experiment except that now the predicted changes would happen to my (A^I 's) body and not (A^I 's) psychology. The predicted changes would be thus:

- i. A^I is subjected to an operation in which a certain device is capable of temporary holding all the information in his brain while enough of his brain (say, both hemispheres) is taken out;
- ii. The above is completed but certain other body parts – limbs, heart, etc – are changed;
- iii. Brain and Body parts are changed and each part is replaced but the surrogate parts are created from new matter and not those of any actual person;
- iv. The same as in iii, except that the body parts are designed to be appropriate to another actual person B^I ;

²³ Williams suggests that the first thought-experiment is artificially described to allow easy description as a body-swap. It is artificial since it ignores other possible scenarios such as that in step v, where there are two survivors with B's psychology. This may indeed offer a new problem for a psychological view, but it does nothing to prevent the described scenario from being a fatal counter-example to the alternative to the psychological view, the physical view. As a result I do not need to discuss his argument here. Whether the problem of reduplication is fatal to the psychological view has been well discussed elsewhere (such as Parfit 1984).

- v. The same as above except that the brain and body parts that are given to A' are actually those of B' but B' is not left the same, since the body parts initially belonging to A' are now given to B' when the operation is conducted in the reverse direction.

As in Williams' account, the man in charge predicts that after each and all of these bodily changes induced in me, that I (me and no other) would be tortured. Once again he is not neutral and we assume as Williams does that he shouldn't be. It is also true that here as in the previous case in which the changes induced applied to my psychology, I should be able to follow the man in charge successfully. And since I do, we would be led to conclude that there is an underlying principle that my fear of future torture is not excluded by bodily changes.

Is it puzzling that we should now arrive at a different principle from Williams' after using a similar description used in the Williams' rendering of the second thought experiment? There is hardly any puzzle here. If Williams was right about his conclusion – that personal identity is constituted by bodily continuity – we should expect that when the *man in charge* predicted these bodily changes for me prior to the torture, I should object and say the resulting person would certainly not be me. In other words, I should not be able to follow him successfully and my fear of torture would be irrational. But I could not object because I was able to follow him successfully and my fear of torture despite these changes was rational. I understood him quite well when he said “you” in spite of the enormous bodily changes that would precede the torture. In a sense, I seem to believe that concern for future torture is at least not concern for this particular body and that I have good reasons to be afraid of future torture predicted for me even if I undergo enormous bodily changes.

Let me summarize my case clearly: the fact that I followed the predictions of the ‘man in charge’ successfully in both versions of thought experiment two in no way indicates a principle that we can establish which we did not already presuppose. In both cases, the ‘man in charge’ says it is ‘me’ that would be tortured despite the changes in the respective cases and I followed him. The failure of William's description (and mine) to be neutral is suspect because the presupposition made implies that we would always arrive at a conclusion that reflects what the ‘man in charge’ takes for granted. When he presupposes that after each bodily change induced it would be me, I followed him successfully; when he presupposes the contrary I also followed him successfully—there seem not to be any underlying principle whatsoever that can be arrived at. But my following in both situations is indicative of the

man's absolute determination of the outcome; it suggests that I pay much more attention to his assertion that it is me, than to the details of the operation he is proposing.

Is it possible to make the same assertions about the description given in the first thought experiment? I do not think so. Unlike the second thought experiment, the description of the first thought experiment can be made in non-question-begging terms. What I have shown is that the description of the second thought experiment undermines its conclusion thus making it less convincing (and perhaps wrong) in relation to the first. If this is the case—and I think it is—then we have a good case against Williams' second thought experiment and good reasons to claim c) b. above. This is the claim that what can be established is that the experiment supporting the physical criterion on personal identity is misleading and most probably wrong. As it is, this is another way of saying that the psychological criterion on personal identity is to be preferred.

2.6. Conclusion

All along my aim has been to undermine whatever support the embodied mind view may glean from Williams' endorsement of the second thought experiment in *The Self and the Future*. To my mind, the analysis of the experiments has shown that any such support is not available. In my view, the experiment supporting the physical criterion on personal identity is misleading. This allows us to further claim that the view that physical continuity is what constitutes personal identity is most probably mistaken.

3. Surviving without a Brain

3.1. Introduction

In the first part of his book, McMahan analyzes and rejects what has been the most widely held view on identity—the psychological view. He denies that the relations of psychological connectedness and continuity are among the relations that constitute identity arguing instead that the continuity of the capacity for consciousness is what constitutes identity over time. It is central to his view that there is continuity of the physical basis of consciousness as what is both necessary and sufficient for identity and egoistic concern is the continuity of *this capacity* for consciousness realised in the *same brain*. This chapter is specifically concerned with the alleged importance of physical continuity to our survival.

To be sure, McMahan builds his case systematically beginning with thought experiments that ostensibly support the importance of physical continuity to survival. He then offers a two-fold response to Parfit’s own arguments against the importance of physical continuity—the second of which targets the plausibility of the notion of psychological continuity. In adopting this strategy, not only does he advance a positive case, mostly on the basis of thought experiments, by recommending intuitions that ostensibly support the importance of physical continuity to our continued existence, he also puts forward a negative case, which attempts to reveal the unsavoury consequences of adopting the notion of psychological continuity used by psychological theorists. This chapter is a response to his positive and negative arguments. I intend to argue that thought experiments recommending the psychological criterion are far more compelling than he allows and that psychological theorists can make do with a notion of psychological continuity that is immune to McMahan’s criticisms.

3.2. Three types of Continuities

Before dealing with the core of McMahan’s arguments, it is important to fill in the essential details of his embodied mind view so as to facilitate an adequate understanding of the issues to be considered. One of the recurring themes in his discussion of the embodied mind view of identity and egoistic concern is the continuity of the relevant areas of the brain which sustains consciousness or a mind. In an attempt to fully clarify the details of the embodied mind view, he differentiates three types of continuity concerning the brain—physical, functional and organizational continuities—that normally underlie mental life. His idea of *physical continuity* of the brain is applicable in either of two ways. It could involve the “continued

existence of the same constituent matter” of the brain in which consciousness is generated and sustained or “the gradual, incremental replacement of the constituent matter of the brain over time” (2002: 68). In the latter sense, the brain, like most other organs of the human body, can survive gradual cellular turnover, which, in his view, is congruent with physical continuity. So, the core principle in preserving physical continuity over time is that there is sufficient integration between the old and new matter, which rules out in advance the compatibility of rapid replacement with physical continuity.²⁴

Closely related to physical continuity is what McMahan calls *functional continuity* of the brain. This sort of continuity broadly requires that the basic psychological capacities of the brain are preserved and carried forward in time. As the brain has a number of capacities, McMahan is specifically concerned with the preservation of the brain’s *capacity for consciousness* as that alone will be enough for minimal functional continuity. There are obvious similarities here with Unger’s (1990) idea of core psychology, which he distinguishes from distinctive psychology. The core of one’s psychology includes basic psychological capacities for thought, consciousness etc. as distinct from the contents of one’s mental life—such as memories and beliefs. A third and obviously less important type of continuity for McMahan is *organizational continuity*. It involves the continued existence of the various tissues of the brain that normally underlie the connections among the distinctive features of a person’s psychology over time. These include the connection between an earlier experience and a later memory of it—the sort psychological theorists appeal to.

The embodied mind view stresses the first two types of continuity of the brain distinguishing itself from other views of identity that insist on the third type of continuity, which emphasizes the connections among the contents of one’s mental life. It asserts that organizational continuity is not necessary for egoistic concern for the future and is not among the relations that constitute our identity over time (McMahan 2002: 74). Accordingly, the criterion of personal identity is physical and minimal functional continuity, in non-branching form, of the brain—more specifically, enough of the relevant areas of the brain in which consciousness is generated and sustained—to be capable of preserving the *capacity* for consciousness (2002: 67–69). On this view, the continuity of the distinctive features of one’s mental life (memories,

²⁴ The operative concept here is similar to the notion of “assimilation” employed by Unger (1990) in articulating the notion of physical continuity. The idea requires that the replacement parts coexist with the old parts for some time and contribute to the process of supporting mental life.

beliefs etc) is incidental to survival. As indicated earlier on, it is central to this view that there is physical continuity of the brain in order to support the continuity of basic psychological capacities—more importantly, the capacity for consciousness. But there are other reasons why McMahan thinks that physical continuity of the brain is crucial to our survival.

Our intuitions about our survival and what matters are best captured, McMahan thinks, only if the relation of physical continuity is among the relations that constitute identity and is part of what grounds egoistic concern. This takes us deeper into considerations about thought experiments that ostensibly support the importance of physical continuity. Arguing against an earlier thought experiment, Parfit's *Teletransportation* (1984) that supposedly undermines the importance of physical continuity, McMahan puts forward a significant number of other experiments²⁵ that elicit conflicting intuitions. In one instance, he invites us to consider:

The Suicide Mission. In a time of war, one has been chosen to carry out a military mission that will involve certain death. Although the operation of the Replicator is very expensive and has therefore been strictly rationed, one's superiors have granted one the privilege of having a replica of oneself made prior to the mission. They will also allow one to choose, prior to the process of replication, whether one will go on the mission oneself or whether the replica will be sent. (Because one is a dutiful soldier, one's replica will be dutiful as well. One knows that if ordered, he will go on the mission) (McMahan 2002: 57).

McMahan's own intuitive response is to choose without any reluctance that his replica should go on the mission and he thinks that most people will be inclined to do the same. His choice is based on the evidence that what matters—physical continuity—is absent in his relation to his replica. On the face of it, the scenario shows that the continued existence of the same brain makes a huge difference to a person's survival. Although, he admits that certain cases give rise to uncertainties about what really matters, he presents other cases that tend to reinforce the importance of physical continuity. Of the others, perhaps the most controversial one is *Double Replication*:

A person steps into one of the earlier-model replicators and presses a button, whereupon his body is simultaneously scanned and destroyed, while two exact replicas are created in adjacent replicating booths. (2002: 58)

²⁵ *Deprogramming* is an example of one such thought experiment that is held to elicit the intuition that physical continuity is crucial to our continued existence and is part of what matters. A clear illustration of this, and one to which McMahan refers, is found in Williams (1970). I have argued in the previous chapter that Williams' thought experiment fails to convince.

There are obvious similarities between McMahan's *Double Replication* and Parfit's *My Division*. In both cases, there are two resulting persons neither of whom is identical to the original person. But surprisingly McMahan claims that there is a relevant difference between division via replication and division via hemispheric separation and double transplantation (2002: 59). The relevant difference, he argues, is that in the former psychological continuity is grounded in physical and functional continuity of the same brain. He then goes on to claim quite controversially that most people would not be persuaded by Parfit had he presented his fission case as involving double replication. There are no independent arguments for that claim—he simply asserts that many people will deny that they have reason to be concerned in an egoistic way for their replicas.

At this point, I wish to point to an important methodological misgiving with regards to McMahan's discussion. The use of the term 'replica' can only help to mislead most people to think that it is not oneself and cannot be one's object of future self-concern. But that in itself seems to suggest that one has an antecedent notion of identity and what matters which rules out in advance the possibility of a replica, psychologically continuous with oneself, being oneself or the object of one's egoistic concern. It seems better (if the experiment is to have some authority) to examine whether the original person in *Double Replication* has reason to care in an egoistic way about what happens to the resulting *persons* rather than his *replicas*. Even so, there is nothing about a replica that compels us to deny that we should have egoistic concern for it as long as there will be continuity of a causal kind between oneself and the replica. That a replica may be oneself as well as an object of egoistic concern is seen in McMahan's thought experiment—*The Nuclear Attack*—which shows that egoistic concern for a replica that is psychologically continuous with oneself is justified.

I will return to the details of *The Nuclear Attack* thought experiment in the last section. For the moment, I will accept that McMahan has done considerable work in persuading us as to the strength of the intuition that physical continuity is crucial to our survival.

3.3. Why Physical continuity is unimportant

The idea that physical continuity is important to our survival is not itself novel. Similar views have been defended in Unger (1990), and Williams (1970). But it is not obvious that our continued existence hinges significantly on the continuity of the physical system that sustains our mental life. Just as there are thought experiments that ostensibly elicit the intuition that

physical continuity is of great importance, so are there others that elicit conflicting intuitions. With the exception of *Teletransportation*, McMahan at no point comments on thought experiments that elicit intuitions of the kind he expressly denies.²⁶ But this is not because there is a shortage of such experiments. Locke (1694) discusses a classical example of a thought experiment in which the soul of a prince enters and animates the body of a cobbler. According to Locke, we should expect that the person in the cobbler's body is the prince having all of his memories, thoughts and beliefs. Most of us will intuitively interpret the scenario as an instance of body swap. The possibility of persons exchanging bodies in the manner suggested undermines in no small way the importance of physical continuity to our identity. That this is so is reinforced in Williams' refined version²⁷ of the thought experiment.

In Williams' portrayal of things, our imagination is thoroughly engaged. Two persons A and B undergo a process, induced by a highly sophisticated machine, during which the contents of person A's brain are transferred into that of person B with a similar operation occurring in the opposite direction. At the end of the process, the person with A's brain has the psychological features that use to belong to B and vice versa. Since A-brain person will remember living the life of B, having her beliefs and desires, our intuitive response is that A-brain person is now B. The same applies to B-brain person who is now A. If we respond this way, it follows that what is involved in our continued existence is not physical continuity of the brain, as McMahan claims, but rather psychological continuity. Suppose that person A was asked to choose beforehand which one of the resulting persons should be tortured and that he goes on to choose \$100,000 reward for B-body person whilst preferring torture for A-body person. Suppose that A's preferences were granted, we should expect B-body person, who now has A's psychology, to remember making that choice and express a great deal of satisfaction over making a wise choice (Williams 1970: 163–164).

Such a response reinforces the claim that the process involves a body swap. It does enormous damage to the view that physical continuity is crucial to survival. These thought experiments (Locke's and Williams'), therefore, tend to support the psychological criterion of identity. But not only does McMahan remain unconvinced by them, he sees no need to include them in his assessment of the significance of physical continuity to our survival. In the last section, I shall

²⁶ It is worth noting though that Parfit does not actually base any argument against physical continuity on Teletransportation.

²⁷ That it is a refined version is not only due to the manner of presentation (it explores the possible responses of the persons involved to predicted torture and reward) but also, as Beck writes, because it removes souls as "the casual mechanism of body-swap" (2001: 160).

return to these issues in an attempt to show that the kinds of thought experiments McMahan ignores are the more persuasive ones.

For now, there is still more that can be said about the unimportance of physical continuity to our continued existence. One of the highlights of McMahan's discussion of the issue is his presentation of Parfit's challenge to the significance of physical continuity. Parfit attempts to show that the difference between there being and there not being physical continuity is insignificant to a person's survival. He writes,

Suppose that I need surgery. All of my brain cells have a defect which, in time, would be fatal. But a surgeon can replace all these cells. He can insert new cells that are exact replicas of the existing cells except that they have no defect. We can distinguish two cases. In *Case One*, the surgeon performs a hundred operations. In each of these, he removes a hundredth part of my brain, and inserts a replica of this part. In *Case Two*, the surgeon follows a different procedure. He first removes all of the parts of my brain, and then inserts all of their replicas. (1984: 474)

Parfit's argument is straightforward. In *Case One*, physical continuity of the brain is preserved (in virtue of the new parts coexisting with the old ones for a time) but not in *Case Two*. However, he denies that this difference in the manner of removing and inserting parts of his brain can be the difference between life and death—i.e. can greatly affect survival. "I cannot believe" Parfit writes "that what would matter for my survival is whether, over some period, the replicas of parts of my brain would be inserted in one of these two ways" (1984: 476). The force of his denial of the significance of this difference to his survival is based on the realization that in both cases psychological continuity will be preserved. Although McMahan concedes that people's intuitions are divided, he is nonetheless convinced that with regards to Parfit's survival, it matters whether there is physical continuity of his brain.

3.4. McMahan's Rejoinder

As I see it, there are two sides to McMahan's rejoinder. He wants to claim that it is not arbitrary to suppose that the difference—there being and there not being physical continuity—between *Case One* and *Case Two* is crucial to survival. He argues that the important difference is concealed both in Parfit's description of things as well as in his defective notion of psychological continuity. I shall consider each point in turn.

The first aspect of McMahan's rejoinder is that if described differently the important difference between the cases can be brought to light. His preferred manner of description is to consider a spectrum of possible cases between *Case One* and *Case Two*. In *Case Two*, which features at one end of the spectrum, the replacement of brain parts is done all at once just as Parfit describes it. In *Case One*, which features at the other end of the spectrum, the replacements occur at equal intervals spanning a period of fifty years. As we move from *Case One* along the spectrum, the number of years decreases. As we approach *Case Two*, and as the number years decrease significantly, McMahan contends that it becomes less clear whether Parfit survives the process. He adds that "between this area of indeterminacy and *Case Two*", where the replacements occur all at once, "the rate of turnover is too great to be compatible with physical continuity" (2002: 72).

However, in the cases between *Case One* and the region of indeterminacy, one may confidently assert that Parfit survives the process precisely because physical continuity is ultimately preserved. In a nutshell, it matters how long it takes for the replacements to occur in *Case One* in order to distinguish it from *Case Two*. In the latter, there is no assimilation between old and new parts. In the former case, there is sufficient assimilation between the old and new parts compatible with physical continuity. Had Parfit described the scenario as a spectrum of cases, McMahan thinks, we would intuitively respond differently with regards to our survival. I doubt that we will respond differently even if the scenario is described in McMahan's way. The reason is that the manner of description does not undermine Parfit's contention that in both cases and whether or not there is physical continuity, there would be psychological continuity between the persons at the beginning and the end of the process; the causal connections among the distinctive features of Parfit's psychology will still hold despite there being no physical continuity in *Case Two*. This sort of continuity is enough to support a positive response (as in the case of the Prince and Cobbler).

We are yet to consider what is possibly the most damaging argument against psychological continuity. This will bring us to the second aspect of his rejoinder. Earlier in his arguments, McMahan had put forward an objection dismissing the notion of psychological continuity as it is used in the literature by psychological theorists. His contention is that the notion of psychological continuity is set out in terms of two criteria both of which yield unsavoury repercussions. The two criteria he has in mind are causal dependency and qualitative similarity of mental states as establishing psychological connections and/or continuity over

time. Thus, a memory caused by an earlier experience or a desire, which is continually had and qualitatively similar to an earlier one, qualifies as a psychological connection over time. McMahan contends that both criteria are insufficient to establish psychological connections—more, he thinks, is needed. As a result, he proposes that a more promising option is to maintain that mental states establish what he calls *real psychological connections* over time if they are grounded in the same physical realizer—i.e. the same brain (2002: 63). My plan is to spell out in some detail his arguments against both criteria and attempt a defence of a view of psychological connections and/or continuity that is resistant to McMahan's charge.

I begin with the criterion of causal dependency, which McMahan claims is otiose. The examples he uses to make his case are illuminating. Take, for instance, his example of Unintended Replication—a modified version of the case of Teletransportation²⁸.

A person dies. Immediately thereafter, the operators of a Replicator program the machine to create a person that they believe would not be a replica of any actual person. But, by an improbable coincidence, the brain and body of the person created in the replicating booth are exact cell-for-cell duplicates of the brain and body of the person who has just died. (2002: 61)

McMahan claims that psychological theorists should claim that in Unintended Replication what matters is present in the relation between person and his replica just as they claim in the case of Teletransportation. He suggests that the only difference between the two cases is that in Teletransportation replication is intended²⁹. But if what matters is present in Unintended Replication, it seems then that psychological theorist must claim that psychological continuity holds between the original person's psychological states and his replica's even though it is clear that in this case the latter is not causally dependent on the former. McMahan concludes that if this is true then the requirement of causal dependency serves no practical role in the notion of psychological connection/continuity (ibid.). He thinks it should be dispensed with.

There is clearly no causal dependency between the psychology of the original person and his replica in Unintended Replication. But this is not because replication occurs unintentionally.

²⁸ This thought experiment is similar to that of Unintended Replication discussed with a few differences, which I shall highlight later on. For a discussion of it see Parfit (1984: 199-201) and McMahan (2002: 56)

²⁹ There is clearly another important difference which McMahan ignores, and as a result of which he misinterprets the concept of causal continuity between psychological states. What he ignores is in Unintended Replication, the data used in the replicating process is not from the psychology of the original person as is the case in Teletransportation.

Rather it is because this instance of replication is not one that involves blueprints scanned from the psychology of the original person as is the case in Teletransportation. It is not intention that is does the work in Teletransportation; in fact, it is immaterial whether the process is intentional or not. If unintended replication had occurred by means of blueprints scanned from the psychology of the original person, then there will clearly be causal dependency between the original person's psychology and the replica's.

McMahan is right that in unintended replication there is no causal dependency between the original person and his replica. If Parfit and other psychological theorists claim that their envisaged psychological connections hold between them, then it is hard to see, as McMahan observes, why causal dependency is an important feature of psychological connection/continuity. But psychological theorists need not claim that there is psychological continuity between the original person and his replica in Unintended Replication. All they need claim is that the psychology of the replica is not caused by blueprints scanned from the original's psychology and this is not continuous in their sense. Accordingly, the replica is not numerically identical to the original (neither is it the case that what matters is preserved in the relation between them) though they are qualitatively similar.

McMahan's ground for dismissing the notion of psychological continuity is based on a tendentious reading of the concept. He takes the concept to mean qualitative identity—i.e. an earlier and a later psychological state constitute psychological connections, in McMahan's terms, as long as they are qualitatively identical to each other. It is for this reason that he claims that psychological theorists are committed to the claim that there is psychological continuity between the original person and his replica in Unintended Replication as they are both qualitatively identical. As I noted earlier, the psychological theorist can deny this: although the original and his replica are qualitatively identical, the psychology of the replica is not caused by the psychology of the original.

We are gradually arriving at and clarifying a more accurate conception of causal continuity, which is resistant to McMahan's charge of otiosity. On this more restrictive conception, a replica's psychological states are causally continuous with mine iff my psychological states provide the blueprints for replication. In this sense, it is immaterial whether the replicating process is intended or not—as long as the replica's psychology is causally dependent on the original's in the way described above, what matters is preserved in the relation between them.

A little more needs to be said about the other criterion that must be satisfied in order for two mental states to qualify as a psychological connection. According to McMahan, it is that the “latter state must have the phenomenological character and behavioural consequences appropriate for a normal successor state” (2002: 60). The criterion requires that there is qualitative similarity between an earlier and later mental state. He claims that this criterion inevitably results in the proliferation of psychological connections and/or continuity across lives. He reasons that “if we grant a compatible latitude in establishing quasi-memories, while at the same time dispensing with the requirement of causal dependency, it seems that psychological connections across different lives will be abundant” (: 61). But, as I have argued, we are not rationally compelled to give up the requirement of causal dependency—at least not the conception of it put forward here. The requirement of qualitative similarity survives McMahan’s charge—since we have no reason to dispense with causal dependency, his envisaged proliferation of psychological connections across lives disappears.

Recall that Parfit’s argument against the importance of physical continuity of the brain, as discussed in the previous section, was intended to show that what matters—psychological continuity—is preserved in *Case One* and *Case Two* of brain replacements. McMahan’s strategy was to show that there is something wrong with that notion of psychological continuity that may hold between persons at different times even in the absence of physical continuity of the brain (as shown in Parfit’s arguments in section three). But he has clearly not established his case against a plausible notion of psychological connections and/or continuity as well as the unimportance of physical continuity.

3.5. Which Intuition is Stronger?

At the beginning, I indicated that McMahan recommends, on the basis of carefully chosen thought experiments, the intuition that our survival involves going wherever our functional brain goes. In the second section, I discussed other thought experiments that elicit conflicting intuitions about our survival—in particular those of Locke and Williams, which support the view that our survival involves going wherever the distinctive features of our psychology go. In view of that, the contending intuitions are either in favour of core psychology or distinctive psychology³⁰. It is worth noting that McMahan takes core psychology to be individuated by brains, hence the emphasis on physical continuity of the brain. In very broad terms, with

³⁰ A detailed discussion of the distinction between core and distinctive psychology is found in Unger (1990, 1992b).

regards to intuitions, it is a choice between physical and psychological continuity both of which normally coincides in actual persons, but has been shown to come apart by various thought experiments. McMahan accepts that they can be prised apart though he denies that psychological continuity is what matters.

My aim in this section is to show why thought experiments eliciting intuitions that recommend the psychological criterion are on the whole more persuasive than conflicting ones. I have already mentioned two of the thought experiments that McMahan thinks support the physical criterion³¹. Of the two, *The Suicide Mission*, in my view, is far more interesting and revealing. McMahan's intuitive response to this case is to choose his replica, rather than himself, to go on the mission. His choice is based on the consideration that with regards to one's replica, what matters is lacking as there is no physical continuity with oneself. The lack of physical continuity seems to make the difference between caring for and not caring, in an egoistic way, for one's replica. I believe that this particular thought experiment shows that there can be disagreements regarding which intuition is the much clearer and stronger one concerning survival. But I suspect that it is the sort of thought experiment or that it is described in a way that reveals our strongly held beliefs about survival.

Consider the following case. I am told that, by means of the brain-state-transfer device, my distinctive psychology will be recorded and my brain will be wiped clear of distinctive psychology. This will only be for a moment. As this process occurs, an exact replica of my body in the resulting state (minus a distinctive psychology) is created from new matter. Before the process occurred, a computer was programmed to construct a distinctive psychology based on the general facts of my life – similar to the scenario envisaged of B in step (iv) of William's second presentation in 'The Self and the Future'. Once the process of copying and clearing is complete, the person operating the computer will flick a switch and the constructed psychology will be fed into my original body while the recorded psychology is fed into the replica body.

If I am asked beforehand which resultant person should be rewarded and which tortured, thinking self-interestedly, which should I choose? Or, using McMahan's scenario, if one of

³¹ The case of *Multiple Replication* is another thought experiment that McMahan discusses. I do not see the need to discuss here as what will be said will apply very broadly to it as well.

the resultant people is to be sent on a suicide mission while the other stays behind, which should I choose to go on the mission?

McMahan's view implies that I should have no hesitation: I should reward my original body and have the copy tortured. In the second scenario, I should send the replica body on the mission and leave my original body at home. But not only do I hesitate in making this decision, I go very firmly the other way. It seems, then, that the presence of my core psychology is not sufficient to draw my commitments to myself when my distinctive psychology is the other option. In McMahan's own discussion of the suicide mission, both people have my distinctive psychology while only one has my core psychology. In that case, my decision goes with the option that offers me more of the original—but it does not show that my commitment to my core psychology outweighs my commitment to my distinctive psychology. When compared with the current scenario that becomes obvious: McMahan is offering the wrong dilemma. And the more appropriate dilemma offers his view no support.

A choice between sending oneself and one's replica on a suicide mission gives one considerable latitude with regards to one's beliefs about the situation. And choosing between oneself and one's replica may turn out to be an easy enough choice for most people. But when it really comes down to it most people will intuitively respond that they can survive as a replica that is psychologically continuous to them. This is illustrated in *The Nuclear Attack* scenario, which McMahan himself describes:

One is an employee at the Pentagon, which has a Replicator capable of transmitting one's cellular blueprint to a replicating booth in Alaska. One receives confirmation that a nuclear missile, targeted on the Pentagon, has penetrated the country's defenses and will obliterate the entire area within a minute. That is just enough time to have oneself scanned and for the data to be transmitted to Alaska. (2002: 58).

McMahan concedes that most people will accept that replication in this sense will offer all, not part, of what matters and they will, without intention of a taking a gamble, opt for it. But, since the response is not compatible with the physical criterion and since it is not in harmony with his intuitive response to, he denies that this is a clear case of survival. There are no arguments for his denial; he only claims that on the grounds that one denies that there are grounds for egoistic concern for one's replica in *The Suicide Mission*, it follows that in the case of *The Nuclear Attack* we should doubt that what matters is present. I believe that this

latter case is an example of one in which people's strongly held beliefs are elicited; when people's beliefs are tested in this way, there is no doubting that they will choose to go with their distinctive psychology. In this case, things are much clearer: there will be someone on the other side—in Alaska—who will have one's memories, beliefs, desires and personality traits. There are good reasons to think that the inheritor of one's distinctive, rather than core, psychology is oneself.

One such reason has been suggested by Simon Beck (2001). He suggests that since there is nothing sufficiently unique about core psychology—which is common to everyone—to constitute identity over time, it is more plausible to maintain that it is distinctive psychology that does the important work for identity (: 166). As my focus in this section has been on thought experiments, perhaps it is apt to present one he discusses and which highlights the point that the notion of the same brain or physical continuity is irrelevant to one's survival and that it is, in fact, the continuity of the distinctive features of one's psychology that is crucial. He writes,

Brown and Robinson have their brains removed from their bodies. Using a brain-state transfer device, distinctive psychologies are swapped between the removed brains. Now Robinson's brain is placed into Brown's body and vice-versa. In Robinson's body is someone who thinks they are Robinson; in Brown's body is someone who thinks they are Brown. The question as to who they really *are* seems to receive a straightforward answer. (Beck 2002: 166)

The answer is that Brown's body person is Brown and Robinson's body person is Robinson. But because they are both lacking their core psychology individuated by its physical realizer (the brain), which McMahan thinks is crucial to survival, our intuitive response to the case is determined by the fact they retain their distinctive psychology. The important point which Beck highlights is that although the intuitive response is that the person in Brown's body is Brown, the core of Brown's psychology as well as any fact about physical continuity of the brain is of no significance to our intuitive response to the case. This is because Brown's brain is now in Robinson's body. The same goes for Robinson as well. McMahan must claim that intuitively Brown's body is now Robinson. But nobody will have that intuition.

It is important to note though that I am not claiming that thought experiments of this sort as well as those of Locke and Williams are the final court of appeal with regards to the

conditions of our survival. But if they reveal anything by way of our strongly held beliefs about survival, then they reveal quite persuasively and decisively that the continuity of the distinctive features of a person's psychology is both necessary and sufficient for survival.

4.6. Conclusion

With regards to the necessary conditions of our survival, McMahan has not shown that our intuitions about what matters are best captured if physical continuity of the brain are among the relations that constitute identity and egoistic concern. Despite certain cases in which uncertainties tend to arise, thought experiments that support the psychological criterion are generally more persuasive. In addition, McMahan's arguments against the plausibility of the notion of psychological connections and/or continuity fail to convince. The arguments fall far short from showing that these sorts of connections are not necessary for survival. Together these points show that physical continuity of the brain is not as crucial to survival as McMahan may be inclined to think.

4. Courting the Enemy: McMahan on the Unity of the Mind

4.1. Introduction

One of McMahan's main objectives is to defend the view that we are essentially minds and that our identity consists in the continued existence of functioning brain to be capable of supporting a mind. Despite its initial attractions, important questions arise for that view with regards to the unity of the mind. This chapter is an assessment of the performance of McMahan's embodied mind view on certain phenomena—in particular, commissurotomy—that challenge the unity of the mind. Available experimental data seem to reveal that a single mind can divide into two independent streams in ways that are incompatible with our intuitive notion of a mind. This phenomenon casts doubts on the possibility of individuating minds. McMahan attempts a solution to the problem the core of which requires us to give up our intuitive identification of same mind with same consciousness.

My aim is to discuss in detail the envisaged problem for McMahan's embodied mind view and proceed to argue that his proposed solution fails: it appeals to the distinctive features of mental life—of the sort employed by his professed enemy, the psychological theorist—which he denies any importance in these matters. In the end, I also suggest that the lack of clarity with regards to certain issues opens up McMahan's account to criticisms that would normally be laid at the door of non-reductionists.

4.2. Why the Standard View is flawed

In recent years, the dominant view in the literature on personal identity has been labelled the “standard view” according to which identity is a matter of a relation of psychological continuity between persons existing at different times. Shoemaker exemplifies the criterion as follows: “X at t_1 is the same person as Y at t_2 if and only if Y is uniquely psychologically continuous with X” (2009: 61).

The usual sort of clarifications to the criterion involve what proponents of the view take psychological continuity and the condition of uniqueness to be. Parfit defines the former in terms of psychological connectedness: the relation of memory to an experience, the continuance of a belief, desire, or character traits between X and Y etc. constitute psychological connectedness over time. On his view, there is psychological continuity when there are “overlapping chains of *strong* connectedness” and since what counts as strong

psychological connectedness is profoundly vague, he claims that the connections are *enough* if there are “over any day, *at least half* the number of direct connections that hold, over everyday, in the lives of nearly every actual person” (Parfit 1984: 206). Likewise, the condition of uniqueness is crucial given considerations that psychological continuity could take a branching form as in Parfit’s much debated *My Division* scenario. The uniqueness clause ensures that there is no other person Z who bears a similar relation to X as Y (Shoemaker 2009: 64). As such, the uniqueness clause ensures that the criterion captures the logic of identity³².

There are good reasons why the psychological view has enjoyed a large following. Its main appeal is one that all versions of the psychological view retain namely, that it accounts well for rational anticipation and self-concern. Suppose identity is indeed what matters, the view implies that I can rationally anticipate the future experiences of or be egoistically concerned for the well-being of the individual that inherits my psychology. Consequently, I cannot rationally anticipate the experiences of or be especially concerned for someone who is not psychologically related to me. In addition, it is compatible with our ordinary criterion of self-identification: I am aware that I am the same person that got into bed the previous night not by observing my body but, as Shoemaker puts it, by “simple introspection... consisting in memories, intentions-to-be-fulfilled, persistence of beliefs/desires/goals, and similarity of character” (2009: 65). Despite these attractions, the psychological view seems to be troubled by a host of other issues, which together explain doubts over its status as the standard view.

I now turn to some specific objections that have been laid at the door of psychological theorists by McMahan citing other philosophers who have raised similar objections. The psychological view is committed to our being essentially persons. In the literature, the conventional use of the term suggests a being with “a rich and complex mental life, a mental life of a high order of sophistication” (McMahan 2002: 45). Since infants lack this complexity, we are led to the apparently ridiculous conclusion that we, who are now persons, were never infants. The problem is generally taken to challenge the assumption that we are essentially persons; we begin and cease to exist as persons. McMahan takes the charge further. He claims since on the psychological view we are mistaken to think that we, who are

³² Note here that when faced with a scenario in which two individuals are psychologically continuous with an original one, proponents of the psychological view like Parfit, McMahan, and Shoemaker often resort to the view that identity is not what matters—which is perhaps the best explanation of the scenario.

now persons, were once conscious infants, it must be the case that the existence of an actual person is usually preceded by a conscious subject. McMahan writes,

There was, in short, a conscious being present both immediately before and after the birth of one's organism—a being whose mental life was in some ways continuous with one's own... But the assumption that this being was actually oneself is, of course, mistaken if the Psychological Account is right... Who, or what, was this conscious being? (2002: 46).

His response is that this must be some sort of pre-person, “a subpersonal subject of consciousness...” (or post-person in end of life cases—e.g. Alzheimer's), lacking the capacity for self-consciousness as well as enough strong psychological connections that are characteristic of nearly every actual person. The objection then is that the psychological view as it stands posits the existence of more new entities, pre- persons, which offends against parsimony—the view that the more new entities a theory postulates, the less plausible it is (2002: 46–47). Similar considerations apply to end of life cases as in Alzheimer's disease when a person's mental life gradually erodes. The suggestion is that in this case one existence is followed by the existence of some post- person.

Other problems emerge from considerations on the *Method of Cases*³³, which is believed to strongly reinforce the psychological view. The strategy of psychological theorists often involves relying on fictional cases like Teletransportation and Fission³⁴ to elicit our intuitions on personal identity. There are two related problems here for the psychological view. First, the cases that involve body-swapping tend to objectify the mind as it assumes that one could transmigrate bodies even without brain transplant. De Grazia suggests that this assumption is plausible only if substance dualism, which most psychological theorists reject, is true (2005: 419). Second, consideration of a wide range of cases reveals that we are not essentially persons but bare subjects—i.e. beings with a capacity for consciousness. The objection bears striking resemblance to Williams' view that concern for the future is not precluded by a profound discontinuity in one's psychology (1970) and, on a similar principle, McMahan's assertion that we could survive as non-persons seeing that we could rationally fear future pain in the absence of psychological connections from day to day (2002: 66).

³³ Besides what I have discussed here as the problems the intuitive case method pose to the standard view, other philosophers have challenged the very possibility of the scenarios described. They take it to be deeply impossible. See Mark Johnston (1997) and Shoemaker (2009).

³⁴ Both cases are discussed extensively in Parfit (1984), McMahan (2002) and Shoemaker (2009).

Clearly, these problems cannot be ignored by the defender of the psychological view and it appears that the standard view is problematic and stands in need of substantial revision.

4.3. An Alternative Account

For McMahan, any attempt to revise the standard view will prove to be unsuccessful. More specifically, that view will prove insufficient in capturing our intuitions about what matters. It seems reasonable then to explore other alternatives, which retain the merits of the standard view. This is precisely what McMahan sets out to do. He proposes the embodied mind view according to which we are essentially minds and the criterion of identity³⁵ is the “continued existence and functioning, in non-branching form, of enough of the same brain to be capable of generating consciousness or mental activity” (2002: 68). Unlike the psychological view, which stresses the continuity of one’s psychological life, the embodied mind view stresses “the continuity of psychological *capacities*—in particular, the capacity for consciousness” (ibid.). McMahan claims that the notion of same consciousness as used here is equivalent to the notion of same mind. Thus, our identity consists in the survival of enough of those areas of the brain in which consciousness is realised to be capable of supporting a mind.

The embodied mind view has some merits of its own. In general, it deals well with some of the problems associated with the psychological view. It is resistant to the newborn problem highlighted earlier. Since it maintains that we begin to exist when the brain develops the capacity to support a mind, it follows that we begin to exist sometime before we were born. The emphasis on psychological capacities rather than the contents of one’s mental life (as in the case of the psychological view) enables the embodied mind view to accommodate the intuition that one continues to exist and is justified in fearing future pain even after psychological connections from day to day have ceased to hold. Further, the insistence on the *same brain* or a physical basis for minds precludes the possibility of reifying minds. But more importantly, it offers the embodied mind view a criterion for individuating minds. As McMahan writes, “a mind, it seems, is individuated by reference to its physical embodiment, just as an individual mental state is.” (2002: 67).

³⁵ Since McMahan prefers a theory that seeks to establish congruence between the theories of identity and egoistic concern, he proposes that physical and functional continuity of enough of the same brain capable of generating consciousness is both necessary and sufficient for egoistic concern (though unlike identity, this is a matter of degree).

Clearly, the suggestion is that the notion of the same mind depends on the continued existence and functioning of enough of the same brain in which it is realised; in other words, this is a case of a mind for a single functioning brain. However, the possibility of a single brain supporting more than one mind raises interesting questions for the embodied mind view and it is to this issue I now turn.

4.4. A Tale of Two Minds?

The embodied mind view accepts that we are essentially minds. On this view, the continued existence of a mind is dependent on the physical and functional continuity of the same brain in which it is generated. Intuitively, our concept of a mind is characterised by the synchronic unity of consciousness; the various experiences and conscious states of a person at a given time are held to be accessible in one unified stream of consciousness. The minds of most people are like this and each person is individuated on the basis of this unity. In view of that, an important question that arises for the embodied mind view is whether the same brain could support the existence of more than one mind at a time or serially. These possibilities would threaten our assumptions about the unity of the mind. In particular, there would be no clear criterion for individuating minds—and therefore beings like us. The embodied mind view attempts to individuate minds by reference to the same brain but the possibility of the same brain supporting more than one mind is precisely what is at issue.

That possibility is not far-fetched. McMahan himself concedes that certain phenomena challenge his account of the unity of the mind. He is specifically concerned about cases of brain-bisection (or commissurotomy)³⁶, which seem to suggest that the same brain could sustain more than one mind at a time—a proposition that spells imminent doom for the embodied mind view.

There is a rich and fascinating literature on the unity of the mind with particular reference to what actual cases of commissurotomy imply about that unity. The starting point of most philosophers working in this area has been psychological studies³⁷ on patients who have undergone commissurotomy. This is a surgical operation in which the corpus callosum, which

³⁶ Cases of multiple personality disorder also pose a similar challenge but McMahan prefers not to comment on these cases as he thinks that there is at yet little understanding of what happens in these cases, and in particular, little understanding of the neurological basis of the disorder pp. 87. I strongly disagree with McMahan's way of dismissing the problem but I will let it go for now.

³⁷ These studies were mostly done in the early 1960s by neuro-scientist Roger Sperry (1968) of the California Institute of Technology.

connects the two cerebral hemispheres of the brain, is severed resulting in an apparent loss of information transfer between the two hemispheres. Following the operation, these patients behave normally and most observers tend not to notice anomalies in their personalities. Even so, it is well documented that they exhibit philosophically puzzling behaviours under certain experimental conditions. Charles Marks (1981) narrates an interesting and revealing case:

A subject, S, is told to fixate a point on a screen before him. ‘Key ring’ is flashed on the screen for a tenth of a second, with ‘key’ appearing to the left of the fixation point and ‘ring’ to the right. Since the time is too brief for eye movement, the information from the right visual field (‘ring’) is projected exclusively to the left hemisphere and information from the left visual field (‘key’) is projected exclusively to the right hemisphere... If S is asked to sort through an array of items (concealed from sight) with both hands and pick out what he saw, the right and left hands work independently. The right hand will pick up and reject a key before settling on a ring; the left hand will pick up and reject a ring before settling on a key. In general, when the response demanded is controlled by the left hemisphere, it indicates that S was aware of ‘ring’ and unaware of ‘key’; when the response demanded is controlled by the right hemisphere, it indicates that S was aware of ‘key’ and unaware of ‘ring’. Someone seems to have seen ‘key’ and someone seems to have seen ‘ring’ and they seem unaware of each other.³⁸ (Marks 1981: 4–5)

Clearly, the ability of the left and right hemispheres to function independently is crucial in order to fully grasp what happens in such controlled conditions. The relevant information here is that the right hemisphere controls movement associated with the left part of the subject’s body and vice versa. That there is no single awareness that *someone seems to have seen ‘key’ and someone seems to have seen ‘ring’* often invite the suggestion that there are therefore two subjects of experience since the sort of mental processes that take place in each hemisphere are similar to that in normal persons, who are usually taken to be subjects of experience (Mark 1981: 7). As a result, it is generally agreed within the literature that there are two streams of consciousness—though some people deny that they are therefore two minds.

Nagel (1971) puts forward five interpretations of the experimental data. I shall follow his taxonomy and in the process locate McMahan’s solution to the split-brain problem in the literature.

³⁸ There are other anomalies generated under similar conditions. For instance, the subject is shown to be unable to state the identity of an object placed in his left hand. The identity of that object is available to the right hemisphere. The subject cannot identify the object in his left hand as its identity is not available to the left hemisphere, which controls speech (and writing). But if asked to state the identity of the object, the subject will smile or frown (as she can’t say it) depending on whether the answer is correct or wrong. This is because it is the right hemisphere that knows the identity of the object.

1. The patients have one fairly normal mind associated with the left hemisphere, and the responses emanating from the nonverbal right hemisphere are the responses of an automaton, and are not produced by conscious mental processes.
2. The patients have only one mind, associated with the left hemisphere, but there also occur (associated with the right hemisphere) isolated conscious mental phenomena, not integrated into a mind at all, though they can perhaps be ascribed to the organism.
3. The patients have two minds, one which can talk and one which can't.³⁹
4. They have one mind, whose contents derive from both hemispheres and are rather peculiar and dissociated (i.e. they are not well integrated as normal minds)
5. They have one normal mind most of the time, while the hemispheres are functioning in parallel but two minds are elicited by the experimental situations which yield the interesting results. (Perhaps the single mind splits in two and reconvenes after the experiment is over.)⁴⁰

Let me begin with a brief comment on the first interpretation before proceeding to state the problems associated with each of the interpretations given. The activities of the nonverbal right hemisphere are taken to be responses of an automaton and thus not conscious. The reason is that the split-brain subject verbally denies any awareness of the activities of the right hemisphere. But, as Nagel argues, that sort of reasoning begs the question as the capacity for speech is controlled entirely by the left hemisphere. He also adds that taking verbalizability as a necessary condition of consciousness cannot be justified in principle (Nagel 1971: 403). In any case, the problem with the first two interpretations is that their denial of a mind to the right hemisphere is arbitrary. In the event that the left hemisphere ceases to function completely, the independent functioning of the right hemisphere will suffice to ascribe a mind to the split-brain patient.

The two-mind interpretation of the experimental data is implicated by the observation that the functions of the two hemispheres are well integrated in everyday life. The fifth interpretation is independently implausible as there are no physiological changes induced in the patients during the experimentation that accounts for them shuttling between one and two minds. For our purposes, the fourth interpretation requires detailed consideration as McMahan's conclusion on the split-brain problem corresponds to it.

³⁹ A variation of this sort response—one which proposes that experimental data reveals that there is one person with two minds—is found in Bogen (1969).

⁴⁰ Nagel (1971: 402–403).

McMahan claims that commissurotomy patients have one mind involving two centres or fields of consciousness. On his view, two or more fields of consciousness are held to constitute a single mind rather than a collection of minds, not merely because they are supported by one brain, but because there is sufficient integration among the various mental events generated in the different fields. The plausibility of recognizing two minds increases as the degree of separation between the fields increases (2002: 87–88). Here, McMahan reasons on the basis of observations that there is no incongruity in the behaviour of split-brain patients in everyday life thus indicating the presence of a single mind. It is for this reason he concedes that it will be plausible to claim that the patients have two minds both supported by a single brain, if and only if it was observed that their behaviours were consistently incongruous.

The obvious problem with this solution is that in affirming a single mind it faces the difficulty of making sense of what is observed under experimental conditions in which is revealed a clear disintegration that is inconsistent with a single mind. It appears that a single mind can attend to different tasks simultaneously without it being aware of what is happening in each independent field of consciousness. As Nagel writes, “...in these patients there appear to be things happening *simultaneously* which cannot fit into a single mind: simultaneous attention to two incompatible tasks, for example, without interaction between the purposes of the left and right hand” (1971: 407). Clearly, the real puzzle comes in there being conscious states to which other conscious states have no access. In supposing that a single mind can encompass two or more centres of consciousness, the solution violates our intuitive assumption about the unity of mind. It is for this reason that McMahan suggests that we give up our intuitive understanding of same mind with same consciousness. He thinks that split-brain cases reveal that our concept of a mind can tolerate some degree of disintegration.

4.5. Some Notes on McMahan’s Solution

In reaching his conclusions on the split-brain conundrum, McMahan glosses over a number of salient issues. He deals mostly with the observation that the patients behave normally in everyday life claiming that if the reverse was true, then commissurotomy would be seen as a case of Division. However, it is precisely the goings-on in the experimental room that generates the interesting philosophical quandaries. Most attempts have been not only to account for the coordinated behaviour of commissurotomy patients but more importantly to make sense of what is observed under experimental conditions. In giving short shrift to the

latter, McMahan makes his proposal of giving up the intuitive identification of same mind with same consciousness appear to be the solution to the problem.

The proposal turns out to be a major leap as the experimental data doesn't necessarily challenge our intuitive identification of same mind with same consciousness but rather questions McMahan's assumption that we are essentially embodied minds. The independent functioning of each hemisphere seems to point to the presence of two minds (in the sense that we normally understand a mind) supported by two hemispheres. The conscious activities associated with each hemisphere are in every way similar to those attributed to a normal mind as we intuitively understand it (supposing that one of the hemispheres ceases to function during the operation; we certainly cannot deny the surviving single-hemisphere patient a mind). I doubt that it is true that whether a field of consciousness qualifies as a mind is dependent on the non-existence of another field. There is nothing in the experimental data of split-brain patients that compels us to deny the presence of two minds or to give up our intuitive identification of same mind with same consciousness. Does the observable behavioural integration of commissurotomy patients in everyday life compel us to do so?

For McMahan, the behavioural integration observed in everyday life undermines a two-mind reading of the split-brain syndrome. Accordingly, he proposes that cases of commissurotomy are best understood as involving two streams of consciousness (a disunified consciousness) both of which are part of a single mind. The notion of a mind with two independent streams of consciousness or a divided consciousness is counterintuitive. Thus, he claims that cases of commissurotomy challenge our intuitive understanding of mind; in particular, our intuitive identification of same mind with same consciousness. If we give up that intuition, then our concept of a mind can tolerate a certain degree of disintegration observable in the experimental situation. The reason he thinks that there is a single mind is that there is sufficient degree of mental integration (as evident in the behaviour of these patients) between the two streams, which indicates the presence of a single mind. As the degree of separation increases, each centre of consciousness may be thought of as an independent mind. It is worth noting that on his view, then, a mind can be individuated not only by reference to the same brain in which it is generated but also by the degree of integration among the various mental events (McMahan 2002: 88). It goes without saying that this is a significant departure from his expressed view.

There are obvious problems with McMahan's solution to the split-brain syndrome. All along we have been led to believe that a mind is a bare substance that begins to exist when a brain develops the capacity to support consciousness; that we, who are essentially minds, can survive as bare subjects/consciousness without any degree of interrelated mental events. Now, the notion of a mind seems to require integration of various mental events of the sort psychological theorists appeal to and which on McMahan's view is not necessary (though sufficient) for identity. The necessary features of our continued existence depend on what we essentially are. On the embodied mind view, the continued existence of a mind depends on a brain retaining its *capacity* for consciousness; the continuity of the distinctive features of one's mental life is not among the relations that constitute identity. Yet, McMahan has now introduced it as necessary for individuating minds.

But even more important is the issue of the dependence of the mind on the brain. For it is not clear now why such an enormous alteration in the anatomy of the brain of commissurotomy patients should bring about two streams of consciousness and yet preserve the unity of the mind, which all along has been held to depend on the physical processes of the brain. Apparently, what we have is the extravagant supposition that an account of identity must include a brain, a stream(s) of consciousness and a mind—assuming that a previously unified stream of consciousness can divide as a result of brain bisection without a corresponding division of the mind. In other words, the mind doesn't seem to be *directly* dependent on the brain. (This is partly why I suggested earlier on that cases of commissurotomy pose a unique challenge to McMahan's embodied mind view).

Recall that McMahan's solution to the split-brain syndrome is based on observation that the patients behave normally outside experimental conditions and this indicates a sufficient degree of mental integration (which indicates a single mind). I think that for the solution to work he needs to provide a plausible account of a mind or else we are not compelled to give up our intuitive identification of same mind with same consciousness. I propose that he has to define a mind as a bare substance that underlies a possibly disunified consciousness resulting from major changes in the anatomy of a normal brain and is picked out by its role in behavioural integration. The definition seems fair. After all, it is his view that the unity of the mind is preserved following brain bisection. The mind, whatever it is, must underlie both the severing of the corpus callosum and the resulting fields of consciousness. Additionally, the unity of the mind is explained by the behavioural integration observed in split-brain patients

(i.e. the coordinated behaviour of the patients suggest a sufficient degree of mental integration of mental events in both fields, which in turn suggests the presence of a single mind).

I doubt that he can define a mind in the way I have suggested without playing down the significance of the brain, which is what gave the embodied mind view its initial appeal. As brain bisection does not divide the mind but divides what was previously a stream of consciousness, the operation itself may be said to have no direct effect on the mind. But the embodied mind view as McMahan presented it rests on the notion of the dependence of a mind on a brain. Given that the undivided existence of a mind underlies brain bisection, it is not clear what sort of mind/brain dependence McMahan relies on. Likewise, if he is right that we plausibly recognise the presence of a mind when there is sufficient mental integration accounting for coordinated behaviour, then what is the relevance of the physical system in which the mental events are generated?

McMahan's solution to the split-brain syndrome is remarkably similar to the one proposed by Charles Marks (1981). As I see it the virtue of the latter's account is that it gives little importance to the centrality of the brain. In fact, it employs a functionalist account of a mind taken to be part of a theory, which intends, amongst other things, to explain behaviour (1981: 33). As such, he claims that the divided brain accounts for the divided consciousness but the coordinated behaviour of the patients in everyday life is a function of a single mind. It is consistent with his view of mind that it does not require a particular physical system in which the various mental events are realised. (I suspect that there are other problems with Mark's arguments, but that is beyond the immediate scope of the chapter). In any case, it doesn't seem as though McMahan can appeal to the notion of mind as that which explains behaviour (as Mark does) without undermining the centrality of the notion of the same brain and the mind's dependence on it. This I suggest is a price he is not willing to pay.

All along, my intention has been to show that McMahan needs to provide a plausible account of mind in order to make sense of the split-brain syndrome—but he clearly hasn't done this. The story of a sufficient mental integration as evidence of a single mind might be useful in accounting for lack of behavioural anomaly in the everyday life of split-brain patients. But it doesn't seem to settle the quandary in the experimentation room. For what is clearly lacking there is precisely the sort of mental integration or coordinated behaviour that is evidence for a single mind. McMahan at no point deals with the issue of how many minds a split-brain

patient has when in the controlled situation. It appears to me that if the coordinated behaviour and therefore mental integration that is evidence for a single mind is lacking in the controlled situation, then there is no justification for the view that there is a single mind. And if it is true that the independent functioning of each hemisphere is equivalent to that of a normal mind, it seems then that McMahan will have to admit that the patients have two minds under experimental conditions and one outside those conditions—a thesis that is hard to believe.

Perhaps, a little more needs to be said about McMahan's appeal to mental integration as doing the important work in the individuation of minds. It is important to reiterate that this is a radical departure from his expressed view, according to which the continuity of a brain's *capacity* for consciousness is what is what individuates a mind. This view does not require the *contents* of mental life, of the sort psychological theorists trade on. In split-brain patients, there are arguably two *capacities* supported by each hemisphere. McMahan's case for a single mind appeals to the integration of the *contents* of the mental life of split-brain patients. But this is precisely the kind of moves he is not entitled to having denied that the *contents* of one's mental life is crucial to one's continued existence. In fact, this is the sort of moves we expect his professed enemy—the psychological theorist who stresses the continuity of the *contents* of one's mental life—to make. Clearly, McMahan is treading on thin ice and playing into the hands of those with whom he had previously and unambiguously refused amity.

The other hard questions that are left unanswered concern his alliances with non-reductionism. It will not be surprising for a non-reductionist to claim that the unity of the mind is a fact over and above the phenomenon of brain bisection. McMahan is clearly a reductionist about identity and so it will be surprising to hear him speak in these terms. But perhaps, the lack of an adequate characterization of the mind; the implicit suggestion that a single mind underlie brain bisection and a divided consciousness; and the suspicion that the unity of the mind is a fact that underlies these enormous anatomical changes in the brain together raises concerns about how much of non-reductionism McMahan's account of the unity of the mind is willing to accommodate. To the extent that these claims are well-substantiated, his discussion on the unity of the mind might risk facing criticisms that will normally target non-reductionism.

5.6. Conclusion

All that can be said, as far as McMahan's discussion on the unity and individuation of the mind is concerned, is that his sympathies lie elsewhere than where he professes them to be. He must appeal to the sort of connections among mental states which he claims is not crucial in these matters and on the grounds of which he criticizes the psychological view. As such, his solution to the split-brain conundrum fails to convince.

Conclusion

I indicated at the beginning that the aim of this research project is to respond to some of the criticisms McMahan levelled against the psychological view and to assess the plausibility of his embodied mind view. In the first chapter, I attempted to accomplish that aim by arguing that McMahan's objections against the psychological view fail to convince. Some of the important insights from that chapter include the claim that he offers a tendentious and less charitable reading of Parfit's view, which is the target of his criticisms. I also argued against the success of the isolated subject charge by analyzing and rejecting two ways of understanding the charge.

The other chapters involve arguments that target the plausibility of the embodied mind view. In the second and third chapters, I argue comprehensively against the importance of physical continuity to our survival since one of the perceived incentives for McMahan's embodied mind view is the assumption that our intuitions about survival are best captured only if the relation of physical continuity is among the relations that constitute identity and egoistic concern.

The fourth chapter reveals the failure of the embodied mind view to deal with the split-brain syndrome. This phenomenon occasions the possibility of the same brain sustaining more than one mind (as we intuitively understand it). This goes contrary to McMahan's thesis that affixes the notion of same mind to the same brain.

As a final point, what remains to be considered pertains to what we should make of the issues and arguments covered in the research project. My main aim has been to unearth some of the problems that plague the embodied mind view as a metaphysical account of personal identity. The dissertation is emphatic in its submission that there are indeed many unresolved issues that McMahan has to deal with. In reaching some of the conclusions, I have made claims and arguments that tend to offer support to the psychological view. However, the arguments in the dissertation do not suggest that the psychological criterion of personal identity is the right one. Where the dissertation defends the psychological view and offers criticisms against the embodied mind view, it only seeks to show that the psychological view may point us in the right direction with regards to our deep commitments about personal identity and egoistic concern.

Bibliography

Beck, S. 1998. Back to the Self and the Future. *South African Journal of Philosophy* 17, 211–225.

Beck, S. 2001. Let's Exist Again (like we did last summer). *South African Journal of Philosophy* 20, 159–170.

Bogen, J. E. (1969). The Other Side of the Brain II. An Appositional Mind. *Bulletin of the Los Angeles Neurological Society* 34, 135–162.

DeGrazia, D. 2003. Identity, Killing, and the Boundaries of Our Existence. *Philosophy & Public Affairs* 31, 413–442.

DeGrazia, D. 2005. *Human Identity and Bioethics*. Cambridge: University Press.

Johnston, M. 1987. Human Beings. *Journal of Philosophy* 84, 59–83

Locke, J. 1694. Of Identity and Diversity in *Essay Concerning Human Understanding*, reprinted in Perry 1975 pp. 33–52.

Madell, G. 1981. *The Identity of the Self*. Edinburgh: University Press.

McMahan, J. 2002. *Ethics of Killing: Problems at the Margins of life*. Oxford: Oxford University Press.

Nagel, T. 1971. Brain Bisection and the Unity of Consciousness. *Synthese* 22, 396–413.

Noonan, H. 1989. *Personal Identity*. London: Routledge.

Noonan, H. W. 2003. *Personal Identity* (2nd edition). London: Routledge.

Marks, C. E. 1981. *Commissurotomy, Consciousness & Unity of Mind*. Cambridge: The MIT Press.

Olson, E. T. 1997. Was I Ever a Fetus? *Philosophy and Phenomenological Research* 58,

95–109.

Parfit, D. 1984. *Reasons and Persons*, Oxford: Oxford University Press.

Schechtman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.

Shoemaker, D. 2009. *Personal Identity and Ethics: A Brief Introduction*. Toronto: Broadview Press.

Shoemaker, S. 1970. Persons and Their Pasts. *American Philosophical Quarterly* 7, 269–285.

Sperry, R. W. 1968. Mental Unity following Surgical Disconnection of the Cerebral Hemispheres. *Harvey Lectures* 62, 723–733.

Swinburne, R. 1984. Personal Identity: The Dualist Theory, in S. Shoemaker & R. Swinburne, *Personal Identity*. Oxford: Blackwell.

Unger, P. 1990. *Identity, Consciousness and Value*. Oxford: Oxford University Press.

Unger, P. 1992a. Précis of identity, Consciousness and Value. *Philosophy and Phenomenological Research* 52, 133–137.

Unger, P. 1992b. A Reply to Reviewers. *Philosophy and Phenomenological Research* 52, 159–176.

White, S. 1989. Metapsychological Relativism and the Self. *Journal of Philosophy* 86, 298–323.

Wiggins, D. 1980. *Sameness and Substance*. Oxford: Blackwell.

Williams, B. 1970. The Self and the Future. *Philosophical Review* 79, 161–180.