1-1-2015

Personal Identity, Survival And What Matters

James Alexander Gromak
Wayne State University,

PERSONAL IDENTITY, SURVIVAL AND WHAT MATTERS

by

JAMES ALEXANDER GROMAK

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2015

MAJOR: PHILOSOPHY

Approved By:

_____

Advisor Date

_____

_____

_____

DEDICATION

For my parents.

Thank you for nurturing my philosophical nature from an early age.

ACKNOWLEDGMENTS

My thanks to the members of my committee Bruce Russell, Lawrence Lombard, Sean Stidd and Brad Roth for the insightful suggestions and criticisms they offered throughout this project. Their guidance has made this dissertation much stronger than it would have otherwise been. I am especially grateful to Lawrence Lombard for his unwavering determination and belief in my abilities to complete this work. I also owe a great deal to the other instructors at Wayne State University with whom I was fortunate enough to have classes. In addition, Stevens Wandmacher, L. Nathan Oaklander, and Simon Cushing from The University of Michigan – Flint were extremely influential in my early philosophical education. I would not be here if it were not for them. Finally, I wish to extend my appreciation to my friends and family for their continued support, reassurance, and encouragement.

TABLE OF CONTENTS

CHAPTER 1 - PERSONHOOD AND PERSONAL IDENTITY

It will be useful to begin with an extensive introduction to personhood in general. Like many other terms, philosophers oftentimes use the word 'person' differently from the colloquial use of the word. This colloquial use is usually meant to be singular for 'people', or to mean 'a human being'. Although philosophers do use the word in these ways, they also use it in yet another way. Our first question, then, I will call The Personhood Question: "What is it to be a person?" That is, what makes persons different from non-persons? What do persons have that non-persons do not have? Are human beings the only candidates for persons or are there (or could there be) nonhuman persons?

Many philosophers throughout history have discussed these questions and suggested answers to them. A common trend from the Early Modern Period of Western philosophy (specifically Descartes and Locke) was to favor the mental aspect as essential to personhood. For Descartes, you are your mind. That is, you are an immaterial substance that thinks. The Cartesian view of personhood is thus associated with the person's soul. Locke similarly described a person as, "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places."[1] According to him, you are a conscious being that persists by means of continued consciousness and memory. For both Descartes and Locke, a person is a conscious agent capable of interacting with and experiencing the world and generating plans or desires upon which to act.

In "The Concept of a Person", Charles Taylor offers his understanding of personhood. He begins by giving an overview of what kind of thing a person is. He says,

---

[1] Locke, John. *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press (original work, 2nd ed., first published 1694); partly reprinted in Perry 1975.

> Where it is more than simply a synonym for 'human being', 'person' figures primarily in moral and legal discourse. A person is a being with a certain moral status, or a bearer of rights. But underlying the moral status, as its condition, are certain capacities. A person is a being who has a sense of self, has a notion of the future and the past, can hold values, make choices; in short, can adopt life-plans. At least, a person must be the kind of being who is in principle capable of all this, however damaged these capacities may be in practice.[2]

Taylor then addresses two views of what it is to be a person, the first of which, he claims, derives from the ones described in the previous paragraph. He says this first view is

> …rooted in the seventeenth-century, epistemologically grounded notion of the subject [personhood]. A person is a being with consciousness, where consciousness is seen as a power to frame representations of things…They [Persons] have the wherewithal to reply when addressed, because they respond out of their own representation of the world and their situation.[3]

Taylor is explaining here that one view of 'person' originates from the historical, epistemological tradition to designate an agent who (a) possesses consciousness and (b) is able to generate representations of the world as well as respond to such representations. He goes on to say,

> What this view takes as relatively unproblematic is the nature of agency. The important boundary is that between persons and other agents, the one marked by consciousness. The boundary between agents and mere things is not recognized as important at all, and is not seen as reflecting a qualitative distinction. This was so at the very beginning, where Descartes saw animals as complex machines…Proponents of this first view tend to assume that some reductive account of living beings will be forthcoming. What marks out agents from other things tends to be identified by a performance criterion: animals somehow maintain and reproduce themselves through a wide variety of circumstances. They show highly complex adaptive behavior.[4]

Taylor is noting that this historical/epistemological use relies mainly, if not entirely, on a performance criterion to determine what an agent is, and hence, what a person is.

---

[2] Taylor, Charles. "The Concept of a Person", *Philosophical Papers. Volume 1*. Cambridge: Cambridge University Press, 1985, p. 97.
[3] Ibid., 98.
[4] Ibid., 98.

Although Taylor does not give an explicit definition of ‘performance criterion’, we ought to be able to formulate a general idea of what he means. It appears this performance criterion is that the agent is able to form and act upon these representations of the world. Such a view of personhood, then, is that to be a person is to be an agent who can perform or act in a certain way (hence, performance criterion). In other words, a being is a person if it can interact with the world in relevant ways based on the representations it forms. It seems that consciousness would be a prerequisite for such a view since it may be difficult to imagine a being that is able to interact with the world based on representations that it forms without its being conscious. However, I believe that Taylor is not actually addressing a performance criterion that includes consciousness; I think he is addressing only a pure performance view. My reasoning for this interpretation will soon be made clear.

Taylor’s problem with this first view is that such a criterion will include too many things as persons. For instance, animals (even the ones we do not want to qualify as persons) interact with the world, respond to stimuli, and exhibit behavior that may be comparable to persons. Even more problematic, machines can be made to function in very complex ways and respond to input and generate a corresponding output similar to that of persons. In short, Taylor’s fear is that by endorsing a performance criterion for personhood, nonperson animals as well as computers could not be distinguished from persons. Since part of Taylor’s concern is that the performance criterion includes machines, even though machines are not conscious, Taylor is addressing only a pure performance criterion as being sufficient for persons.

Taylor then introduces the second view, which he favors, of what it is to be a

person. It is a significance-based view, expressed in the following passage,

> What is crucial about agents is that things matter to them. We thus cannot simply identify agents by a performance criterion, nor assimilate animals to machines. To say things matter to agents is to say that we can attribute purposes, desires, aversions to them in a strong original sense…[likewise] there are matters of significance for human beings which are peculiarly human, and have no analogue with animals.[5]

These attributions of purpose are an important aspect of what Taylor believes draws the line between the two concepts of person that have been discussed. As he says, "if you take it seriously, then you can no longer accept a [pure][6] performance criterion for agency, because some agent's performances can be matched derivatively on machines."[7] In other words, if one takes the importance of purposes seriously, the first concept of person will not do. Hence, Taylor favors the second concept that places heavy importance on matters of significance. He sums up his position well at the end of the first section of "The Concept of a Person":

> Consciousness…cannot be understood simply as the power to frame representations, but also what enables us to be open to these human concerns. Our consciousness is somehow constitutive of these matters of significance, and does not just enable us to depict them…The centre [of personhood] is no longer the power to plan, but rather the openness to certain matters of significance. This is now what is essential to personal agency.[8]

Interestingly, there is some overlap in Taylor's original overview of what kind of thing a person is with Mary Anne Warren's account. Although she does not give a precise definition of what it is to be a person, she gives a list of five criteria for personhood. They are as follows:[9]

---

[5] Ibid., 98-102.
[6] My addition to indicate that he is only addressing a view that does not include consciousness.
[7] Ibid., 98.
[8] Ibid., 104-105.
[9] Warren, Mary Anne. *On the Moral and Legal Status of Abortion*. Monist 57:1 : 43-61, 1973. Reprinted in Mappes and DeGrazia 2001, pp. 456-463.

1. Consciousness (of objects and events external and/or internal to the being; in particular, the capacity to feel pain)
2. Reasoning (the developed capacity to solve new and relatively complex problems)
3. Self-motivated activity (activity which is relatively independent of either genetic or direct external control)
4. The capacity to communicate, by whatever means, messages of an indefinite variety of types, that is, not just with an indefinite number of possible contents, but on indefinitely many possible topics
5. The presence of self-concepts, and self-awareness, either individual or racial, or both.

She does not say a being must possess all five to be a person, but she does think a being needs most of them. She leaves it very open ended, however, as to which ones are necessary and admits there may be multiple combinations that qualify as being as a person. It is interesting how closely her criterion matches how Taylor initially described a person. Recall that Taylor, at first, claimed,

> A person is a being who has a sense of self, has a notion of the future and the past, can hold values, make choices; in short, can adopt life-plans. At least, a person must be the kind of being who is in principle capable of all this, however damaged these capacities may be in practice.[10]

Warren's fifth condition is found in Taylor's view when he refers to having a 'sense of self'. Her self-motivated activity and reasoning conditions may be construed as similar to Taylor's requirement for making choices and decisions that are in the person's interests. And finally, her consciousness condition certainly plays an important role in Taylor's later development of his significance-based view.

Although Taylor and Warren's views are quite similar to each other, not all views of personhood are so alike. Consider Peter Strawson's very influential view. According to Strawson, the concept of a person is fundamental and irreducible. He states that persons are particulars with consciousness; they have physical characteristics (being located within space and time, having a certain height, color, shape, weight, etc.) as well as

[10] Taylor, 97.

mental characteristics (thinking, remembering, seeing, deciding, feeling emotions, etc.). But the concept of person is primitive, or simple, in the sense that it cannot be further analyzed. As he explains, "The concept of a person is the concept of a type of entity, such that both predicates ascribing states of consciousness and predicates ascribing corporeal characteristics, a physical situation [etc.] are equally applicable to a single individual of that single type."[11] Strawson is saying the physical and mental aspects can only be attributed to persons; and it is this, as opposed to just a Cartesian mind or a Lockean consciousness, that makes you who and what you are.

Although the philosophers mentioned above have given some indication as to what they find important in determining an answer to The Personhood Question, they have not given a clear, precise analysis of 'person'. Warren perhaps comes closest to doing so by providing her criteria; but she herself says not all of the five conditions are necessary. This leaves us uncertain as to where to draw the line, how much of each of the five will suffice, and so on. Among the accounts above, common elements emerge such as the presence of consciousness and a sense of self, as well as things mattering or being important to persons, and the ability to form and act upon life-plans. All of this provides a rough, "ballpark" idea of what it is to be a person, but we still do not have a clear-cut definition.

Harry Frankfurt builds upon the above notions of personhood and attempts to give a strict definitional answer to our first question: what is it to be a person? He acknowledges there are creatures that have psychological and material properties yet are not persons in any normal sense of the word. According to him, then, some animals,

[11] Strawson, Peter. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959, pp. 101-102.

including some members of the human species, would not be persons. Yet he is not committed to saying that the conception of personhood must be species specific. He says that we generally acknowledge only members of the human race to be persons, but conceptually there could be other forms of intelligent life that would be accurately classified as persons also. But for our purposes, we will focus on human personhood. The question now becomes: what key feature(s) separates creatures from persons and nonpersons? For Frankfurt, the key difference is the structure of the creature's will. He concedes that humans are not the only beings capable of having desires and making choices; it seems that some creatures even deliberate prior to acting on a desire. These desires to do or not to do something he refers to as "first-order desires". However, what is unique to persons is the ability to form what Frankfurt calls "second-order desires."[12]

While the object of a first-order desire may be some object, action, or mental state, the object of a second-order desire is a first-order desire. For instance, a person may have a first-order desire to do X, yet he may wish he did not have the desire to do X; so he would have a second-order desire to not want to do X. Likewise, a person may not have a desire for Y, but wish he or she did. This person would have a second-order desire to desire Y.[13] Frankfurt notes that no beings other than humans seem to have this self-reflecting ability to form such second-order desires. One may object that the cause of having such second-order desires need not be self-reflection. One may claim that one could be brainwashed into wanting to want heroin, for example.[14] Although it is true that

---

[12] Frankfurt, Harry G. The Importance Of What We Care About. *Freedom Of The Will And The Concept Of A Person.* New York, NY: Cambridge University Press, 13th Printing 2007, pp. 11-12.

[13] Wanting to satisfy one first-order desire before satisfying some other conflicting first-order desire is another kind of second-order desire. This will be addressed in the next paragraph where I discuss what Frankfurt calls second-order volitions.

[14] This objection was raised by Bruce Russell.

one could be brainwashed into having certain second-order desires that one would not otherwise have, this misses the point. Frankfurt is not saying that the only way we could come to possess second-order desires is by self-reflection; it is the fact that we have this ability to reflect and form higher-order desires *at all* that is important. That is, the fact that one has the ability to self-reflect on one's first-order desires and form higher-order ones upon this reflection is what is relevant to being a person according to Frankfurt.

Frankfurt then specifies that when one has a second-order desire that a certain first-order desire be one's will, then this individual has a "second-order volition." As Frankfurt explains, "people characteristically have second-order desires concerning what first-order desires they want, and they have second-order volitions concerning which first-order desire they want to be their will."[15] The distinction between second-order desires and second-order volitions is as follows: a second-order *desire* is nothing more than a desire to *have* (or not have) a certain first-order desire, whereas a second-order *volition* is a desire to *act on* (or not act on) a certain first-order desire. This is what Frankfurt means in the previous passage when he discusses someone who wants a first-order desire to be one's will. It simply means that the person wants to act upon their corresponding first-order desire and in virtue of that, has a second-order volition. It is the ability to have these second-order volitions that is the essential feature of personhood for Frankfurt and serves as the foundation for what he calls "the self."[16]

There is an objection that may be raised against Frankfurt's position that I must address. This objection, again, concerns a person that has been brainwashed.[17] Suppose a person has been brainwashed into wanting his desire to take heroin to become his will.

[15] Ibid., *Identification And Wholeheartedness,* 164.
[16] We will see the complete formulation of this Frankfurtian account of the self later on.
[17] It was, again, brought to my attention by Bruce Russell.

This is supposed to be a counterexample to Frankfurt's view that what it is to be a person relies on one's second-order volitions. Just as with the former objection regarding a brainwashed person, I find this objection to be misplaced. When Frankfurt says the essential feature of personhood lies in our higher-order desires, I do not think we should interpret that as, "whatever higher-order volitions we happen to have no matter how we happen to get them." I think that what he really means is, "the higher-order volitions we generate as a result of active self-reflecting on our set of first-order desires." It seems wrong to dismiss the means by which we obtained these higher-order desires as unimportant. Frankfurt's whole point about what makes persons special, or different from non-persons, is that we have the ability to *generate* higher-order desires, not that we merely *have* them. Only if we interpret him in the latter sense would he fall victim to the brainwashing counterexamples. But since the importance of higher-order desires, as it relates to personhood, rests on the forming of such desires being *of our own doing*, I think we can avoid these objections. It is our ability to look inward on ourselves, reflect on our desires, and as a result, generate higher-order desires based on that act of self-reflection that is important to being a person. The reason we have these higher-order desires as a direct product of our own self-reflection is what makes these higher-order desires genuinely ours. It is these desires that are relevant to being a self and not ones that have been implanted via brainwashing or any other external means. The reason for this, of course, is simply that the former are products of our own doing while the latter ones are not.

There is another issue I wish to address before continuing. One may wonder if Frankfurt's account of personhood is just another performance criterion and, hence,

subject to Taylor's objection to such a view of personhood. Recall that Taylor's concern was that holding a performance criterion amounts to saying that what it is to be a person lies in being able to act and interact with the world in a certain way. I do not believe Frankfurt's view is merely a performance criterion in this sense, because forming second-order volitions amounts to much more than acting and interacting with the world. Being able to generate higher-order volitions via self-reflection requires a creature to be capable of differentiating among its lower-level desires and, moreover, having the ability to decide which one of those it wants to be his or her will. Clearly this requires that the creature be capable of self-reflection and internal deliberation that far exceeds mere interaction with the external world.

Even if one insists that Frankfurt is suggesting a performance criterion, it certainly does not have the consequences claimed by Taylor since it is not a *pure* performance criterion. Taylor's fear, remember, was that such a criterion would include too many things as persons, since even single-celled organisms like amoeba are capable of interacting with the world and responding to stimuli. However, these beings are not conscious in the same sense that persons are. Moreover, it ought to be very clear from what Frankfurt has said that his view does not commit him to such things being persons. Being able to generate higher-order volitions amounts to a creature deciding that it wants a certain lower-level desire to be his or her will; no simple organism is able to do this. Frankfurt acknowledges that some animals are capable of forming lower-level desires but the forming of higher-order volitions is something they cannot do. Frankfurt may even have an edge over Taylor since Taylor seems to be saying that the property of

personhood is uniquely human[18]. This implies that the only persons there are or even could be are human persons. This is implausible. It is reasonable to think that sufficiently intelligent aliens could count as persons. Frankfurt's view, although it is concerned mainly with human persons, is open to the possibility of nonhuman persons.

Now that we have had an introduction to what it is to be a person, we may now extend our inquiry to the core questions regarding personal identity (PI) that will be considered throughout the dissertation. We may begin with The Personal Identity Question. Where The Personhood Question simply asks what it means to be *a* person, The Personal Identity Question asks what conditions must be satisfied for person X and person Y to be the *same* person. From here, we may consider a closely related question, The Persistence Question: What accounts for identity through time? That is, what does it take for a person–the *same* person–to persist? Consider statements like, "when *I* was five years old, *I* did x". This sentence would be false, if persons did not persist through time, since the first 'I' in that sentence would not refer to the same being as the second 'I'.

Before we even begin thinking about answers to these questions, we must become clear on precisely what each of them concerns. To help clarify the distinctions, let us formalize the questions to specify precisely what each are asking:

> The Personhood Question: Under what conditions is X a person?
>
> The Personal Identity Question: Suppose X and Y are persons. Under what conditions X = Y?
>
> The Persistence Question: Under what conditions does person X persist?

[18] Remember that Taylor said, "there are matters of significance for human beings which are peculiarly human, and have no analogue with animals." Also recall that Taylor found these attributions of purpose to be the important aspect that draws the line between the two concepts of person that he discussed. Moreover, he favored the second concept that places heavy importance on matters of significance. If he believes these matters of significance are "peculiarly human, and have no analogue with animals", then he seems to be implying that personhood is unique to humans.

As previously stated, The Personhood Question simply asks what distinguishes persons from non-persons. The Personal Identity Question however, deals with under what conditions person A and person B are numerically identical. The Persistence Question concerns this also but applies it across time. That is, it asks what conditions must be satisfied in order for a person to persist through time. It should be clear that this question is not about qualitative identity, but rather numerical identity. To say person X and person Y are numerically identical is to say that they are literally the same person, i.e., there is only one person and not two. Two numerically distinct things are qualitatively identical if they are exactly similar. Since The Persistence Question is about numerical identity, it does not follow that a person must remain qualitatively identical at all times in order to remain the same person.[19] Since persons change (they get bigger or smaller, gain new experiences and forget past ones, etc.), the Persistence Question does not ask what it takes for a person in the future or the past to be merely qualitatively just like you, but what it takes for that being to actually *be* you, or to be numerically identical with you as opposed to being a person that is *not* you, or numerically distinct from you. That is, the Persistence Question asks what changes a person can undergo without ceasing to exist.

With this understanding of The Persistence Question, it is reasonable to ask: how much of a change could one survive? In addressing this question, we must choose our words carefully. People often say things like "you're a totally different person" when they see someone they have not seen in a long time. Or someone may say, "I'm not the person I used to be," perhaps after a religious conversion of being "born again." These expressions cannot be taken literally. When we reunite with a friend whom we have not

[19] There are, however, some philosophers that believe PI is not a question of numerical identity. They conclude that all changes to a person result in a numerically distinct person.

seen for ten years, we do not literally mean they are a numerically different person. We simply mean that the person has changed in some significant way from how the person used to be.

That is not to say that a person could never, under any circumstances, cease to be the same person after some radical change. It is certainly possible that if certain things were to happen to you that drastically change, in the relevant ways (whatever they may be) who you are, the resulting person after these changes may be someone else; that is, *you* may not be able to survive these changes, and have thus ceased to exist. Derek Parfit addresses this when he says, "…certain kinds of qualitative change destroy numerical identity. If certain things happen to me, the truth might not be that I become a very different person. The truth might be that I cease to exist–that the resulting person is someone else."[20] This would mean that the changes person X underwent were so substantial that X could not persist through the change. That is, X has ceased to be and some new person, call him or her Y, now exists.

But this is not normally what we mean when we say things like, "you're a totally different person." We simply mean that the person has changed in some relevant or important way from the way the person used to be, but not so drastically that the person is literally no longer the same person. This is our concern when we ask how much change one could survive. It is this issue that Parfit addresses in the following passage:

> We might say, of someone, 'After his accident, he is no longer the same person'…We claim that *he*, the same person, is *not* now the same person. This is not a contradiction. We merely mean that this person's character has changed. This numerically identical person is now qualitatively different…I may believe that, after my marriage, I shall not be the same person. But this does not make

[20] Parfit, Derek. *Reasons And Persons*, New York, NY: Oxford University Press, 1986, p. 202.

> marriage death. However much I change, I shall still be alive if there will be some person living who will *be* me.[21]

What must constitute PI that allows a person to persist through these changes? Although there are many views that try to answer this question, most of them can be grouped into two types. The first type places the ultimate importance upon the body (or at least some part of it, namely the brain) whereas the other places it upon the psychological aspects of a person.

Body-theorists believe that PI amounts to the continuity of one's body (or at least a part of it, namely the brain). Thus a body-theorist's answer to The Personal Identity Question would be X = Y just in case there is bodily (brain) continuity between X and Y. It is easy to see how this type of view would answer The Persistence Question: one persists through the continued existence of one's body (brain). This means that there must be bodily (brain) continuity in order for the person to persist through time.

Mind-theorists on the other hand believe that one's PI amounts to the continuity of one's mind, or at least some psychological feature of it (those often pointed to being consciousness and memory). Mind-theorists answer The Personal Identity Question by claiming that X = Y just in case there is mind continuity between X and Y. They answer The Persistence Question as follows: one persists through the continued existence of one's mind in the sense of psychological continuity of consciousness or memory.

Rather than exploring these two types of views of PI and determining which is likely to be correct, it will be more useful to take a step back and consider things from a different perspective. Since the entire discussion of PI revolves around the identity of a person (whether it be at a single time or at different times) it is difficult to address these

[21] Ibid., 201-202.

issues without presupposing that identity is maintained. I would like to propose an alternative approach to discussing the topic of PI (at least initially). This alternative approach is from the perspective of what I shall call '*continuance*'. I will use 'continuance' to refer to some kind of 'continuing life'[22] that is embodied in some person or persons. The term will be used as a neutral term for discussing the continuity of a person without any implications of identity. Moreover, this continuing life is to be understood abstractly. Consider the following passage regarding the president's changing parties: "In 1965, The President was a Democrat. In 1975, The President was a Republican. Oh my God! The President changed parties!"[23] Of course, the conclusion being derived is false. We are not to understand 'The President' as referring to one person. Rather, 'The President' refers to an abstract entity. This is how I shall use the term 'continuance'. In some cases, continuance will refer to a continuing life of one person but in other cases it will refer to a continuing life of more than one person. In the literature on PI up until now, there has been no such neutral term. As a result, when considering the various cases in the literature, it is difficult to talk about the resulting person in those cases without presupposing that he is identical with the original person. The use of the term 'continuance' will allow us to talk of the resulting person without such presuppositions. This will be further explained and become clearer once we consider a couple examples.

There is a question lurking here: Is identity/persistence just one kind of continuance, or is it that the only way for there to be continuance is for some one thing to persist? Initially, it may be easier to grasp this distinction (and its importance) with an

[22] What kind of continuing life will be addressed later.
[23] My thanks to Richard Sharvey through Lawrence Lombard for this example.

example concerning an inanimate object rather than a person.[24] Suppose we take a buzz saw and slice a table directly down the middle cutting it into two halves. Now suppose the left half re-grows a new right half and the right half re-grows a new left half. We now have two tables. Is either of them identical with the original table? My answer is 'no'. To show why, let me present a couple of similar examples.

Suppose, as before, we slice a table directly down the middle cutting it into two halves. But now, suppose we destroy one half. The remaining half re-grows the rest of a table. Some may think this table is identical with the original. But I do not. The reason for this is neither half is a table. In order for there to be continuity of a table, the object must be a table at every moment. But since half a table is not a table, there is discontinuity of existence between the tables; hence, the table that results after the re-growth cannot be the same table as the original one. There is, however, continuance between the resulting table and the original since the resulting table grew from part of the original table.

In order to further persuade you, consider the following extreme cases. If only a quarter of a table is cut off and destroyed (such that the table's legs are still intact, it still functions as a table, etc.) and the remaining three quarters grows a new quarter, would this resulting table be identical with the original? Yes, since the object is a table at every stage, it is at least possible for it to be the same table. Moreover, there is also continuance in this case. Next, consider a similar example: Suppose we cut off one leg of a table and then destroy the rest of it. Out of the severed leg, the rest of a table grows. It should be clear this is not the same table as the original. Again, the reason for this is that a table leg is not a table. Despite there not being continuity between the tables, there is continuance; once again, because the resulting table grew from the leg of the original one.

[24] This strategy was suggested to me by Lawrence Lombard.

Return to the first case where we slice a table down the middle and each half re-grows another half. I asked whether or not either of the resulting tables are identical with the original. We can now see why my answer was 'no'. Since neither half of the table is a table, it is impossible for either of the resulting tables to be identical with the original one[25]; the original table has gone out of existence once it was cut in half. But note that there is continuance between each of the two tables with the original one since the two resulting tables began from one half of the original table.

The answer to the lurking question should now be easily answerable. Identity/persistence is just one kind of continuance. In other words, there can be continuance with or without identity. This has been demonstrated through the above table examples. In all of these examples, there was continuance, but there was only continuity in the cases where the table persisted. This is what I had in mind when I said that the term is neutral in regard to identity. It allows us to use the term to talk about such cases without any implications as to whether or not the object persists. These examples about physical objects will serve as a starting point for later on when we consider similar fission/split-brain cases involving persons.

In order to have an adequate theory of PI, we must first have an adequate theory of continuance. This is because the continuance of *a* person is necessary (though not sufficient) in order to have the persistence of that person. My first task then, is to consider

[25] I am here assuming Locke's claim that "one thing cannot have two beginnings of existence," as found in John Locke's *An Essay Concerning Human Understanding*, book 2, chapter 27. This principle is, of course controversial, and will not work for every physical object. For example, consider a watch that has been taken apart to be cleaned. When it is taken apart, the watch does not exist. But once it is put back together, I believe that is the same watch as the one in need of cleaning. So here is a case where a physical object existed, temporarily did not, and then existed again. Hence, Locke's principle is false in this case. I am not endorsing this principle as universally true; I am just assuming its truth for sake of argument in regard to the table to help explain my example about continuance.

competing continuance theories. From there, we may add whatever further stipulation(s) is/are necessary to formulate a theory of PI.

As stated above, most theories of PI can be grouped into two main types: physical and psychological. These theories place the determining factor of whether or not a person persists on physical continuity and psychological continuity, respectively. We can address them in terms of continuance in the same manner. My task in the next three chapters is to determine what kinds of continuance are relevant to persistence. I shall achieve this by considering numerous cases and examining our intuitions regarding whether or not the person persists in those cases. By observing what kinds of continuance are present in cases of persistence and those that are not, it will become evident what kinds of continuance are necessary and sufficient for persistence. I shall begin with physical continuance.

CHAPTER 2 - PHYSICAL CONTINUANCE

As I alluded to in the last chapter, there are several types of physical continuance that must be considered. The first is physical continuance of the body (by which, I mean the entire body). There are many arguments that attempt to show that persons are not identical with their bodies. These arguments may seem to have some intuitive bite. However, by employing the term 'continuance' as described in the previous chapter, we may sift through these examples to show that they still leave the possibility of bodily continuance as a genuine option for persistence.

Consider Franz Kafka's *The Metamorphosis*[26]. In this novella, a man named Gregor Samsa falls asleep in his bed one night. In the morning, a being awakens in that bed possessing the body of a large, cockroach-like insect. We are led to believe that Gregor's body has transformed[27] into this cockroach body. The cause of the transformation is never revealed, but this is not the point. The point, for our purposes, is to determine whether or not the person that awakens with the insect body is Gregor. The rest of the novella deals with the resulting person's attempts to adjust to this new situation. The person feels he is a burden to Gregor's parents and sister, who are repulsed by his condition. He grows more frustrated as he tries to adapt because he remembers (or at least seems to remember) how Gregor's life was before the metamorphosis. Moreover, he is severely depressed since he is unable to satisfy the desires Gregor once had. In short, the entire story operates under the assumption that the person with the insect body *is* Gregor. What makes this so? The entire story bases it on the fact that the person with

[26] Kafka, Franz. *The Metamorphosis.* Kurt Wolff Verlag, Leipzig, 1915.
[27] I wish to emphasize that we are considering a case involving the literal *transformation* of body, not whole bodily replacement. That is, we are supposing that Gregor's original human body has *transformed* into that of cockroach body. His human body was not merely *replaced* by the cockroach body.

the insect body has Gregor's entire psychology. This story assumes that no matter what happens to one's body, the survival of the person lies in psychological continuity.

If the title of the novella is any indication as to how the person in the insect body came to be in such a body, we may assume that there is physical continuance here. Despite the fact that the resulting body is not even human, since the human body *transformed into* (since that's what it means to metamorphose) the insect body, there actually is bodily continuance. So this example does not show what it needs to show in order to rule out bodily continuance as a possibility for being necessary or sufficient for persistence of the person. In order to rule out bodily continuance, we would need a case where there is not bodily continuance but there is persistence.

Lynne Rudder Baker[28] supplies us with a similar example: Suppose a person's organic parts were gradually replaced by non-organic parts. Suppose further that this replacement is completed in such a way that the person's psychological life is continuous. She thinks that in this case, "the person would persist but the organism would not" and that "(t)he possibility of cases like these rules out identification of a person with the organism that is her body."[29] As with the previous example, there is certainly physical continuance here. Although the organic parts are being replaced with non-organic parts, there is still continuance of body. Moreover, there is still the continuation of a life–indeed, the same life–within this body. Hence, this example, as did the previous one, fails to demonstrate that bodily continuance is neither necessary nor sufficient for persistence.

Although the threats posed by the cases just considered were avoided, I think there is a case that does show that bodily continuance is neither necessary nor sufficient

[28] Baker, Lynne Rudder. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press, 2000.
[29] Ibid., 19.

for persistence. The example is an updated or modernized version of John Locke's[30] classic example involving the prince and the cobbler. I say "an updated or modernized version" because, in the original example, Locke asks us to imagine that the soul (by which he means consciousness) of a prince enters and animates the body of a cobbler. Let us suppose that, rather than just the soul moving from the prince to the cobbler, we swap their brains. So the brain of the prince is in the body of the cobbler and vice versa. In this thought experiment, the person in the cobbler's body possesses all of the memories, thoughts, beliefs, desires, etc., of the prince and vice versa. Most have the intuition that the prince and the cobbler have simply swapped bodies. That is, the prince now has the body that used to be the cobbler's and the cobbler now has the body that used to be prince's.

If this is right, then both the prince and the cobbler have persisted through the brain swap. Notice also that there is not bodily continuance for either the prince or the cobbler. Although it is true that each of them still has *a* body, there is not continuance because the original body each person had is distinct from the body they currently have. Remember that "continuance" is used to refer to some kind of continuing life. Moreover, the term will sometimes refer to a continuing life of one person but in other cases it will refer to a continuing life of some other person. In this case, the continuance of the prince's body is the continuing life of two persons. Before the brain swap, that life was the life of the prince but after the brain swap, it was the life of the cobbler. A similar explanation can be given of the cobbler's body. So if it is possible for the prince and the cobbler to persist with bodies different from the ones they initially had, and there is

[30] Locke, John. "Of Identity and Diversity" in *Essay Concerning Human Understanding,* (first published 1694); reprinted in Perry 1975, pp. 33–52.

clearly no continuance between the body either of them had before the swap with the one they had after, then bodily continuance must not be necessary for persistence of the person.

In addition, this also shows that bodily continuance is not sufficient for persistence. For if it were, then the continuance of the prince's body should entail the persistence of the prince with that body. In other words, if bodily continuance were sufficient for persistence, and since there is the continuation of a life in the prince's body, then the life should be that of the prince. But it is not; hence, bodily continuance must not be sufficient for persistence either.

The existence of two resulting persons may make the conclusion less clear than if there were only one resulting person. So let us modify Locke's example further. Rather than swapping brains, which results in two persons, suppose we removed both the prince's and the cobbler's brains. We then dispose of the prince's body and the cobbler's brain and transfer the prince's brain into the cobbler's body. This is similar to the modified Lockean example given above except here we only end up with one person rather than two. As before, most will say that the resulting person is the prince in the body of the cobbler. Also as before, there is not bodily continuance. Although it is true that the prince still has *a* body, there is not continuance between his original body and the one he currently has since they are distinct bodies. Also note that there is no continuance of the prince's body since it was destroyed. Yet the prince still exists despite the lack of his original body. This shows that bodily continuance is not necessary for persistence. Likewise, there is continuance of the cobbler's body yet the cobbler no longer exists. Hence, bodily continuance is not sufficient for persistence either.

Although I find the above modification to Locke's example to be quite conclusive, there is an issue that I feel should be addressed. How are we to account for our temptation to say that a person does persist solely as a result of having the same body? For instance, suppose a person's psychology changed quickly and radically[31] due to a mad scientist altering his brain. Or in an even more extreme case, imagine that the mad scientist erases the entire content of the person's mind (memories, beliefs, desires, etc.). It may be tempting to say that the person persists in these cases. We can imagine the resulting person being given a series of photographs of "him" from the distant past up to the present allowing the resulting person to see that his body is the same as that of the person depicted in the photos. However, the basis for saying this lies in the fact that the resulting person has the same body. But as the modified Lockean example has already shown, not even the continuity of an entire body is sufficient for persistence. So although it may be tempting to say that the resulting person is the same as the original, we are being misled into thinking so simply because so often 'same body' is connected to 'same person'.

There is a comparable real life scenario that may illustrate the point better. Consider those whose mental faculties have greatly deteriorated as a result of advanced stages of dementia or Alzheimer's disease. Let us stipulate that there is no psychological continuance at all. In these cases, bodily continuance does seem to provide some evidence for survival. This is because, as I said above, sameness of person so often goes hand-in-hand with bodily continuance. That is, we see the body of an individual in an advanced stage of dementia or Alzheimer's disease and still recognize that individual as

[31] I say "quickly and radically" because it may be the case that a person can survive radical changes to one's psychology provided that they occur over a sufficient length of time. This will be discussed in more detail in a later chapter.

being our loved one. We do this because we are so used to associating that body with the person since physical continuance, psychological continuance and persistence so often go together. But I believe we are simply being misled. If the psychological aspects are diminished enough, I do not believe the person persists despite bodily continuance.

Now, it is true that in some cases of dementia or Alzheimer's disease there may still be psychological continuance. For example, the person may not be able to remember what happened, say, five minutes ago but he may be able to remember what happened four minutes ago. Furthermore, that person could remember what happened four minutes before that and so on. In a case like this, it may be correct that the person does persist. But here, there is psychological continuance as well as physical continuance and I believe the reason we are inclined to believe that the person persists is due to the psychological continuance.

My point is to show that bodily continuance alone is not sufficient for sameness of person. If it were, then even cases of dementia or Alzheimer's disease where there is no psychological continuance we would say the person persists. This seems false, so bodily continuance alone must not be sufficient for persistence.

We have thus ruled out the possibility of the first kind of physical continuance (whole body continuance) as being either necessary or sufficient for persistence. The second kind of physical continuance that may now be considered is that of the brain. There are two types of brain continuance that may be considered: continuance of the whole brain and continuance of part of the brain.

Let us address the whole brain first. As I just argued, the whole body (which obviously includes the whole brain) is not sufficient for persistence. If the whole body

including the brain is not sufficient then certainly the whole brain by itself is not sufficient either. Indeed, it would be strange if the whole brain were sufficient for persistence but the whole body, which includes the whole brain, was not. If this were true, it would mean that continuance of the whole brain alone results in the persistence of the person but if there is bodily continuance in addition, then the person might not persist. That the addition of bodily continuance to brain continuance could potentially destroy one's chances for survival is very bizarre, especially since, as I pointed out above, oftentimes bodily continuance and brain continuance go together. Hence, it must be the case that if the whole body is not sufficient, then the whole brain is not sufficient either.

Along a similar line of argument, if the whole brain is not sufficient for persistence, then it certainly follows that part of the brain is not sufficient either. It would be odd if part of the brain was sufficient for persistence but the whole was not. This would mean that, in a split-brain case, if the entire brain was transplanted, then the person may not persist (depending on what may be done to the psychology of that brain); but if only part of the brain were transplanted (regardless of what may be done to the psychology), then the person definitely would. This is a very difficult notion to understand because it seems that if two things together are not conceptually sufficient for persistence, then neither can either of those two things alone be sufficient. Hence it must be the case that if the whole brain is not sufficient for persistence, then part of the brain is not sufficient either.

Let us now consider whether continuance of the whole brain is necessary. Although there are some cases in the literature that seem to show that *identity* of the whole brain is not necessary, this does not mean that they show *continuance* of the brain

is not necessary. It is important to note that the success of the arguments against continuity does not result in the success of those against continuance. An example of such an argument is given by Frances Kamm[32] when she considers possible situations in which a patient with dementia recovers. Suppose a patient in an advanced stage of dementia recovers by being given a treatment that allows her body to grow a large number of brain cells to replace ones destroyed in such a way that is not consistent with physical continuity.[33] Kamm argues that one who thinks continuity of brain is necessary for survival would be forced to say that the patient does not survive. This is not the answer most intuitively have. After all, there is nothing extremely hypothetical about the scenario. The treatment simply allowed the patient to rejuvenate their cells in bulk as opposed to one-by-one as it normally and naturally would. Since the process is the same, just hastened, there seems to be no reason why one would conclude that the patient, who is now out of the hospital and functioning normally, has not survived. She concludes from her example that brain continuity is not necessary for persistence.

Notice, however, that although there is not continuity in her case, there certainly is continuance. In the same way that there was continuance of the original table after growing another half, so too is there continuance of the brain in Kamm's example since each of the new brain cells replaced the destroyed ones. So although Kamm's example may have shown that *continuity* of the brain is not necessary for persistence, it does not show that *continuance* of the brain is not necessary for persistence.

---

[32] Kamm, Frances. "Jeff McMahan's Ethics of Killing: Problems at the Margins of Life." *Philosophical Review* 116, 2007 (2), pp. 273–280.

[33] Note that most advocates of physical continuity views hold that cellular replacement (whether it be by the natural process of cell replication or by some procedure) allows for physical continuity only when done in small amounts at a time. On their view, if all, or a very large portion of cells are replaced at once, they deny there is physical continuity.

Despite this, I believe we can show that continuance of the whole brain is not necessary for persistence. Consider a case similar to the most recent modified Lockean example except, instead of transplanting the whole brain, we split the brain into two hemispheres and only transplant one while destroying the other.[34] We may suppose that all the contents of the person's mind would be housed within the transplanted hemisphere so no psychological features would be lost. If we believed that the resulting person persisted in the previous example, then certainly we ought to believe the same in this example. After all, the only difference between them is how much of the brain is transplanted. But since we are supposing that the entire content of the mind is present in each case, then this difference is negligible; whether there is the entirety of the brain present or only a part plays no role. We may conclude from this that continuance of the whole brain is not necessary for persistence since the above example just demonstrated that persistence can be secured without it.

I conclude from the cases examined thus far that no physical continuance theory alone (neither body, whole brain, nor part of brain) is sufficient for persistence. I also conclude that neither body nor whole brain is necessary. But could part of the brain be necessary? In the cases that have been discussed so far, the continuance of the psychological aspects of the person have been largely what tracks the persistence of that person. In other words, when considering the examples above, what seemed to be doing the majority of the work was the continuance of psychological aspects; physical continuance seemed to be playing virtually no role. So it may be possible that no kind of physical continuance is necessary for persistence. It is this possibility I now wish to explore.

[34] Split-brain cases (also called fission cases) will be explored in more detail in a later chapter.

In all of the examples so far, there has been continuance of at least part of the brain along with psychological continuance. In the examples to come, I will consider cases where there is not even continuance of part of the brain to see if we may have psychological continuance without it. If we can, and if we believe the person persists in such cases, then I will have shown that not even continuance of part of the brain is necessary for persistence. If this is true, then this will lend at least some support to those who believe a purely psychological view of persistence is correct.

Let us first examine an example from Parfit. 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.[35]

It should be clear that there is continuance of the brain in *Case One*. This is because the surgeon replaces only a hundredth part of the brain until the entire brain has been replaced. Since the new parts coexist with the original ones for a time, then there is brain continuance. But in *Case Two*, there is not continuance of even part of the brain since the entirety of the original brain is removed and then all the replica parts are inserted.

Let us consider the possible responses to this scenario regarding whether or not the person survives: (a) we may say that the person survives in both cases, (b) that the person survives in *Case One* but not *Case Two*, (c) that the person survives in *Case Two* but not *Case One*, or (d) that the person survives in neither. It seems (c) and (d) can be readily dismissed. If this is not immediately clear, it will be easy to show. Note that in both cases, we are supposing that psychological continuance is maintained. As already

[35] Parfit, 474.

stated, in *Case One*, brain continuance is preserved. But in *Case Two*, there is not brain continuance. In light of this, it would indeed be strange to believe option (c). If one were to have this response, it would mean that psychological continuance alone is sufficient for survival but if one has brain continuance in addition, then that somehow destroys the chances for survival. This would certainly be a strange view to hold.

Option (d), although not necessarily strange, does suggest a radically different understanding of persistence than the one this dissertation is addressing. Option (d) is that one does not survive in either case. This would amount to saying it does not matter if physical continuance (of any kind) or psychological continuance is preserved; moreover, even if both are preserved, that is not enough for survival. This type of response would be from someone who believes something entirely different than physical or psychological continuity is necessary for survival. An example of such a condition would be requiring the exact same matter. On such a view, a change in the matter that makes up the brain means a change in the brain, which in turn results in a change of the person. Advocates of this view say that "a person" is actually a series of momentarily different persons. In other words, "a person" is made up of "time slices of persons" that are causally linked across time.

Requiring the exact same matter in order to remain the same person, I believe, is too strong a requirement. In fact, I find it is even too strong for just the identity of physical objects.[36] Suppose one thought that the identity of physical objects requires the identity of matter. This would mean that no physical object could remain the same object after losing one of its parts, even if that part was replaced by a qualitatively identical one. Consider the Ship of Theseus. In this example, we are to imagine the parts of a ship being

[36] Although mereological essentialists hold such a view.

replaced one-by-one over an extended period of time. After the first part is replaced, we are asked if the ship is identical with the original ship. Of course it is. Moreover, it seems wrong to say after the removal of any part that the ship is no longer the same ship as the original one. So we conclude that the ship is still identical with the original ship even after the last original part is replaced. The resulting ship now has all new parts but is still numerically identical with the original ship with all the original parts. This shows that same matter is not necessary for same object. In the same way, requiring the exact same matter of the brain in order for a person to remain the same across time is also too strong a requirement.

We are now left with option (a) and (b). Since each of these options say that the person survives in *Case One*, the difference comes down to whether or not the person survives in *Case Two*. Being committed to holding that the person survives in *Case One* should not be discouraging. After all, the resulting person still has the same body (the brain is not being swapped into a different body as in the modified Lockean example) and although all the parts of the brain have been replaced, they have been replaced in such a way that allows for continuance of the brain as well as retaining the entirety of the original person's psychology. So we end up with a whole brain that is a physical continuant of the original one (as opposed to only a partial continuant as in the split-brain case). Since the intuition was that the person persists in the modified Lockean example as well as in the split-brain case, the same intuition should arise here that the person persists in *Case One*.

Now, what ought we say about *Case Two*? Since the original brain is entirely removed in one hundred parts and is then replaced by one hundred replicated parts, it is

clear that the resulting person's brain is nothing more than a mere duplicate. This is not the same as what occurs in the Ship of Theseus example. In that case, the parts of the ship are replaced one-by-one over time which allows for the ship to assimilate the new parts. But in *Case Two*, the parts are not being replaced; rather, all of the one hundred parts are being removed and then the one hundred duplicates are being placed in. I see this as no different from a person's brain being replicated (such that it contained the same mental contents as the original brain), the original brain being removed and destroyed, and the replicated brain being placed in their empty head. Just as I think we should say this person has not persisted, since his brain is a mere duplicate, neither has the person in Parfit's *Case Two* since his brain is also a mere duplicate.

Although Parfit agrees that the person does not survive in *Case Two*, he does not think this matters. As he says,

> I cannot believe 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. I cannot believe that, if the surgeon alternates removing and inserting, this will be just as good as ordinary survival, while if he does all the removing before all the inserting, this will be nearly as bad as ordinary death.[37]

In this passage, Parfit is not disputing that the person would not survive in *Case 2*; only that what happens to the person in *Case 2* "will be just as good as ordinary survival" and will not be "nearly as bad as ordinary death."[38]

Thus, I endorse option (b) that the person survives in *Case One* but not *Case Two*. Although *Case Two* was one in which there is not continuance of even part of the brain, there is not persistence either. So it does not provide us with an example that shows that continuance of part of the brain is not necessary for survival.

---

[37] Parfit, 476.
[38] I will offer critical analysis of Parfit's view concerning what matters in Chapter Five.

Although the above example by Parfit was not one where he intended to show that continuance of part of the brain is not necessary for survival, Bernard Williams gives an example where he is attempting to show precisely this. Suppose persons A and B undergo a procedure whereby the contents of person A's brain are transferred into person B's brain and the contents of person B's brain are transferred into person A's. After the procedure, the person with A's brain has the psychological features that used to belong to B and vice versa. The resulting 'A-brained/bodied' person will seemingly remember living the life of B, as well as having his beliefs desires, etc. In short, the 'A-brained/bodied' person will have B's mind. And similarly, the 'B-brained/bodied' person will have A's mind. Williams says that our intuitive response is that the 'A-brained/bodied' person is now B and the 'B-brained/bodied' person is now A. If this is our intuitive response, then what is relevant regarding continued existence is not continuance of any part of the brain, but rather pure psychological continuity.

The problem with Williams' account is that it takes the "contents of brains" to be like "marbles in a bag"[39] that can be transferred from one brain/bag to another. To remove the metaphor, Williams is imagining that memories, beliefs, desires, etc. are all objects within a person's brain that can be removed from one brain and transferred into another. But the contents of a person's brain–memories, beliefs, desires, etc.–are not physical objects that can be moved. Rather, what composes all of these mental states in a person's brain is merely the configuration of that person's brain–which neurons fire in what ways, etc. So when Williams asks us to imagine that the contents of person A's brain are transferred into person B's brain and vice versa, he is asking us to imagine something that is impossible. The only way in which Williams' scenario could be

[39] To borrow an expression by Lawrence Lombard.

achieved would be by recording the configuration of both A and B's brain, then reconfiguring B's brain to have the configuration that A's brain had and vice versa. This is the closest possible method to achieve what Williams is asking of us. But in this case, the contents of A's brain have merely been replicated in B and the contents of B's brain have merely been replicated in A. Is our intuition here the same as Williams'? That is, are we certain we want to say that A now resides in B and B now resides in A? I certainly do not. It is clear to me that A and B no longer exist and the two persons that do now exist are mere duplicates of A and B. The A-duplicate is now in the B-body and The B-duplicate is now in the A-body.

It is also interesting to point out that those who hold Williams' view say that it matters how the brains come to have their contents. That is, if the B-bodied person's brain has A's contents as a result of scientists scanning A's brain to record its structure, and then caused the same structure to be formed in the B-bodied person's brain, then the B-bodied person would be A. But if the scientists imposed a random structure on the B-bodied person's brain that just happened to be the same structure as found in the A-bodied person's brain, then the B-bodied person would not be A.

Consider an analogy to a copy machine.[40] If a copy machine produced a piece of paper with the contents of the original printed on it because of what was printed on the original that was scanned by the machine, then it is a copy. But if the copy machine malfunctioned and printed some random content on a piece of paper that just happened to be the same contents as the original, then is it a copy of that original? Does the difference in how the content got there matter in determining whether or not it is a copy of the original? Regardless of how we answer this question, what is obvious is that in both

[40] This analogy was suggested to me by Bruce Russell.

cases, the piece of paper is certainly a mere copy of the original and not the original itself. Hence, in the case that Williams gave regarding how person A and person B's mental content is transferred, regardless of the method by which this is done, all that exists at the end of the procedure are duplicates, or copies of A and B; neither are the original A and B.

William's also adds a twist to his example. He says that before the procedure, suppose person A was asked to choose which of the resulting persons should be tortured and which should be awarded with $100,000. Now suppose person A opted for the award to go to the 'B-brained/bodied' person, while the 'A-brained/bodied' person would be tortured. Post procedure, if A's preferences were granted, should we not expect the 'B-brained/bodied' person, who now has A's mind (or one qualitatively identical to it), to remember (or seem to remember) making that choice and express a great deal of relief and satisfaction over making what turned out to be a wise choice? [41] If this is right, then it reinforces our original intuitions that the process was essentially a body swap. This can only be true because we take the person who expresses relief and satisfaction over making a wise choice while inhabiting body B to be *the same person* as the person who made that very same choice while inhabiting body A.

Again, Williams' example does not show what he hopes it to show. For we can adequately account for everything above given our understanding of how this procedure would be implemented. The resulting person in B's body would only seem to remember making that choice since his mind was just reconfigured to be like A's. So it would not be the case that he would be glad that *he* made that choice; rather, he would be glad that *his predecessor* made that choice. He would only *seem to remember* making that choice but

[41] Williams, Bernard. "The Self and the Future". *Philosophical Review* 79, 1970, pp. 161–180.

it would not actually be him who made it. Rather, it would be his predecessor that made it. Still, the resulting person in B's body would express the same degree of relief and satisfaction over the choice that was made since it would mean that he would be awarded the $100,000 rather than being tortured. It just would not be true that *he* was the one that made that decision.

This case is similar to the one above by Parfit in which although there is not continuance of part of the brain, there is not persistence either. Williams' case was supposed to be one where there is not continuance of part of the brain, but there *is* persistence. The first part was satisfied, but the second one was not. Hence, this case does not provide us with a counterexample to the view that continuance of part of the brain is necessary for persistence.

A similar example to Williams' is given by Oritsegbubemi Oyowe[42] as a response to Jeff McMahan. The dialectic begins with McMahan's case:

> *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).[43]

Obviously, there is both physical (whole body including whole brain) as well as psychological continuance for the original. But for the replica, there is not physical continuance of any kind because the replica is formed out of new matter. However, there

---

[42] Oyowe, Oritsegbubemi. "Surviving without a Brain: A Response to McMahan on Personal Identity". *South African Journal Of Philosophy*, 29, 2010.

[43] McMahan, Jeff. *Ethics of Killing: Problems at the Margins of Life*. Oxford: University Press, 2002, p. 57.

is psychological continuance with the original since the replica has the entire content of the original's mind.

Most people's intuition (and I confess it is my own as well) is to send the replica on the suicide mission. This example (and ones like it common in the literature) is used to pull on the strings of our egoistic, self-interested concerns. Advocates of physical continuance views, like McMahan, say that if your intuition is to send the replica, then that must mean you do not believe you are your replica. They then claim that the reason you hold this belief is that the replica is not a physical continuant of you. In other words, they believe the lack of physical continuance between you and the replica is what accounts for your not caring for the replica as much as you care about yourself. Thus, they see this example as a victory for their view, and conclude that it shows physical continuance is necessary for survival.

As I mentioned above, Oyowe thinks that claiming this as a victory is too hasty. He gives another version of this example in which the procedure is altered. The alteration, he believes, draws out an important difference between the two cases that brings to the surface what truly motivates our intuitions. The result he hopes to show is that the original case does not really offer any support to physical continuance even though it may seem to at first. Oyowe's modified version requires the use of a distinction that Peter Unger uses between what he calls 'core psychology' (CP) and 'Distinctive psychology' (DP). CP is the fundamental psychological *capacities* for thought, consciousness, and other functions of the mind and is grounded in continuance of the brain. Moreover, CP is common to all persons in the sense that we all have the same psychological capacities. In this sense, my CP is not different from yours. However, this

does not mean that everyone does not have their own CP. Psychological capacities in Unger's sense are types. Instances of those types are, if one is a materialist, particular neurological configurations. That is, if something happens to my brain, I might lose my CP but you, of course, do not lose yours. DP however, is the *content* of the mind (e.g., memories, beliefs, desires, etc.), which obviously is unique to each person in the sense that my memories, beliefs, desires, etc. are my own and not the same as yours.[44]

I will summarize Oyowe's modification to *The Suicide Mission* as follows. Suppose your DP is removed and immediately after an exact replica of your body in its current state (i.e., without any DP) is formed out of new matter. The device then generates a new DP qualitatively identical to the original. Simultaneously, your original DP is then placed in the replica body and the computer generated DP is placed in your original body. Suppose you were asked beforehand which resulting person would be the one to be sent on the suicide mission. Thinking in the same egoistic, self-interested way, whom should you choose?

If one held that brain continuance was necessary, then the answer should be clear: the person with the replica body and brain should be sent and the person with the original body and brain should remain here. But Oyowe suspects that is not what most people's intuition would be. In this modified version, his intuition is that the person with the original body and brain ought to be sent.

The reasoning behind this will be revealed upon a closer look at who does and who does not possess CP and DP in each version of the case. In the original case given by McMahan, one has your DP and CP. This is obviously the one with the original body

---

[44] Unger, Peter. "A Reply to Reviewers". *Philosophy and Phenomenological Research* 52, 1992, pp. 159–176.

since only that one has the original brain; the other has a duplicate brain and hence a duplicate DP and CP. The reason for our likely unanimous agreement to send the replica in this version is due to the fact that the replica has neither of the two types of psychology where the original has both. The one with both is more clearly and obviously you. But in Oyowe's version, he claims that one resulting person has your original DP but not your original CP, whereas the other is vice versa.

Since in most cases our CP and DP are within the same person, when given the option of choosing between a person who has both as opposed to a person who has neither (as is the case in McMahan's version), then of course we will choose the one that has both. As Oyowe says, "…my decision goes with the option that offers me more of the original..."[45] The true test, however, is when we are faced with choosing one type of psychology over the other. When faced with this, Oyowe thinks what is truly necessary for survival becomes clear. It does not have anything to do with physical continuance. He continues, "…it does not show that my commitment to my CP outweighs my commitment to my DP."[46] Yet this is what McMahon's example would need to do in order to show that physical continuance is necessary for survival (since one's CP is grounded in brain continuance). But since it does not do this, Oyowe argues that McMahan's case, and cases like it, do not truly support bodily theories despite their *prima facie* appeal.

The problem with Oyowe's modification of McMahon's example is that it asks us to imagine the same impossible things that Williams' example asked us to imagine. Since the basis of Oyowe's modification is that we 'remove' the person's DP (which is just the

[45] Oyowe, 283.
[46] Ibid., 283.

person's memories, beliefs, desires, etc.) to be later placed into another brain, he is asking us to imagine that we can remove these mental states as if they were physical objects and implant them into another brain. But again, what composes all of these mental states in a person's brain is merely the way that person's brain is configured. Just as in Williams' example, these mental states cannot be removed without removing the brain itself. Or using Oyowe's terminology that he borrowed from Unger, there is no way to remove one's DP without removing one's CP since it is grounded in the brain itself.

How then, are we to make sense out of Oyowe's modification? I do not believe that we can. The example, as it stands, is unintelligible. The closest we could get would be to remove the DP from the original person's brain, essentially wiping it clean, and then reconfigure it to have the same configuration it had before. The brain found in the replica's body would then also be reconfigured to have the same configuration as the original brain. The question now becomes: are either of these resulting persons the same person as the original? That is, has the original person persisted in either of these two resulting persons? The resulting person in the replica's body would obviously be a mere duplicate. This is because the replica would only have a DP qualitatively identical to the original person's and moreover it would be embodied in different matter. But what of the resulting person in the original body? Although this person has physical continuance of the brain and body, the original person has not persisted since his original DP was wiped clean and replaced with a qualitatively identical DP. That is, the DP that is currently in the original brain was created by copying the DP in that original brain, wiping it clean, and then imposing the same configuration of that DP onto the blank brain once again. If this is how the process is completed, then the DP housed in the resulting person's brain is

not the original DP; it is only a copy. So it seems the only answer we can give to the closest approximation of what Oyowe is asking us to imagine forces us to a similar result to the one found in Williams' case. Namely that the original person does not survive through either of the resulting persons.

Once again, we have examined a case where there allegedly was not continuance of even part of the brain, but there was persistence. But through analysis of the case, and by discovering that it assumed the impossible, it became clear that in the event of no continuance of any part of the brain, the original person did not survive. Still, we have yet to find a case where continuance of part of the brain is not necessary for persistence.

At the end of his article, Oyowe asks, "if our distinctive psychologies were to be secured without physical continuity, would we continue to attach the same importance to physical continuity of the brain?"[47] Where Oyowe says, "our distinctive psychologies" I take him to mean that they are truly our own and not a duplication of them. The problem with this is that in order to secure our own distinctive psychologies, they must be grounded in our brain. Any other manifestation of them would not be ours, but rather, that of a duplicate. I think the underlying difficulty in this matter rests on the tendency of advocates of psychological continuity theories to think and speak of 'contents of minds' like memories, beliefs, and desires as physical objects that can be moved from one brain and placed into another. If one understood the contents of the mind to be this way, it is easy to see how one could slip into the view that no sort of physical continuance is necessary for persistence.

In this chapter, I have shown that no physical continuance theory alone is sufficient for persistence. Moreover, although I have shown that continuance of the body

[47] Ibid., 286.

and of the whole brain is not necessary, I have been unable to show that continuance of part of the brain is not necessary. It seems that in every case where there is persistence, there is also continuance of at least part of the brain. Also, any attempt to separate the psychological aspects of a person from that person's brain results in a duplicate rather than the persistence of the person. For these two reasons, I conclude then that continuance of part of the brain is necessary for persistence. But since it is not sufficient, there must be some other necessary condition for persistence as well. In the next chapter, I will attempt to determine what this second necessary condition is.

CHAPTER 3 - PSYCHOLOGICAL CONTINUANCE

To determine what the other necessary condition for persistence is, we only need to reconsider what we tracked when considering the numerous examples discussed in the last chapter. It appears that what determined a person's survival was largely due to whether or not there was psychological continuance. Although it became apparent that in order for there to be psychological continuance in a manner in which the person persists, continuance of at least a portion of that person's brain was also necessary. But given some physical substratum, what was relevant for persistence was psychological continuance. We may think of psychological continuance as the continuation of a psychological life. But what kind of psychological life? What psychological aspects are we to favor when considering what kind of psychological continuance is relevant to persistence? Throughout this chapter, I will examine different psychological continuance views: the memory theory, Derek Parfit's theory of continuity and connectedness (which comes in degrees) no matter how it was caused, Bruce Russell's view of a continuous stream of consciousness where the cause is important, and finally my suggestion based on a Frankfurtian desire structure.

Before I begin, I want to be clear that in this chapter I am only considering what psychological aspects are important to the continuation of a psychological life. There is a distinction that must be made between considering what psychological aspects are important *given* persistence and what psychological aspects are important *to* persistence. The former will include psychological aspects that we regard as being important to the type of person we are regardless of whether or not we survive whereas the latter concerns which parts of our psychology are relevant to whether or not we survive. During the

course of this discussion, it will be easy to slip from one to the other. But this is not a maneuver I am making. My aim in this chapter is merely to consider different psychological continuance views that incorporate various aspects of our psychology that we regard as being important to us in this first sense, not in the second. I am making no inference that these features of our psychology are relevant in terms of determining whether or not a person persists, only that they are relevant in terms of a continuing psychological life. Considering which of the following psychological continuance views are necessary and/or sufficient for persistence (given a physical substratum) will be the topic of the next chapter.

It is also important to note that when discussing these mental or psychological states, regardless of what kind (memories, beliefs, desires, etc.), I am referring to particular states. That is, I am referring to tokens rather than types.[48] For example, say that you and I have the exact same set of mental states after undergoing some procedure. This does not commit one to say that you and I are the same person. The reason for this is that although each of us have the same set of mental states when it comes to type, we do not share the same tokens. In other words, you have your mental states and I have mine. The particular instances of the mental states between you and me are distinct; therefore, no view is committed to saying that you and I are one person.

With that in mind, let us consider Locke's memory theory.[49] This view holds that the kind of psychological continuance that amounts to the continuation of a psychological life is to be found in memory. For example, suppose a person right now has the same memories that you have of being in Paris five years ago. (Note that whether this person

[48] See Peirce, Charles Sanders, "Prolegomena To An Apology For Pragmaticism", *Monist*, vol.16 (1906), pp. 492–546 for this distinction.
[49] Locke, John. *An Essay Concerning Human Understanding*, 33-52.

really remembers or merely seems to remember is relevant to persistence, but not to continuance. The reason why will soon be explained.) If this were the case, there would be continuance between you from five years ago and this person. In other words, when the memories of persons at different times are linked together in this way, we have the continuance of a psychological life. Whether this continuation of a psychological life is the life of one person or two depends on whether or not the person actually remembers or only seems to remember. If the person really remembers, then you are identical with that past person. Here is a case of continuance and identity. But if the person only seems to remember, there is continuance but not identity.

A famous criticism of Locke's memory theory called the Brave Soldier Paradox was given by Thomas Reid.[50] Suppose a young boy was flogged at school. That young boy became a brave solider who, at the time of being the brave soldier, could recall being flogged as a boy. The solider then became a general much later on in life and could then remember being the brave solider, but could not remember being flogged. According to Locke's memory theory, the young boy is the brave solider, the brave solider is the general, but the general is not the young boy. This, however, is not possible. If x and y are one and y and z are one, x and z cannot be two. This is, of course, because identity is transitive: if x is identical to y, and y is identical to z, then x is identical to z. But Locke's view of memory continuity, as it stands, violates transitivity.

This is easily remedied by appealing to indirect links of memory as opposed to direct links. That is, we may still say that the old general is the same person as the young boy even though he can not remember being the young boy because he can remember

[50] Reid, Thomas. "Of Mr. Locke's Account of Our Personal Identity." Perry, John. *Personal Identity.* Berkeley: University of California Press, 2008. pp. 113-118.

being the brave soldier. And since the brave soldier can remember being the boy, this indirect path of following the links of memory back is enough to allow us to conclude that the old general is identical with the young boy.

A more serious problem for the memory view is pointed out by Joseph Butler.[51] The problem is that Locke's memory view is circular. This is because only you can actually remember your own experiences. Moreover, you can *only* actually remember your own experiences. Any alleged-memories you have other than your own are not real memories. Similarly, any real memories you have can only be yours. For if someone else had them, they would not be real memories because this other person did not have those experiences. So it is not your memory of an experience that makes it yours, but rather, you remember it only because it is already yours. In short, Butler's objection is that although memory can disclose your identity with a person who experienced some past event, the memory does not make that person you.

David Shoemaker's response to this objection is to introduce a notion called 'quasi-memory'[52], which is just like real memory but removes the identity requirement. It amounts to saying that I have a quasi-memory of an experience just in case I seem to remember an experience and that experience happened to *somebody*. How then is this different from merely seeming to remember? We add the requirement that my quasi-remembering of this experience is caused 'in the right way'. What this actually means is never explicitly stated by Shoemaker but it requires something like the cause of my quasi-remembering the experience must be connected to the experience itself as opposed to connected to something else (such as a hypnotist or scientist that implanted the quasi-

[51] Butler, Joseph, "Of Personal Identity," in *The Analogy of Religion*, 1736. Reprinted in Perry 1975, pp. 99–105.
[52] Shoemaker David W., "Selves and Moral Units," *Pacific Philosophical Quarterly*, 80 1999: pp. 391–419.

memory in my head). By appealing to quasi-memory, advocates of the memory theory are able to avoid Butler's objection because it avoids the circularity of presupposing that the person who has the memory must already be the same person as the one who experienced the event. That is, saying 'I quasi-remember doing x' does not presuppose I am identical with the person who did x since quasi-memory only requires that *someone* did x, and that this remembering was caused in the right way by the experience itself.

An even more serious objection to the memory view is that of dreamless sleep. During dreamless sleep, I cannot remember anything at all. It is also not possible to follow the links of memory back as we did as a response to Reid since I am not even indirectly linked to some past person during this time. So the memory theory seems to have the bizarre implication that I do not exist at any time "I" am dreamlessly sleeping or in any state of unconsciousness for that matter. In short, this view implies that I have never had dreamless sleep, nor have I ever been unconscious.

Parfit's view is another psychological theory, but is quite different from any other. His view involves what he calls Relation R, which amounts to psychological continuity and connectedness. The connectedness consists in many connections of psychological states (including memories, intentions, beliefs, etc.) and the continuity consists in overlapping chains of these strong psychological connections.[53] Obviously, this connectedness comes in degrees. The degree of connectedness between a person at two different times depends on the number and significance of the direct psychological connections between the person at those times. For example, I will be more closely connected with myself yesterday than ten years ago. This is because the number and strength of psychological connections between myself now and myself yesterday will be

[53] Parfit, Derek. *Reasons and Persons*, pp. 206-207.

much greater than the number/strength between myself now and myself ten years ago (assuming that I do not lose my short-term memory). But of course this does not mean I am not the same person now as I was ten years ago. The reason for this, Parfit explains, is that there are intermediate selves between then and now that are appropriately psychologically connected with each other; and since these intermediate selves all connect to each other, there is a continuous chain of selves from then to now that are all linked together to form one person. Parfit also says that there is no fact of the matter about whether I am the same person now as some earlier person if the degree of connectedness between these persons is neither strong nor weak. But I am that person if the degree of continuity is strong, and only one existing person has this strong connection, whereas, if the degree is weak, or more than one person can legitimately claim such a connection, then I am not that person.

Another important aspect of this view is that psychological connectedness is non-transitive (the fact that A is psychologically connected with B and B is psychologically connected with C does not imply that A and C are psychologically connected with each other). Thus, Relation R is not the same as identity. Parfit believes that if a person is replicated and then destroyed immediately after, in such a way that the replica has the same psychological continuity and connectedness as the original person, then Relation R holds between the original person and the replica even though identity does not. In terms of continuance, there would be the continuation of a psychological life between the original and the replica but there would not be identity.[54] He also thinks that the cause of Relation R obtaining does not matter. Of course, the normal cause of Relation R

[54] Parfit also says that it does not matter that identity is not preserved. His view is that all that matters is Relation R. This will be addressed in more detail later.

obtaining is a result of physical continuity. That is, the normal way we have our psychological continuity and connectedness is by having our own brain and body. But as Parfit says,

> I believe that physical continuity is the least important element in a person's continued existence. What we value, in ourselves and others, is not the continued existence of the same particular brains and bodies…I believe that what fundamentally matters is Relation R, even if it does not have its normal cause.[55]

For him, all that matters is the continuance of a psychological life through the preservation of Relation R, regardless of the cause of this continuation. The normal means by which Relation R is maintained is by the continued existence of our brain/body and by the persistence of the person. But if the person was destroyed, yet there was continuance of Relation-R by means of a replica, this would be "about as good as ordinary survival."[56] Again, this is because the original person is R-related to the replica. I will provide critical analysis of Parfit's view on what matters in Chapter 5.

Bruce Russell's view is similar to Parfit's. However, for Russell, the cause by which the psychological life continues does matter. Russell describes his view in terms of a continuing stream of consciousness in an unpublished essay from June 2001 entitled *Being John Malkovich*:

> From the standpoint of concern only for yourself, it is rational to want a stream of consciousness, part of which has been your stream of consciousness, to continue, but not rational to want a copy of your stream of consciousness to exist. Both, in some sense, have had their origins in you, but it matters how they have had their origins in you. If things mattered to rivers, a river should not care if its flow were divided. But it should care if the water from its source were pumped dry and at the same time another river somewhere else that looked just like it began to flow through territory that looked just like the territory it had previously flowed through.

---

[55] Ibid., 284-285.
[56] Ibid., 285.

So although Parfit would say that the continued existence of this copy of your stream of consciousness, since it is R-related to you, is about as good as your ordinary survival, Russell disagrees. For him, this duplicate stream of consciousness is not good enough because it is merely a copy. It is not really a continuous stream of consciousness part of which used to be yours; rather, it is just a *copy* of a stream of consciousness part of which used to be yours.[57]

Although these psychological views certainly have merit in determining what psychological aspects are significant to the continuation of a psychological life, I believe a great deal of other important features of our psychology are being neglected. However, they can be brought out through a thorough understanding of Frankfurt's notion of personhood and his resulting theory of what he calls "the self". Recall that Frankfurt built upon the rather fluid notions of personhood that were discussed in Chapter 1 to form a strict definitional answer to The Personhood Question: under what conditions is X a person? His answer is that X is a person if and only if X is able to form second-order volitions by engaging in self-reflection. Also recall that second-order volitions are defined as desires to act on (or not act on) first-order desires.

From this view of personhood, Frankfurt adds further conditions and ultimately constructs an analysis of what he calls "the self". As I said, I believe a great deal of what comprises this Frankfurtian self can be extrapolated to a theory of psychological continuance. In order to see this, we will need to spend some time laying out Frankfurt's view of the self in its entirety. The next portion of this chapter follows Frankfurt along this journey.[58]

---

[57] An application of Russell's view will be examined in the next chapter.
[58] I owe a great deal of my understanding of Frankfurt's complex view of "the self" to Sean Stidd.

With Frankfurt's view of personhood serving as a foundation upon which to construct his theory of the self, the first aspect he adds is called "identification." We identify with a desire when we want that particular desire to be the motivation of our will. Consider a smoker who is trying to quit. He has conflicting first-order desires: one desire is to refrain from smoking and the other is to have another cigarette. The person wants the former desire to be what motivates his will. That is to say, he has a second-order volition to refrain from acting on the desire to smoke. But suppose his desire for another cigarette proves too strong and, although he would prefer not to act upon this desire, he finds himself smoking another cigarette. Frankfurt describes this as a lack of coherence or harmony of the will. Since the desire that prevails is one the smoker would prefer not to act upon, Frankfurt says his will is not under his control. Moreover, not only is his will not under his control, but more importantly, it is not the will he wants to have. Since the prevailing desire is not one the person would choose to have, Frankfurt says this desire is not one the person identifies with. Identification, then, applies to those desires we want, after reflection, to be the governors of our will.[59] Or in other words, it applies to those desires about which we form second-order volitions. These second-order volitions are actually desires, specifically, a special subset of the second-order desires.

One may ask: what is the difference between a strong desire and an identified-with desire? The difference is that the identified-with desire creates a reason to alter your other desires to accommodate the one you identify with in case the two come into conflict. Suppose, for example, you have a strong desire to eat very greasy, fatty foods that are not at all good for your health. Now suppose after going to the doctor for a

[59] Regardless of what actually *does* govern our conduct, we identify only with those desires we *want* to govern us.

check-up, you are told you are at a high risk of a heart attack and need to radically adjust your diet. Even though you had a strong desire to eat these unhealthy foods, this desire may dissipate upon hearing the doctor's warning. Now suppose I am a professional chef that specializes in making greasy, fatty foods. If I were to find out from my doctor that I am at risk of a heart attack as well, and if I identify with the desire to eat these unhealthy foods, it may be more difficult for me to give it up. Since you do not identify with the desire to eat greasy food, the desire fades more easily for you. But since I am a professional chef that identifies with the desire to eat these foods, the desire may create a reason to alter other desires I have in order to sustain the desire to eat these kinds of foods. For example, I may decide to start going to the gym and taking vitamins to allow me to still eat what I want, and want to want.

In some sense, all of our desires are a part of us, and thus a part of our self. But the desires we identify with are more crucial to the self than those with which we do not identify. In other words, your non-identified-with desires can be understood as not really a part of you, or as being external to the self.[60] As Frankfurt explains, "…the desire is in the fullest sense his [when] it constitutes what he really wants – when he identifies himself with it."[61] In discussing the character of a person, or what constitutes the self for a person, he says "it concerns whether the dispositions at issue [i.e., the person's desires]…are characteristics with which he identifies and which he thus by his own will incorporates into himself as constitutive of what he is."[62]

---

[60] Frankfurt. *Identification And Wholeheartedness,* pp. 164-165 and Frankfurt, Harry G. Necessity, Volition, and Love. *The Faintest Passion,* New York, NY: Cambridge University Press, 1999, p. 105.
[61] Frankfurt. *Identification And Wholeheartedness,* 170.
[62] Ibid., 171-172.

From these two passages it is clear how important identification of desires is to the self. A useful metaphor to help illustrate this would be to think of the self as being divided into two parts: the inner and outer. The outer self contains all our desires while the inner self contains only those with which we identify. The two together make up the whole self, but according to Frankfurt, the inner self, which houses our identified-with desires, is more vital to the self than those non-identified-with desires.

There is something troublesome about this, however. Returning to the unwilling smoker example, Frankfurt says this person identifies himself with his desire not to smoke. That is the desire he wants to be the motivator of his will. Understood in this way, it seems this person should identify himself in accordance with his desire not to smoke and say he is a nonsmoker when asked if he smokes. This seems not only strange but also false. When asked if one is a smoker, to formulate an answer, one does not consult whatever desire one wants to be the motivator of one's will. The answer lies in whether or not the individual smokes, not in whether or not the individual wants to smoke, nor in whether the individual *wants* to want to smoke. I believe there is more to be said about this that may serve as an adequate answer, but I shall address it later.

Regardless of this minor (and I believe temporary) befuddlement, we now have a strong start to Frankfurt's conception of the self. But more development of the notion of identified-with desires needs to be formulated before we have his full view. The self does not consist merely in those desires with which one identifies, but additionally, they must also be desires with which the person is satisfied. When a person identifies with one desire as opposed to another, Frankfurt explains that the goal is not to eliminate the conflict between those two desires; nor is it even to reduce the tension between them. It is

possible that this may happen–the other desire may be weakened or perhaps eliminated altogether from the person's list of desires–but this need not be the case. It is possible that the conflicting desire will remain even after the person identifies with its counterpart. So a person's identifying with a desire does not necessarily eliminate the conflicting desire, nor does it eliminate the conflict between the desires themselves (since the desire to do x and the desire not to do x still conflict regardless of whether or not I am conflicted about which I want to satisfy). Rather, identifying with a desire eliminates the conflict *within the person* as to which desire the person wants to be the motivator of his will. When this happens, the person is no longer uncertain about what side he is on in the conflict between the two desires. This commitment to the identified-with desire is thus a wholehearted commitment, which is to say, the conflicting desire, although it may still be present among the person's desires, is not one with which the person identifies. Thus, wholeheartedness amounts to none of one's identified-with desires being in conflict. To desire something wholeheartedly, on Frankfurt's formulation, is to be identified with that desire and moreover for that desire not to conflict with any other identified-with desires.[63]

Frankfurt sees wholeheartedness as a kind of unity of the will. If a person had two conflicting desires that he identified with, then each desire would generate a reason for the person to act in two conflicting ways. But since the desires are in conflict, only one can prevail on a given occasion. Only when a person identifies with a desire wholeheartedly, has the internal conflict of the will been settled. As Frankfurt explains,

> Wholeheartedness does not require that a person be altogether untroubled by inner opposition of the will. It just means that he must be resolutely on the side of one of the forces struggling within him and not on the side of any other. Concerning

[63] Ibid., 173.

> the opposition of these forces, he has to know where he himself stands. In other words, he must know what he wants [or perhaps better: what he *wants* to want].[64]

When asking in what does wholeheartedness consist concerning psychic elements of some feeling or attitude towards a desire, Frankfurt answers that,

> It consists in his being fully satisfied that they, rather than others that inherently (i.e., non-contingently) conflict with them, should be among the causes and considerations that determine his cognitive, affective, attitudinal, and behavioral processes.[65]

In other words, it is being fully satisfied that these identified-with, wholehearted desires, as opposed to others of the same kind, are what determine one's will. This unity of the will is an important aspect of the self according to Frankfurt.

Additionally, I believe this may clear some of the confusion on something that seemed problematic earlier. Remember the question that was posed to the smoker about whether or not he smokes. Suppose this was asked of a smoker who is happy with the fact that he smokes. His answer would be a simple 'yes' whereas the unwilling addict's answer would be a '…yes…' with a sense of shame. This feeling of shame, when the governor of one's will is not one which the person wants to have, serves as a motivator to replace the unwanted desire with one which the person does identify. In other words, Frankfurt describes the unwilling smoker as possessing a lack of coherence or harmony of the will. Hence, he feels shame. This resulting shame functions as a kind of response mechanism to the person's having a will he does not want to have as a way to generate a reason that motivates the person to obtain the will he does want to have.

---

[64] Ibid. *The Faintest Passion*, 100.
[65] Ibid. 103.

The kind of conflict just discussed concerned which competing desires should earn a position on the list of identified-with desires. But there is another kind of conflict that Frankfurt discusses, although briefly, that also seems relatively important to his conception of the self. This other conflict concerns how high a position these identified-with desires ought to take on that list. It is not enough to simply pick out which of all your desires will qualify as your identified-with desires; we must also determine the hierarchical order of these desires. Moreover, Frankfurt acknowledges that sometimes, some of these identified-with desires may be in competition with each other regarding their position on the list. When a conflict of this kind is resolved, Frankfurt says, "…the competing desires are integrated into a single ordering, within which each occupies a specific position."[66] Also that the resolution of both kinds of conflict is what "creates a self out of the raw materials of inner life."[67]

There is one final element Frankfurt discusses in his construction of the self. Thus far, he has been focused entirely on desires, or what people want. He notes, however, that some of the things people want, or desire, are not things the person really cares about. In fact, even if a person wants one thing more than another, it does not entail he cares about it more. This is because he may not care about it *at all*. Moreover, it is not enough to appeal to the reason a person wants the thing. That is, one may be inclined to argue that if one wants something purely for its own sake rather than merely as a means to achieve some other desire, then this 'wanting it for its own sake' implies that the person cares about it. Not so according to Frankfurt. As he explains:

> When I want an ice-cream cone simply for the pleasure of eating it, that pleasure is for me a final end. I desire it for its own sake. But this hardly means that it is

[66] Ibid. *Identification And Wholeheartedness,* 170.
[67] Ibid. 170.

> something I care about. Very likely, the pleasure of eating the ice cream is something that I truly *do not* consider at all important to me. There is no incoherence in appraising something as intrinsically valuable, and pursuing it actively as a final end that is worth having in itself, and yet not caring about it.[68]

What then, does it mean to care about something? In exploring this question, Frankfurt gives an example about a concertgoer. Suppose that someone is a devout music lover of a particular band and plans to attend their upcoming concert. Suppose, next, that a close friend asks for a favor such that doing the favor will make him unable to make it to the concert. He agrees to do the favor, so it is clear doing the favor is more important to him than going to the concert. However, he is disappointed he will not be able to go. If forgoing the concert will be some kind of loss to him, or result in disappointment, that must be because going to the concert is something he still very much wants to do despite his decision to help his friend instead. For if he no longer had a desire to go to the concert, he would have no reason to be upset about not being able to attend.

Frankfurt explains that the concertgoer cares about the concert, even after he agrees to forgo it, because "he continues to desire to attend the concert - and therefore to be susceptible to pain caused by the frustration of this desire - despite the fact that he now feels that satisfying the desire is less important to him than doing the favor for his friend."[69] So caring consists partly in the persistence of a desire. However, persistence alone is not enough to constitute caring. Frankfurt also thinks "the desire must endure through an exercise of [one's] own volitional activity."[7071] In other words, the negative impact that results from forgoing the concert must be in some way his own doing. That is,

---

[68] Frankfurt, Harry G. Necessity, Volition, and Love. *On Caring*. New York, NY: Cambridge University Press, 1999, p. 159.
[69] Ibid., 160.
[70] Ibid., 160.
[71] Note that Frankfurt recognizes this volitional activity for the desire to remain something he wants can be fully conscious or deliberate, but it need not be. This allowance of the non-deliberate volition is necessary to explain how we can care about something that we wished we did not.

the persistence of his desire to go to the concert must be due to his unwillingness to give up that desire. A person cares for something then, when he or she has, and moreover identifies with, a persistent second-order desire about some first-order desire that he or she volitionally (be it deliberate or not) supports or maintains. Furthermore, if the desire were to begin to dissipate or be forgotten, the person would actively seek to replenish its importance so as it will not be lost. Caring about a desire expresses a kind of commitment to that desire to ensure it always holds some meaningful position among one's set of desires.[72]

The person who wants ice cream does not care about it in the Frankfurtian sense because he does not satisfy all of the conditions for Frankfurt's account of caring as described in the previous paragraph. Suppose a person finds that he does not have enough money to buy ice cream. He would still want it and we may suppose he would be disappointed that he was not able to get it. Since the person wanting ice cream has a persistent second-order desire about some first-order desire that he volitionally supports or maintains, he satisfies the first condition for caring. However he does not satisfy the other conditions. That is, after a short time, we would expect the person to get over not being able to get ice cream. As this happens, his desire to get it would fade. Moreover, his desire to obtain the ice cream, even if only for its own sake, would not be one that the person would actively seek to replenish. It would be strange if the person, as he found himself losing his desire to have ice cream, actively tried to keep this desire in place to ensure it did not fade away. Since caring expresses a kind of commitment to a desire that

[72] Ibid., 162.

ensures it always holds some meaningful position among one's set of desires, it is clear that this is not something that would occur to the person wanting ice cream.

A useful analogy may be drawn to friendships. We generally have friends that hold varying degrees of importance to us. For some of our friends, we may not care if we lose touch, grow apart, and the friendship ends. But for others–our closest friends whom we deeply care about–if the friendship began to fade, we would take an active role in trying to maintain it to ensure it does not deteriorate further. This active role of doing whatever it takes to keep it alive is the mark of caring for a friendship. Similarly, the mark of caring for a desire is the active role of doing whatever it takes to keep that desire in a meaningful position on one's list of identified-with desires.

We have now given an account of Frankfurt's conception of the self adequate for the purposes of this dissertation. To recapitulate, the first element of the Frankfurtian self consists in having second-order volitions. This entails determining which first-order desires the person wants to be his will. This is the essential attribute of personhood for Frankfurt, and it serves as the foundation for his construction of the self. Next are the desires the person identifies with. All desires a person has belong to him in some sense, but only those with which he identifies are ones that are part of the inner self or are internal to his will. Hence, it is these identified-with desires, and not the non-identified-with desires that are crucial to the self. The position or preferential ordering of these identified-with desires is important as well. Then wholeheartedness, which involves one's knowing which identified-with desire takes precedence for all those that are in conflict, is what gives harmony or unity to the will. The final ingredient is caring. One cares about a first-order desire if one has an identified-with second-order desire that it not be

abandoned. These features together are what comprise the Frankfurtian self.

Before concluding my discussion of the Frankfurtian self, I want to mention something that I believe is implicit in Frankfurt's view: consciousness. It may be surprising that consciousness was not mentioned since in all the earlier views of personhood that we examined, it seemed to be the predominate feature. Although not explicitly stated as a necessary condition anywhere in Frankfurt's account, I take it to be implied that it is necessary. Since the view is that a person is a being who is capable of forming second-order volitions, and what constitutes that person's self is wholeheartedly identifying-with and caring-for those second-order volitions, it seems that to actually be able to do this (as opposed to merely simulating it, like a robot or a zombie perhaps), being conscious would be a prerequisite. In fact, a creature's mere ability to have a desire seems to imply that it is conscious. However, consciousness alone is not sufficient for being a person, for lots of creatures (e.g., dogs, cats, etc.) are conscious but are not persons. Again, although consciousness does not explicitly appear in Frankfurt's account, I believe we may charitably attribute it as a necessary condition to his view of personhood and hence, to the self.

At the end of Chapter 2, we had determined that some physical substratum was necessary for persistence. But since no kind of physical substratum alone was sufficient, it was apparent that some other component was also necessary. When deciding whether or not a person survived in the cases from Chapter 2, what carried most of the load was whether or not psychological continuance was present. Hence, I concluded that some kind of psychological continuance was the missing component necessary for persistence in addition to the physical substratum.

In this chapter I have given the standard psychological continuance views as well as my own Frankfurtian continuance view. As I stated at the beginning of this chapter, I have only been considering what psychological aspects are important to the continuation of a psychological life independent of whether or not the person persists. I warned that it is easy to slip from one to the other but I have not made this inference. My aim in this chapter was merely to consider different psychological continuance views that incorporate various aspects of our psychology that we regard as being important to our psychological lives regardless of whether or not there is persistence. In the next chapter, I will consider what kinds of psychological continuance are necessary and/or sufficient for PI given some physical substratum.

CHAPTER 4 - PERSISTENCE

Now that we have considered different psychological continuance views, each of which incorporate various aspects of our psychology, we may now explore which of these views are necessary and/or sufficient for persistence (given a physical substratum as already determined in Chapter 2).

I find Frankfurtian continuance to be an interesting and compelling criterion. That is, given that Frankfurtian continuance incorporates so many features of one's psychology, and since we have determined that some form of psychological continuance is necessary for persistence, I find Frankfurtian continuance to be an interesting candidate.

Consider an analogy to a chair.[73] There may be some essential properties to being a chair but there may also be other considerations like its history in space and time that although not essential to it, do play a role in its being the *same* chair. In the same way, even if it is not an essential feature of being a person that one have a robust Frankfurtian desire structure, it is possible that the features that compose Frankfurtian continuance are relevant to the persistence of that person. In other words, in the same way that a chair's history in space and time is not essential to it yet may be relevant to its persistence, so too can a person's Frankfurtian desire structure be relevant to that person's persistence despite these features not being essential to that person.

Some time must now be spent explaining what role Frankfurtian continuance will play when considering the persistence of a person. After I have done that, I will attempt to determine which kind(s) of psychological continuance are necessary and/or sufficient for persistence, given some physical substratum, by considering a variety of cases.

---

[73] My thanks to Bruce Russell for this useful analogy.

To begin, I find the relation between Frankfurtian continuance and persistence to be quite natural. Since on Frankfurt's view, what it is to be *a person* is simply to be able to form second-order volitions, then whatever second-order volitions one has, specifically those one wholeheartedly identifies with and cares about, ought to have some relation to whether or not that person persists. We see instances of similar inferences elsewhere. For example, sets are things with members; so to be the same set is to have the same members. Similarly, to be a physical object is to be a thing that occupies spaces at times; so to be the same physical object is to occupy the same space at the same time. So then, if we say: to be a person is to have second-order volitions, then it is only natural to suppose: to be the same person is to have the same set of second-order volitions[74].

Obviously, it would be too stringent to require a person to maintain the same Frankfurtian desires at all times in order to remain the same person. This would mean that a person could not change one's desires without ceasing to be the same person; this is certainly absurd. Returning to my example concerning the phrase "you're a totally different person" will be useful here. Recall that we do not normally take such statements literally, that is, to imply numerical distinctness. We simply mean that the person has changed in some relevant or important way from how they used to be. On the Frankfurtian view, this would simply mean that the person has changed a significant portion of his Frankfurtian desires. Yet it certainly seems possible to change one's desires, no matter how important, (a career path, a religion that largely governs one's life, etc.) yet still be numerically the same person and be psychologically continuous with one's earlier self. That is, these would not be the kinds of changes that would result in

[74] Of course it does not follow from any of this that these components cannot change i.e., it does not follow that sets cannot change members, that physical objects cannot move, or that persons cannot change second-order volitions.

one's ceasing to exist and the resulting person being numerically distinct from the original person. In other words, one's changing a significant portion of his Frankfurtian desires does not result in the person ceasing to exist and a new person coming into existence.

If one can change one's desires and still be the same person, how will Frankfurtian continuance be useful? A more detailed look at caring will lead to an answer. Remember that a person cares for a desire if he has a commitment to actively see to it that that desire maintains a significant position among his set of identified-with desires. That is, should a desire begin to fade away, the person, if he cares for this desire, actively halts its fading and restores it to a meaningful place on his list of desires. According to Frankfurt, this is what gives continuity of the self. As he explains,

> Suppose we cared about nothing. In that case, we would be creatures with no active interest in establishing or sustaining any thematic continuity in our volitional lives. We would not be disposed to make any effort to maintain any of the interests, aims, and ambitions by which we are from time to time moved.[75]

Frankfurt acknowledges that this lack of caring may leave our ability to have desires and volitions of the second-order fully intact. That is, we would surely still have wants or desires in our lives and we would also be able to will which desires we have. Moreover, we could still identify with some desires but not others, and further still, wholeheartedly identify. Frankfurt even says that some of these desires might tend to endure and provide some sort of consistency. However, this stability would be mere happenstance; it would not be of our own doing. Without caring, the things we deem important to us could change from day to day without any reason. Our desires would be random and chaotic without any sense of order. The importance of caring, then, is that it

[75] Ibid. *On Caring*, 162.

binds our desires together and gives us a meaningful consistency to our lives. Frankfurt explained it perfectly when he said "caring is…the indispensably foundational activity through which we provide continuity and coherence to our volitional lives."[76]

It is with this understanding of caring that I find Frankfurtian continuance useful when considering the persistence of persons. Although I believe that the entirety of one's Frankfurtian desire structure is relevant, continuity of caring is most pertinent. By 'continuity of caring' I do not mean that a person must care about all the exact same things, for we have already seen that this is not necessary for numerical identity. Rather, there must be some preservation or continuous development of person X's caring about certain desires at time T_1 and X's caring about certain (possibly different) desires at time T_2. Although the set of desires does not need to remain unchanged, nor do all the cared-for desires need to remain unchanged, there does need to be some causal link or connection between the desires from T_1 to T_2. The act of caring seems to be what provides such a link or connection. By 'continuity of caring' I mean that there is continuity of the person's care-structure over time. What that means is that it has been deliberately changed on the basis of certain acts of caring from one time to another. For example, a person has a certain care-structure at T_1, then through acts of caring, his care-structure changes and now this person has a different care-structure at T_2. The person's care-structure at T_2 has been formed from his previous care-structure at T_1 in such a way that the person exhibits an evolving care-structure over time.

I think there is some intuitive appeal to a view that places such importance on Frankfurtian desire structure. For instance, suppose that after a car accident or a botched brain surgery, a person loses all of his former second-order desires. It may not seem

[76] Ibid., 162.

unreasonable to say that the resulting person is literally a numerically distinct person. As noted in an above passage from Parfit, "…certain kinds of qualitative change destroy numerical identity. If certain things happen to me, the truth might not be that I become a very different person. The truth might be that I cease to exist–that the resulting person is someone else."[77] Moreover, there may be some reason to believe that Parfit himself thinks that continuity of desires or caring is necessary for persistence. As Frederick Doepke says, "After radical changes in ideals and affections…he [Parfit] holds that it is relatively trivial to think that the same person still exists..."[78] Doepke cites the Nineteenth Century Russian example as the basis for his claim:

> In several years, a young Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade. To guard against this possibility, he does two things. He first signs a legal document which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, 'Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent'. He adds, 'I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks.[79]

This example shows that Parfit seems to think that such a radical loss in one's ideals results in a substantial change to the person; one in which the person does not survive. In the same way, if one were to lose all of one's wholeheartedly identified-with cared-for second-order desires, it may be tempting to think he is no longer the same person. After all, all of his desires, values, and aspirations would be gone. The things he enjoys and his motivations to do certain things and act in certain ways that he regards as important and constitutive of who and what he is would be eradicated.

---

[77] Parfit, 202.
[78] Doepke, Frederick. "The Practical Importance of Personal Identity." *Logos: Philosophic Issues in Christian Perspective*, 1990, pp. 83-91.
[79] Parfit, 327.

Does this mean that a person's Frankfurtian desire structure is necessary for persistence? What if the resulting person's memory remains completely intact? Is this sufficient despite the loss of desire structure? To answer these questions, we must consider cases in which one's memory is held intact but Frankfurtian desire structure is not (and vice versa) to see which way our intuitions about identity go.[80]

An objection to the idea that continuity of caring is necessary can be found in the example of Saul on the Road to Damascus.[81] Suppose Saul experiences a dramatic religious experience and, as a result, not only does he change his name to 'Paul', but also undergoes a conversion in which all of his Frankfurtian desires are dropped and he acquires new ones. There is no continuity of caring between Saul and Paul, and hence, if we were to suppose that continuity of caring was necessary, we would be committed to saying that Saul is not the same person as Paul. This is a problem, of course, because people sometimes do undergo religious experiences that result in a complete change of their Frankfurtian desire structure; yet we do not believe them to be a numerically distinct person afterwards. It appears, then, that continuity of Frankfurtian desire structure is not necessary for persistence. What accounts for Saul and Paul being the same person rests on the fact that Paul can remember doing what Saul did. Since the person's memory remains completely intact, this is what accounts for Saul and Paul being the same person. In addition to showing that continuity of caring is not necessary, the Saul case also shows that memory is sufficient since it is intact but Frankfurtian desire structure is not.

From here, it may be tempting to think that memory is also necessary. To aid us in

[80] From here on, when considering what is necessary and/or sufficient for persistence, I will be assuming that a physical substratum is necessary since that has already been established in Chapter 2. Thus, the mention of a necessary physical substratum will be omitted.
[81] This objection was brought to my attention by Bruce Russell.

this, let us analyze the main character's condition in the film *Memento* (Newmarket Productions, 2001; film by Christopher Nolan). After being hit over the head by an intruder that broke into his house and raped his wife, Leonard Shelby (Guy Pearce) is unable to form new long-term memories. He can remember the attack and what happened before it, but is unable to remember anything after it for more than a few minutes. To accommodate for his condition, he writes himself messages, and even tattoos his body to remind him of things he would otherwise be unable to remember. The problem is that Leonard's system is flawed. As Russell points out in an unpublished essay from January 2001 on this film entitled *Memento*: *Memory and Personal Identity*, "what Leonard writes down are injunctions and, for the most part, reminders of the facts, not statements of the facts themselves."[82] Of course, this has monumental epistemological implications concerning what Leonard knows and what he is justified in believing. Russell spends a great deal of time addressing these issues in his essay, but he also addresses the metaphysical question of PI: is Leonard the same person he was before his head injury and is he the same person from one day to the next after that injury? Russell argues that Leonard is "a shattered self"[83] and that the resulting person is not the same person as before the attack, nor is the resulting person the same from one day to the next. The basis for this conclusion, of course, stems from the important role memory plays on Russell's view.

Since Leonard is still able to remember lots of things about his life before the injury, Russell thinks that "…there is enough psychological similarity, and memory connections, between the earlier and the later Leonards *for Parfit* to say that he is the

[82] Russell, Bruce. *Memento*: *Memory and Personal Identity*, p. 2.
[83] Ibid, 16.

same person after the injury as before..."[84] This is because Parfit thinks PI is determined by the degree of psychological continuity and connectedness (Relation R) and whether only one person has this relation. Russell thinks this relation holds between earlier and later Leonards enough, for Parfit, to say they are the same person. Here is his argument:

> Suppose P1 is Leonard in the present. Suppose P2 is Leonard a few minutes before then, but enough minutes before then that the current Leonard would not remember what happened to P2. Suppose P3 is Leonard a few minutes before P2, and so on, until we get back to PL, Leonard just before he lost his short term memory. Now my earlier argument is supposed to establish on Parfitian grounds that P1 = PL, P2 = PL, and so on. But if P1 = PL and P2 = PL, then P1 must = P2. This will hold for all of the *n* numbers from P1 back to PL. So according to this line of argument, Leonard is the same person as all the intermediate selves between P1 and PL, and the same person as PL.[85]

Although Russell believes there is enough psychological connection between the earlier and later Leonards to claim that they are the same person on Parfitian grounds, Russell himself thinks otherwise:

> …the lack of connection via memory between post-injury Leonards suggests that there are thousands of new Leonards after the injury. Perhaps the right thing to say is that the later Leonards are not the same person as the earlier Leonard but only a short continuation of the stream of consciousness that was Leonard…But the thousands of later Leonards are not identical to the earlier Leonard, and so not identical to each other…Leonard is a shattered self whose consciousness continues in thousands of dead ends.[86]

It appears that, for Russell, in order to be identical to an earlier person, one's connection via memory to this earlier person must stretch across some minimal length of time. He admits there is no determinate answer as to how long is long enough, but he does feel confident enough to say that "ten minutes is too short and a few years is long enough."[87]

---

[84] Ibid, 13.
[85] Ibid, 13.
[86] Ibid, 15.
[87] Ibid, 16.

I believe Leonard is the same person both before and after the attack, as well as the same person from one day to the next after the attack, despite his inability to form new long-term memories. A deeper look into Leonard's desire structure will show what motivates me to say this. Since post-attack-Leonard is able to remember his life up to the point of getting hit in the head, he remembers, among other things, that he has a wife, who she is, what she is like, and perhaps most importantly, that he loves her. Along with these memories comes certain desires like to keep his wife safe, to protect her, and presumably to avenge her death if she was ever wrongfully killed. These desires, it is reasonable to presume, are ones that Leonard has (or would have if the situation arose) before the attack. After the attack, although he is unable to form new long-term memories, he still has these same desires within him. Every day when Leonard wakes up, despite being unable to remember anything that happened between then and the attack, he is still driven by the desire to find the person who attacked his wife.

Most importantly, in terms of persistence of the person, is that this is a desire Leonard cares about. I argued earlier in this chapter that the most important aspect of the Frankfurtian self when it comes to the persistence of the person is caring. Caring is an active, volitional commitment to maintain those desires most important to us and it provides us with a prolonged unity, or sameness of self across time. It is apparent that Leonard cares about this desire. Moreover, it is clear that this is the number one desire on Leonard's hierarchical list of Frankfurtian desires. As Russell explains, "Leonard's life is dominated by motives of revenge. His main goal in life is to avenge what he thinks is the rape *and murder*[88] of his wife,"[89] and "…his pursuit of what he thinks is his wife's

[88] Italics are Russell's because there is reason to believe the person that raped his wife did not kill her.
[89] Ibid, 2.

murderer organizes and gives purpose to his life. If he had thought that he had already avenged her rape and murder[90], what reason would he have to go on living?"[91] Obviously, this is the motivating force that pushes his life in the direction it does indeed go. Without this desire, Leonard's life would be astronomically different. As Russell says, he would not even have a reason to go on. Whether or not Leonard would be the same person without this desire is a separate question. My point is that the fact that Leonard does have this desire before and after the attack (and the fact that he cares about it and that it is on the top of his Frankfurtian desire hierarchy) gives us sufficient reason to believe Leonard is the same person after his injury as he was before it.

I also think Leonard is the same person from one day to the next after his injury because Leonard possesses the desire to find his wife's attacker each day. Russell points out that an argument could be made against this position that, "…Leonard's intention [or desire to find his wife's murderer and kill him] does not persist but is created anew each day and so does not unify his life."[92] Also that, "…in Leonard's case it seems that he must create anew each day and hour his intention [or desire] to avenge his wife's death since he cannot recall having had it."[93] Even if this particular desire is created anew each day, we must ask how we can account for why the very same desire is recreated every single day. The reason is that there is the persistence of other Frankfurtian desires from Leonard before his injury to after. As I argued above, Leonard's desires to keep his wife safe, to protect her, and to avenge her death if she was ever wrongfully killed are ones that he has both before and after the attack. These desires certainly have persisted; and it

[90] In the film, we are given some reason to believe that Leonard has already avenged his wife but forgotten he has done so.
[91] Ibid, 4.
[92] Ibid, 14.
[93] Ibid, 14.

is the persistence of these desires that causes him to recreate the desire to find and kill her attacker each day. Whether or not the particular desire to find the person who raped and killed his wife persists from one day to the next is arguable. If it does persist, then there is all the more reason to believe that Leonard after the injury is the same as Leonard before it. But even if we are to suppose this particular desire does not persist, and is recreated anew each day, the only reason it is recreated anew each and every day after the attack is because certain other desires did persist. This persistence of other Frankfurtian desires is enough, I believe, to show that Leonard is the same person both before and after the attack, as well as the same person from day to day after it.

If what I say is correct, the Leonard case may show that Russell's memory view is mistaken but it does not show that Parfit's is. On his memory theory, Leonard continues to exist since, for Parfit, there is enough continuity of memory after the attack, and connectedness between Leonard after and before the attack. However, I believe there is a case that can refute all views claiming that memory is necessary for a person to persist.

Suppose a mad scientist kidnaps you and intends to put you through one of two procedures. Procedure 1 removes all of your memories but leaves your Frankfurtian wholeheartedly identified-with cared-for second-order desires intact. Procedure 2 removes all of your Frankfurtian wholeheartedly identified-with cared-for second-order desires but leaves your memories intact. We may suppose that you will not lose consciousness throughout either procedure as to avoid complications regarding discontinuity of consciousness.

Let us further explore the implications of each procedure. If you were to undergo Procedure 1, the resulting person would have a severe case of amnesia. This person

would not remember anything about you, your family, friends, etc. However, he would still enjoy all of the same things you do (although he would have to rediscover that he does, in fact, enjoy those things). Suppose the mad scientist then worked towards re-exposing the resulting person to the things you like and enjoy. After some 'therapy' of this sort, it seems likely that he would be able to reestablish his awareness of having the wholeheartedly identified-with cared-for second-order desires that he does, in fact, have. After all, the procedure only removed your memories, all of your wholeheartedly identified-with cared-for second-order desires are still there; the resulting person has merely forgotten he has them. It does not seem unlikely that after being exposed to these things, the resulting person would eventually be able to realize that he actually has the Frankfurtian desires that he has. From there, through the proper guidance, the other aspects of the Frankfurtian self can also be reestablished: identification, wholeheartedness, caring, etc. It seems that a person who undergoes Procedure 1 would not endure as drastic qualitative change as it may initially appear provided that he is afterwards assisted in relearning what Frankfurtian desires he has. After completing such therapy, the resulting person would be extraordinarily similar to how you were before the procedure took place. The only thing he would not be able to get back is his memories. Now, this is certainly no minor encumbrance, and I am not suggesting that the resulting person has not changed drastically; of course he has. But I do not find this drastic enough to result in the emergence of a numerically distinct person. Although it is true that the resulting person would not even remember his spouse, nor why he loves her, since the resulting person would still desire all the same things, he would presumably be able to fall in love with her again. This is because as all the things that made him fall in love with

her the first time are still things that the person after the procedure would find attractive.

What if the resulting person is unable to reestablish a connection with all of the Frankfurtian desires that he has? Suppose the desires are still there, but even after being re-exposed to the corresponding activity, the resulting person fails to recognize it is something he enjoys. Does this pose a threat to the resulting person being you? I do not think so. I find this no more problematic than maintaining the position that a person may persist after gaining or losing a desire. Surely having the exact same set of Frankfurtian desires is not necessary for a person to persist, so why would it be detrimental to the persistence of a person if he is unable to realize one of his desires after Procedure 1?[94]

Again, I am not saying the qualitative changes you underwent via Procedure 1 were not drastic changes; indeed they were. But they do not seem drastic enough to enable one to conclude that you were destroyed and a numerically distinct person was created. Therefore I maintain that after Procedure 1, the resulting person is still you.

Let us see if the same can be said of Procedure 2. If the mad scientist were to perform this procedure on you, the resulting person would remember everything about you, your past, your family, friends, etc. He would also remember enjoying everything you enjoyed, desiring everything you desire, caring for everything you care for, etc. However, he would no longer enjoy, desire, or care for any of these things. Since Procedure 2 wiped away all of your wholeheartedly identified-with cared-for second-order desires, none of the things the resulting person remembers caring about means anything to him anymore. Moreover, simply re-exposing him to these things will do nothing to bring these Frankfurtian desires back.

[94] Precisely how many desires a person can lose and still remain the same person will, obviously, not have a clear cutoff. However, I will say that as long as there is continuity of caring among the resulting person's desires from before and after the Procedure, the person has persisted.

In the case of Procedure 1, the desires were still there; the resulting person merely forgot he had them. But in this case, the desires are not there at all. It seems clear that it would not be easy for the resulting person to reestablish the same set of wholeheartedly identified-with cared-for second-order desires in this case. We can imagine the mad scientist trying to convince the resulting person that he really does enjoy the things you enjoyed. It is true that the resulting person would have memories of enjoying these same things, but *he himself* would not enjoy them. The resulting person would have to acquire new wholeheartedly identified-with cared-for second-order desires since his desire slate had been wiped clean. Now, it may be the case that he forms some desires that you had before the procedure, but if this does occur, it would be merely by happenstance.

It should be clear there is no persistence of desires between you and the resulting person just because the resulting person happens to form some of the same desires you had. But there was persistence of desires in Procedure 1. Take any one desire that you had before Procedure 1 and the resulting person (who I also believe is you) also has that desire after Procedure 1. The reason the resulting person in the first case has that same Frankfurtian desire as you did before the procedure is that desire persisted *through* the procedure. Since the desire persisted, there was continuity. Now take any one desire that you had before Procedure 2 and the resulting person also has after Procedure 2. In this second case, the desire did not persist; the desire was eradicated as a result of the procedure. It just so happens that that same desire was reformed in the resulting person.

Since the resulting person's Frankfurtian desires after undergoing Procedure 2 cannot be reestablished (or if they can be, not in the right kind of way), then the rest of the Frankfurtian self cannot be reestablished either. It seems that the resulting person

could not be reconstructed to be very similar to how you were before the procedure took place. This may lend some reason to believe that the qualitative changes that result from Procedure 2 are drastic enough to destroy the original person and result in the creation of a new one.

However, we have already seen from an example given above that memory is sufficient for persistence. Since your memory has remained intact after Procedure 2, the resulting person will be able to remember everything you did before the Procedure took place. Unlike in Procedure 1, the resulting person will be able to remember everything about you, your family, friends, etc. This ability to remember everything that happened to you just as you were able to before the procedure took place gives sufficient reason to believe that the resulting person in Procedure 2 is you.

I think we may conclude from this example that identity is preserved in both Procedures 1 and 2–just in different ways. That is, identity is preserved in Procedure 1 via the persistence of the Frankfurtian self and identity is preserved in Procedure 2 via the persistence of memory. I expect most people's intuitive response to be to reject this conclusion. The prominence of the memory criterion over all other psychological criteria and the deep-seated belief among its advocates that memory is necessary as well as sufficient will be what motivates this rejection. In other words, the tendency people have to place such importance on memories will lead them to deny my claim that memory is not necessary.

To help illustrate this further, consider the following scenario: Suppose the mad scientist explained to you the results of each procedure and gave you the option of choosing which one he would perform. I expect that most people's initial response would

be to choose Procedure 2. The thought of losing one's most precious memories is not something that most would be able to bear. I am sympathetic with this *prima facie* response. However I believe it is ultimately superficial. In order to see why, it will be useful to follow the reasoning that would lead one to such a response.

We can imagine a person faced with the choice between the two procedures thinking that he cannot imagine choosing to give up his most cherished memories. For example, we may consider the memory of his honeymoon. We can suppose this was the first major vacation this person ever went on with his spouse and that it was spent in the most beautiful place this person has ever been: Hawaii. We may further suppose that this person loves warm weather and sunbathing on the beach. With all these elements combined, it is easy to see why this particular memory would be so important to this person as well as why he would be so reluctant to give it up. But we must examine *why* this memory is so important. Clearly, it is because the *content* of the memory is important. That is, the only reason why he values this memory is that he values warm weather, lying out in the sun on a beach, and being with his spouse. If the person no longer valued any of these things, it is difficult to see why the person would value the memory. However, even if the person does not value the memory, just his possession of that memory is enough for the resulting person to be the same as the original person. I am not disputing this point; my aim is just to provide an account for why people may be tempted to opt for Procedure 2 even though identity would actually be preserved in both Procedures 1 and 2.

One may object that the fact that we are still able to take pleasure in our memories of doing things that we used to enjoy but no longer do is a counterexample to my claim

that the resulting person who undergoes Procedure 1 would be unable to place value in his memories.[95] For example, suppose someone used to like to ski, but no longer does. This person remembers the great pleasure he took in skiing and can still take pleasure in the fact that he participated in an activity that he took great pleasure in, even if he no longer does. This certainly seems true. If it were not, then when we come to change our desires, we should also lose all attachment of value to the memories that we associate with the desires we previously had, but no longer do have.

I see a significant difference between the resulting person after undergoing Procedure 2 and the normal way in which we lose interest in activities as just described in the example of the former skier. Notice that in the case of the person who used to like to ski, there is continuity of Frankfurtian desire structure between the person that likes to ski and the person who used to like to ski but no longer does. Since there is this continuity, this explains how the person can still attribute value to the fond memories he has of skiing. The only reason these memories are fond to him in the first place is because he still attaches value to them since it is something that he used to enjoy. Once his desires changed, even though he lost the desire to ski, he still retained the value that he assigned to those memories. But in the scenario of Procedure 2, if all of his desires were completely wiped clean, he would lose all value he once had about not only the desires, and what the desires were about, but also the value he attributed to the memories he had of doing those things. There is certainly nothing inconsistent about having fond memories of doing something that a person used to like to do but no longer has any desire to do. But this is only because there is still continuity of desires.

Although the particular things one desires may change, if there is still continuity

[95] This objection was brought to my attention by Lawrence Lombard.

(as there is in normal cases), this accounts for how one may retain the pleasure of thinking back on one's fond memories of skiing. But in the case of the person who endures Procedure 2, there is no continuity of desires at all. So once they are eliminated, all value he attaches to these memories will disappear too. He would just be left with bare remembrances of being on a slope but have no attribution of value to this memory one way or another, as opposed to a memory of skiing to which you attribute value as being something you, at one time, enjoyed. This latter attribution of value is what makes it a fond memory; and this is something that in ordinary circumstances we are able to do. This is how we are able to still take pleasure in thinking back on doing something that we used to enjoy but no longer do. But this is something the resulting person from Procedure 2 would not be able to do. He would not be able to think back fondly on his memories because that would require some attribution of value to these memories. But since his entire Frankfurtian desire structure has been annihilated, he would not be able to do this.

It seems there must be continuity of desires in order for memories to hold any value. Since there is continuity of desires in normal cases, like the person described above who used to like to ski, this accounts for how we are able to think back on memories of doing things that we used to enjoy but no longer do. But if one's desires were wiped clean, as is the case in Procedure 2, the resulting person would not be able to do this. He would just think back to the bare remembrance of engaging in some activity but have no emotional response to it. Or at the very most, he may remember that he enjoyed doing it at the time, but it would hold no value to him now.

Returning to the original scenario, if Procedure 2 was performed, and the person persists, it is true that the person would still have the memory of his honeymoon in

Hawaii, but he would no longer value any of the things about that memory since all of his desires have been wiped clean. This person would no longer enjoy warm weather, sunbathing on the beach, or being with the person that is his spouse. I am sympathetic with not wanting to lose one's memories, but the only reason we place value in our memories in the first place is that we value and care about the contents of those memories. But Procedure 2 eliminates all of one's Frankfurtian desires, which eliminates everything one cares about, including the importance one places on the contents of one's memories, thereby removing the importance of the memories themselves. So although one would still have all of one's memories after the procedure, the person would no longer care about them.

Although it may appear that when forced to choose between the procedures, Procedure 2 is the one that should be chosen, I believe this is the result of a failure of the person to realize that the value one places in his memories would be eradicated by that procedure. The persistence of the memories alone is sufficient for the person to persist but the value one places on these memories would be wiped clean. The person would be left with bare remembrances of past events, which though sufficient for being the same person, does not entail the value that we normally apply to our memories. I think the reason most have the *prima facie* response to opt for Procedure 2 is that they are presupposing that these attachments of value will go along with their memories. But as I have just argued, this will not be the case.

Hence, if it is the value that people place on their memories that makes them opt for Procedure 2, but they are mistaken in thinking that this procedure will actually maintain this, but it will be maintained in Procedure 1, then this gives ample reason why

Procedure 1 would be sufficient for the person to persist. Indeed, the only way to retain any of the importance of what we care about is to retain our Frankfurtian desires. This is done through Procedure 1 but not Procedure 2. It is unfortunate, however, that in order to retain one's values in our hypothetical mad scientist example, one must give up all of one's memories. But I believe this would result in the persistence of the person just as Procedure 2 would–only in a different way.

Again, my aim has not been to show that one ought to opt for one procedure over the other. I have argued that the person would survive in both cases, just in different ways; Procedure 1 by means of continuity of Frankfurtian desire structure and Procedure 2 by means of continuity of memory. I am only trying to provide an explanation for why most people, I expect, will be temped to deny this claim that identity would be preserved through both procedures. Since I have provided such an explanation and shown it to be unfounded, I hope to have dispelled the reason to deny my claim that both memory and continuity of Frankfurtian desire structure, specifically continuity of caring, are sufficient for persistence, yet neither are necessary.

Before moving on, let us consider one final case where the mad scientist performs both Procedures 1 and 2. In such a scenario, all of the resulting person's memories and his entire desire structure have been wiped clean. I find it clear that the person has not survived. Since there is virtually no psychological content left in the person's mind, I do not believe the person has persisted. If it is at all tempting to say that the person has survived, this must only be because of the persistence of the whole body. If one is tempted to hold such a position, I refer back to Chapter 2 where I considered this very case. I determined that one is being mislead simply because so often, 'same body' is

connected to ‘same person’. However, the modified Lockean example of swapping two person’s brains has already shown that whole body is not sufficient for persistence. But although it is not sufficient, remember it was established that some physical substratum is necessary. This is also what allows us to differentiate the resulting person after undergoing Procedure 1 or 2 from being a mere duplicate that either seems to have the same desire structure or merely seems to remember. That is, the fact that we require a physical substratum is what allows us to say that the resulting person (in either procedure) really is the same person that underwent the procedure and not merely a duplicate of him.

We may conclude from this chapter that, given some physical substratum (namely of part of the brain), neither memory nor Frankfurtian desire structure are necessary psychological components of persistence. However, although neither are necessary, I believe both are sufficient.

CHAPTER 5 - SPLIT-BRAIN CASES

Now that I have determined which physical and psychological conditions are necessary and sufficient for persistence I must now contend with different versions of split-brain cases. The standard split-brain case is as follows: suppose my brain is split into two hemispheres and that each hemisphere retains the entire content of my mind. Suppose further that one half of my brain is transplanted into the empty skull of a body while the other half is destroyed. It is clear that there would be continuance of a psychological life between myself and the resulting person in this case. But, is this the continuing life of some one person (namely, me) or two different persons? If we assume that the resulting person would retain all of the contents of my mind, it is almost unanimously agreed (at least among psychological continuity theorists) that the resulting person would be me.

The version of the split-brain case just described is the simplest of its kind. The more difficult version is one in which we suppose the other half of my brain is not destroyed but rather placed into another brainless body at the same time as the first half. In terms of continuance, this version is no more difficult than the former version. Here, we have two resulting persons each of which have their psychological origin from me. Thus, there is continuance between the psychological lives of each of the resulting persons and me. The harder question is the one concerning identity: which of the resulting persons, if either, *is* me? Assuming as before that none of the mental content of my mind is lost in either half of my brain, we now have two resulting persons each with qualitatively identical minds. Moreover, each of the resulting persons' minds are also qualitatively identical to *my* mind. This seems to lead to the conclusion that both of the

resulting persons *are me*, that is, numerically identical with me. But this is impossible; two distinct persons cannot be numerically identical with one person, for then they would be identical to each other. So is one of them me but not the other? If so, which one is me and which one is not? Moreover, why that one as opposed to the other one? There is no relevant reason to pick one over the other. Or are neither of them me? If this is the case, then I have ceased to exist; fission is death.[96]

Before I respond to each version of the split-brain case, I feel that an exploration of the metaphysics of fission is in order.[97] It is unanimously agreed that if the brain is split and each half is put into a different empty head, the result is two numerically distinct persons that are qualitatively identical with each other as well as qualitatively identical with the original person; but neither of which is numerically identical with the original person. What makes the second conjunct true is simply the laws of logic concerning numerical identity. Since one entity cannot be numerically identical with two distinct entities, it follows that the original person cannot be numerically identical with both of the resulting persons. But what makes the first conjunct true? That is, what makes it true that after fission, the result is two numerically distinct persons that are qualitatively identical with each other as well as with the original person? The answer is that both of these resulting persons, we are asked to suppose by the nature of fission cases, possess the entire content of the original person's mind within their respective half of the original person's brain. An important question that is rarely addressed, however, is: how does this happen? How does the entire content of the mind end up in each half of the brain?

This is not an irrelevant question. The method by which the entire content winds

[96] Whether or not death by fission is as bad as ordinary death will be addressed in the next chapter.
[97] My thanks to Lawrence Lombard for showing the need of such an endeavor.

up in both halves may alter whether or not the resulting persons are numerically identical with the original. Remember that those who employ split-brain cases do so with the presupposition that, in the single transplant version, the original person will have survived and in the dual transplant version, the original person will have not survived. But do all split-brain scenarios have this consequence? I will argue below that they do not. Sometimes splitting the brain will result in the continued survival of two persons, and in dual-transplant cases, sometimes one rather than none. Sometimes splitting the brain will not result in the survival of the original person in even a single-transplant case. Does it ever result in the survival of the original person in the single-, but not in the dual-transplant case, as those who employ the split-brain examples presuppose? My worry is that if the only ways both resulting persons could end up possessing the entire content of the mind are ways that force answers other than the standard ones, then we have reason to question the legitimacy of any conclusions drawn from arguments that make use of split-brain cases. Thus, it is necessary to explore the metaphysics of fission to determine in what ways both halves of a fissioned brain may end up possessing the entire content of the mind.

I foresee several possibilities as to how this could occur:

(1) All the contents were present in each half already.

(2) All the contents were present only in one half and when the brain was split, the contents were copied from the half that had them to the half that lacked them.

(3) All the contents were scattered in different places in the brain (some in each half) and when the brain was split, the missing features were respectively copied from the half that had them to the half that lacked them.

(4) All the contents were scattered in different places in the brain (some in each half) and when the brain was split, each half began to accommodate the functions of the missing half and reestablished the corresponding psychological features.

It may seem that if (1) is true, we face a coincidence problem. If the result is two persons after fission, but all the relevant features of what makes these two persons distinct (namely the presence of all the mental contents in each half of the brain) was present before fission took place, then it seems that the two persons were present before fission took place. To draw an analogy to the table example as discussed in Chapter 1, it would be as if two tables that had been pushed together[98] to give the appearance of being one big table were pulled apart but each table was there all along. It seemed that there was only one because they were alongside each other to give the impression that it was one big table. In the same way, there had always been two distinct persons; it just appeared that there was only one since each half of the brain was connected. But once they were fissioned, it became apparent that there had always been two.

Under this interpretation of option (1), if the single transplant version of the split-brain case were to occur, only one of the original persons will have survived. The one that survives will, of course, be whichever one's half brain ends up in the resulting person's body. In the dual-transplant version of the split-brain case, both persons will persist. If this were so, both persons would survive fission. What is interesting about this method of how to understand fission cases is that it does not result in either of the original persons ceasing to exist after fission.

This is one possible outcome if (1) is true, but there is another. Suppose persons are "maximal"[99] – by which I mean that each person is a maximal collection of whatever it is that makes something a person (i.e., their mental content). If this is the correct way to

[98] To make this more analogous to the split-brain case, we can imagine the tables were also latched together in some way to physically connect them.
[99] This possibility was suggested to me by Lawrence Lombard.

understand (1), then there would not be two persons before fission but only one. This is because the maximal collection includes all of the qualitatively distinct mental contents that are present in both halves of the brain and the person is that maximal collection; hence, there is only one person pre-fission.

If this were so, how ought we respond to the single and dual transplant variants of the split-brain case? In the dual transplant version, it is obvious that neither of the resulting persons would be identical with the original. This is because if they were identical, each of them would be qualitatively identical with the original and so with each other; but they are distinct. In the single transplant version, we ought to say that the original person does survive. This is because the resulting person has the maximal collection of what it is to be that person since the maximal collection is present in the one half of the brain that was transplanted. After all, there is not anything in the other half (the one that was destroyed) that is missing in the half that was transplanted. Since the interpretation of (1) that involves persons being maximal suggests that all the contents of the mind were present in each half, then the maximal amount of what makes a person a person is present in both halves. Thus, if only one half of the brain was transplanted, and that half contains the maximal collection of mental content of what it is to be a person, then the resulting person would be numerically identical with the original person.

For example, say the mental contents we are concerned with are x, y and z. These three contents are what make the person a person. So the maximal collection of psychological content that makes the person a person would be x, y and z. If x, y and z are present in each half of the brain, then each half possesses that maximal psychological content. If we understand persons to be maximal under (1), and we suppose that x, y and

z are present in both halves, then as long as only one of the halves is transplanted, the maximal would be present in the resulting person. Hence, the original person will have survived.

If (2) is true, the answer is simple. If only one half is transplanted and the other is destroyed, whether or not the person survives depends on which half is transplanted. This may be easier to grasp if we consider the case where both halves are transplanted. It should be clear that the resulting person with the half that originally contained the mental contents would be identical with the original person and the other resulting person would be a mere duplicate. The reason for this is that this resulting person's brain has only a copy of the mental content rather than the original content that is present in the other half. Applying this to the single transplant version, if the half that is destroyed is the half with the copied content, then the resulting person is identical with the original because he has the half of the brain that originally possessed these psychological features. Whereas if the half that is destroyed is the one with the original content, then the resulting person is a mere duplicate of the original since the psychological features in his brain are just a copy of the original.

If (3) is true, the answer to whether or not the resulting person (or persons, if we are considering the dual transplant version) is (are) numerically identical with the original would depend on the quantity and significance (in terms of how relevant the psychological features in question are to the persistence of the person) of the mental content that are being copied as opposed to the quantity and significance of those that are originally found in that half of the brain. For instance, if the majority of these mental contents are found in the left half, and only a small number of rather insignificant

contents are found in the right, then in a single transplant fission case, if the left half was the one that was transplanted, the person will survive. While if the right half was the one being transplanted, the person would only be a duplicate. If both halves were transplanted, then obviously the resulting person with the left half would be numerically identical with the original person and the resulting person with the right half would be a duplicate. I find this to be a clear scenario.

There are more difficult ones, however. What if the quantity and significance was divided evenly in each half of the brain? The dual transplant version is not so difficult; I find it intuitive that we should regard each of them as duplicates. The reasoning behind this is similar to what has been said above, namely, that there would be no reason to believe one, rather than the other, is the original person. Moreover, since both cannot be the original person, we ought to conclude that neither is. The single transplant version is much harder. If only one half was transplanted and the other was destroyed, it is not clear whether or not the original person has survived. The resulting person would only possess half (in terms of quantity and significance) of the relevant psychological features of the original person. Is this enough? There may be no definitive answer. While it is certainly true that a person may survive the loss of some of these mental attributes, precisely how much loss he can sustain is indeterminate.

It may be that if each hemisphere has exactly half of the relevant mental content, the resulting person would be identical with the original provided that only one half was transplanted–regardless of which one. But if both halves are, then neither is the original person. Or, it may be that half of the relevant mental content is not enough for the person

to survive even if only one half is transplanted. Regardless of which is true, it is clear that there is no determinate answer as to where to draw this line.

Option (4) is similar to (3) in that the entire mental contents are scattered in different places in the brain. However, where (3) says that the missing features were simply copied from one half to the other, (4) says that the brain halves each begin to take on the tasks of the other (now severed) half. I am not suggesting that either half of the brain regrows any part of the other half. What I have in mind is similar to what we sometimes see in cases of brain damage. When certain parts of the brain are damaged to the point that they cannot function properly (or at all), sometimes other still functioning parts of the brain will begin to take on the functions of the damaged parts. For example, after head trauma, a person may have damaged the portion of his brain that controls memory. But it is possible that the remaining undamaged portions of the brain may begin to take on the function of controlling memory even though these portions of the brain previously did not have this function. (4) suggests that this is what goes on in each half of the fissioned brain; the left half will begin to perform the functions that the right half used to have and vice versa. Moreover, this will continue until each half of the brain once again contains the entire contents of the unfissioned brain.

If (4) is true, we may give a similar answer to the one found in (3). Depending on the quantity and significance of the psychological features that are being reestablished, we may give different answers regarding whether or not the resulting person(s) is (are) numerically identical with the original. For instance, if ninety-nine percent of these mental contents are found in the left half, and only one percent is found in the right, then in a single transplant fission case, if the left half was the one that was transplanted, the

person will survive. While if the right half was the one being transplanted, the person would only be a duplicate. If both halves were transplanted, then obviously the resulting person with the left half would be numerically identical with the original person and the resulting person with the right half would be a duplicate.

The difference between (3) and (4) is in how the missing features end up in both halves of the brain. This difference of the missing features simply being copied as opposed to being reestablished by the half brain accommodating for its missing half will likely make a difference in some borderline cases. Recall that in some cases considered when discussing (3), it was unclear whether the quantity and significance of the psychological features being copied were enough to render the resulting person numerically identical with the original person. I think that in some of those cases, the fact that the features are not being copied in (4) but rather are being reestablished by the brain may tip the scales enough to lead one to conclude that the resulting person is numerically identical with the original.

For example, suppose that the left half of the brain contained seventy percent (again, in terms of quantity and significance) of the mental content whereas the right half contained thirty. Suppose the left half is transplanted and the right half is destroyed. I am inclined to think that this is a borderline case under option (3). In this scenario, seventy percent of the relevant content would be original content (i.e., not duplicated) while thirty percent is not original (since it was copied from the other half prior to being destroyed.) Although seventy percent is more than half, there is still a large portion of this mental content that has just been copied. Hence, I am unsure whether or not the resulting person

will have survived.[100] However, if the same scenario was to be considered under option (4), my intuition changes. If seventy percent of the mental content was originally present in the left half of the brain and that half of the brain managed to reestablish the missing thirty percent originally found in the other half, I think the original person has survived. I think there is enough content originally present in the left half to disregard the fact that an admittedly sizeable amount had to be reestablished in the way I described in (4). But if this same amount was simply copied over, as described in (3), I am less certain. So I believe that the method by which the missing psychological features end up in the brain halves do play a role in whether or not the person(s) survives.

By exploring these four possible methods by which each half of a fissioned brain may wind up possessing the entire content of the original person's mind, I have determined that the method by which this occurs most certainly plays a role in whether or not the original person survives. In the single transplant version, some methods result in the original person surviving whereas others do not. Similarly, in the dual transplant version, some methods result in neither of the resulting persons being identical with the original whereas some result in one but not the other of the resulting persons being identical with the original.

As I stated at the start of my discussion on fission, those employing split-brain cases are presupposing that in the single transplant version, the original person survives and in the dual transplant version, the original person does not. My concern was that the method by which each half of the brain comes to contain these mental contents may play a role in whether or not we arrive at these conclusions. I was further concerned that if the

[100] If you have a firm intuition on these percentages, replace them with whatever proportions you find to be a borderline case. My aim is simply to show that the method by which the missing contents end up present in the other half of the brain may indeed be relevant.

only ways in which both halves could come to possess the entire mental content are ways that force us to deny the standard conclusions, then any conclusions drawn from the typical interpretations of split-brain cases are called into question. My first concern was shown to be legitimate. We have seen that the method by which each half of the brain comes to possess the mental content is, indeed, important. In some examples of the single transplant version (specifically (2)), whether the original person survived depended on which half was transplanted and which half was destroyed. Similarly, certain examples of the dual transplant version (also (2)) resulted in one of the two resulting persons being identical with the original whereas the other was not. These are different conclusions to reach than those established via the typical way of understanding split-brain cases.

However, my further concern can now be discarded. Although there are certain scenarios where we come to different results than those who employ fission cases, there are others where we arrive at precisely those results (as in some examples of (3) and (4)). I conclude from this that the type of scenario that advocates of fission cases are asking us to imagine is metaphysically possible.

Now that this concern has been dealt with, I may give and defend my preferred answers to the two different variants of split-brain cases. In what is to follow, please keep in mind that I am supposing that the method by which the entire content of the original person's mind wind up in both halves of the brain is a manner that allows for the conclusion that advocates of fission cases endorse. That is, the manner is one in which the resulting person is identical with the original in the single transplant version and neither of the resulting persons are identical with the original in the dual transplant version.

Concerning the single transplant variant, as I said at the beginning of this chapter, it is almost unanimously agreed that the resulting person would be identical with the original. I agree with the majority in this case. In light of what I said in Chapter 1 about a similar example involving sawing a table in half, this answer may be unexpected. Recall the table example: we slice a table directly down the middle cutting it into two halves and destroy one half. The remaining half re-grows the rest of a table. I claimed that this table is not identical with the original because half a table is not a table, and in order for there to be continuity of a table, the object must be a table at every moment. Since half a table is not a table, there is a gap of existence between the two tables; hence, the table that results after the re-growth is not the same table as the original one.[101]

It may appear that in order to remain consistent, I must give the same answer in the split-brain case, since the cases are extraordinarily similar. However, there is a significant difference. The example regarding a table is about a physical object with no psychological properties. In order for these two cases to be completely analogous, in which case I would be forced to give the same answer, the split-brain case would have to be about a physical object also with no psychological properties. In split-brain cases, we are concerned with whether or not the same *person* exists after fission; and a person is a physical object that, as I have argued in previous chapters, must have psychological properties.

This is relevant because what determines the persistence of a physical object without any psychological properties is different from what determines the persistence of a physical object that must have psychological properties, i.e., a person. The difference

[101] Although the watch example given in a footnote in Chapter One is a counterexample to this Lockean principle that physical objects cannot have two beginnings of existence, recall that I am not endorsing the principle universally. I am only employing it as a means to explain the table example.

between the cases is that while cutting a table in half results in destroying the table, cutting a brain in half does not result in destroying the person. I agree this sounds strange at first, and I foresee the objection being that it is ludicrous to presume that a person may still exist even though his brain has been cut in half and is not housed in a body. Although it is true that the person may not be mentally active and is not embodied while the half of the brain is outside of a skull, I see this as no more problematic than a case of dreamless sleep or a person under anesthetics. In such cases, there are gaps in the mental activity of the person, yet certainly persons survive these gaps. Otherwise, no person has ever survived a night of dreamless sleep or an operation where they were given anesthetics. This does not seem different in any relevant way than what is taking place in the split-brain case. There is merely a gap in mental activity from when the brain is split and removed from the original skull until it is placed in the new skull and the resulting person reawakens. I see no reason to accept that persons can survive dreamless sleep or going under anesthesia but not the split-brain case (at least, this version of the split-brain case). It seems, then, that discontinuity of a person's mental life does not imply discontinuity of that person. This suggests more reason to believe that a purely psychological theory of persistence cannot be correct. The continuity of a physical substratum that was deemed to be necessary back in Chapter 2 also plays a role here in accounting for how and why the discontinuity of mental life does not imply discontinuity of person.

Another difference is that in the table case, we are removing a necessary component of what makes the table a table. That is, we are removing one half of the table itself. In the fission case, however, we are not removing a necessary component of what

makes the person a person. Remember that it was determined in Chapter 2 that only part of the brain is necessary. If we held that the whole brain was necessary, then we would not have a necessary component since we no longer have the whole brain in the fission case. But since we only require part of the brain, and since part of the brain is maintained in the fission case, we are not losing a necessary feature. The other necessary feature is some psychological aspect. Since the split-brain case stipulates that none of the mental content is destroyed via fission, and since I spent some time arguing that this is metaphysically conceivable, then this necessary feature is not lost either. The fission case would be more alike the table example if we required that the whole brain was necessary or if splitting the brain in half resulted in the loss of half the contents of the mind. If this were the case, and if the loss of half the contents of the mind resulted in the individual ceasing to be a person, then we would have a case that works more analogously to the table example. But since it is stipulated in the case that fission leaves the entirety of the person's mental content intact (which is metaphysically possible), and since we do not require the whole brain, I maintain that the two examples are not comparable.

Let us now address the dual transplant version. As I also said at the beginning of this chapter, it is unanimously agreed that neither of the resulting persons would be identical with the original. I agree provided that they each come into existence at the same time. But if one half of my brain is transplanted and the other half, rather than being disposed of, is put "on ice" for, say, a year, and is then transplanted into an empty body, this may make a difference. In a delayed split-brain case such as this, I believe that the first resulting person is me but the second one is not. The second is simply qualitatively identical with how I used to be at the time of fission. If I hold that transplanting only one

half results in my survival, then as long as only one half is initially transplanted, then the resulting person would be me. If at some later time, the second half is transplanted into another body, it would be strange to say that the existence of this *other* person determines whether or not *I* continue to exist. This would make whether or not I continue to exist depend on something completely external to me. This seems wrong because once my persistence has been secured in the first resulting person, the mere existence of some other person cannot take it away.

A similar explanation is found in my preferred answer to the Ship of Theseus example as previously given in Chapter 2. Recall that in this example we are to imagine the parts of a ship being replaced one-by-one over an extended period of time. After the first part is replaced, we are asked if the ship is identical with the original ship. Of course it is. Moreover, it seems wrong to say after the removal of any part that the ship is no longer the same ship as the original one. So we conclude that the ship is still identical with the original ship even after the last original part is replaced. The resulting ship now has all new parts but is still numerically identical with the original ship with all the original parts. Now suppose the original parts are reassembled in exactly the same manner as they were initially. Which is the original ship? My answer is that if the ship with all new parts was identical to the original ship at every moment as the parts are being replaced, then even when all the original parts are put back together, the ship with all new parts is still numerically identical with the original ship. The existence of this second ship (the one with all the original parts) does not somehow strip away the identity of the original ship from the first one (the one that had its parts replaced one-by-one). The second ship is simply qualitatively identical with how the original ship *used* to be. Just as

the first ship does not cease to exist when the second ship comes into existence as all the original parts are reconstructed, neither do I cease to exist when the second person comes into existence as a result of the second half of my brain being put into another body.

How long the delay is between transplanting the brain halves may also be relevant. I said "a year" in my example because I find that to be a clear case that enough time has elapsed to confidently say that the existence of the second person does not hinder my survival.[102] On the other end of the scale, what if the delay was only a split second? What if the second resulting person came into existence a split second after the first? Is this a case where the first resulting person is me but the second one is not? Or is the timeframe so close that we should regard neither of them as being identical with me? I am inclined to say the latter but then I find myself at the mercy of a sorites paradox. There will be no sharp cutoff as to precisely when enough time has elapsed between the transplanting of brain halves to ensure my survival via the first half being transplanted.

Suppose in a delayed fission case, the first resulting person (who is me), dies before the second half of my brain is placed in a brainless body. What are we to say of the resulting person once the second half of my brain is transplanted? Like before, I think it depends on how long I have lived before the second half of my brain is transplanted. For example, suppose I am only alive for a split second after the first half of my brain is transplanted. Once I die, the second half of my brain is transplanted and the resulting person lives on. I think this resulting person is me. However, suppose the first half is transplanted while the second is kept on ice and I continue to live for another sixty years before dying of old age. Upon my death, the second half of my brain is then placed into an empty skull. I do not believe this resulting person would be me. I say he would merely

[102] If you find this not to be long enough, replace it with a sufficiently long duration.

be qualitatively identical to how I used to be at the time of fission. Since I had lived so long, I would have gained a great deal of new memories, experiences, desires, etc. as well as lost many of these traits since the time my brain was split. Since so much of my psychology at the time of my death would be different from the newly resulting person with the second half of my brain, I do not think there is enough psychological similarity for him to be me. Also like before, there will be no sharp cutoff as to precisely how long I have to live before the transplanting of the second half of my brain results in a person that is only qualitatively identical to how I used to be as opposed to him actually being me.

Most psychological continuity theorists take what is called a "non-branching" approach to split-brain cases. This amounts to saying that one's PI is determined by whatever theory of continuity they prefer, plus the added condition that there can only be one person that has the appropriate continuity. This allows them to claim that a person will survive if only one half of their brain is transplanted, but if both halves are, then neither of the resulting persons is identical to the original person. In other words, the original person has not survived, and that the dual-transplant variant of fission results in death. My view is slightly different. I believe that branching is acceptable provided that it does not occur at the same time. That is, if both halves of the brain are placed in brainless bodies and the resulting persons come into existence at the same time (or perhaps within a few seconds of each other), then this type of branching does not allow for the persistence of the original person. However, if the branching takes place over a longer period of time (say, a year), as in the delayed fission case, then this kind of branching does allow for the persistence of the original person. In short, my view takes a "non-branching" *or* "sufficiently delayed-branching" form.

CHAPTER 6 - WHAT MATTERS?

I would be remiss to write a dissertation on PI without addressing Derek Parfit's very influential work in this area. He has two theses, one is negative, the other positive. The negative thesis is that identity is not what matters.[103] In saying that it does not matter, he means, "…not what makes our survival good, but what makes our survival matter, whether it will be good or bad. What is it, in our survival, that gives us a reason for special anticipatory or prudential concern?"[104] The positive thesis is an account of what does matter, which Parfit suggests is to be found in our deep-seated intuitions of what we ultimately care about when discussing the importance of personal survival.

I shall address his negative thesis first. From the split-brain case addressed in the previous chapter, he draws the conclusion that identity is not what matters. Recall that if only one hemisphere were transplanted, the original person would survive; but if both hemispheres were transplanted, then neither of the two resulting persons would be identical with the original person. That is, the original person would not have survived. As I said before when describing the dual transplant version of the split-brain case: fission is death.

I now turn to the question Parfit raises from this example: is death by fission as bad as ordinary death? Parfit's answer is no. He argues that it does not matter whether each half of the brain is transplanted or only one. The reason for this is that in both cases, the resulting person(s) is/are psychologically continuous and connected with the original person. Remember, Parfit's view is that psychological continuance is the preservation of Relation R, which amounts to psychological continuity and connectedness. The

[103] This has already been touched upon in several places throughout Chapter 3 but will be more thoroughly examined here.
[104] Parfit, *The Unimportance of Identity*. Oxford University Press, 2011, p. 430.

connectedness consists in many connections of various psychological states (including memories, intentions, beliefs, etc.) and the continuity consists in overlapping chains of these strong psychological connections. Parfit's positive thesis is that all that matters is the continuance of a psychological life through the preservation of Relation R, regardless of the cause of this continuation. Since this relation holds in both the single and dual transplant cases, Parfit argues that, keeping other things equal and putting extraneous considerations aside, there is no reason to prefer the single transplant to occur over the dual transplant. But since the person does not survive in the dual, but does in the single, transplant case, it must not matter whether the person survives. Hence, identity is not what *really* matters.

Why then do we have this deep rooted intuition that survival does matter? Parfit's explanation is that in everyday life the persistence of the person is the only way to preserve Relation R. But the split-brain case serves as a counterexample to that commonly held intuition. It shows there are ways in which one may retain psychological continuity and connectedness yet not survive. Hence, survival is only contingently what matters but not what *ultimately* matters. In other words, although it appears on the surface that PI matters, there is something deeper that really matters, namely, psychological continuity and connectedness (Relation R). That is all that really matters, and since it is maintained in both the single and dual transplant versions of the fission case, one should not care which version occurs.

Parfit realizes that it is hard to believe that identity is not what matters, so he offers an analogy to help us better grasp his claim:

> Imagine a community of persons who are like us but with two exceptions. First, because of facts about their reproductive system, each couple has only two

> children, who are always twins. Second, because of the special features of their psychology, it is of great importance for the development of each child that it should not, through the death of its sibling, become an only child. Such children suffer psychological damage. It is thus believed, in this community, that it matters greatly that each child should have a twin.
>
> Now suppose that, because of some biological change, some of the children in this community start to be born as triplets. Should their parents think this is a disaster, because these children don't have twins? Clearly not. These twins don't have twins only because they each have *two* siblings. Since each child has two siblings, the trio must be called, not twins, but triplets. But none of them will suffer damage as an only child. These people should revise their view. What matters isn't having a twin: it's having at least one sibling.
>
> In the same way, we should revise our view about identity over time. What matters isn't that there will be someone alive who will be me. It is rather that there will be at least one living person who will be psychologically continuous with me as I am now, and/or who has enough of my brain. When there will be only one such person, he can be described as me. When there will be two such people, we cannot claim that each will be me. But that is as trivial as the fact that, if I had two identical siblings, they couldn't be called my twins.[105]

With this in mind, let us return to the notion of prudential concern as it relates to what matters. Parfit's view is that what makes the prudential concern one has for oneself tomorrow rational is not the fact that the person today who has the rational concern will be identical with that tomorrow-person who is the subject of the concern, but rather that the two persons will be psychologically continuous and connected in the way that Parfit describes. Through the split-brain case, he tries to show that what ought rationally matter to us when we consider future persons is not that the future person be numerically identical with us, just that the future person be psychologically continuous with us in the right sort of way.

Parfit himself acknowledges that in at least one case his view is difficult to believe. He calls it the Branch-Line Case. Imagine that a person enters a scanning device that creates a replica of the person inside after a button inside the device has been

[105] Ibid., 440.

pressed. Now suppose that the scanner does not destroy the original person upon completion of the scanning process, but it does damage his heart such that he will die within a few days. Suppose the replica assures the original person that he will continue the original person's life when he dies of heart failure. Parfit asks what should one's attitude be in such a case? Remember that, for Parfit, the crucial question is whether or not the original person and his replica contain what matters. In putting himself in the situation, he explains that the replica would be "…fully psychologically continuous, not with me as I am now, but with me as I was this morning when I pressed the button. Is this relation about as good as survival?"[106] Although he ultimately says that it is, Parfit does not give a direct argument for it. Instead, he draws analogies to this case from two other cases in an attempt to show that if the other two cases contain what matters, then so does this one.

The first analogy he draws is with a case called The Sleeping Pill. He explains that,

> Certain actual sleeping pills cause *retrograde* amnesia. It can be true that, if I take such a pill, I shall remain awake for an hour, but after my night's sleep I shall have no memories of the second half of this hour… Suppose that I took such a pill nearly an hour ago. The person who wakes up in my bed tomorrow will not be psychologically continuous with me as I am now. He will be psychologically continuous with me as I was half an hour ago. I am now on a *psychological branch-line*, which will end soon when I fall asleep. During this half-hour, I am psychologically continuous with myself in the past. But I am not now psychologically continuous with myself in the future. I shall never later remember what I do or think or feel during this half-hour. This means that, in some respects, my relation to myself tomorrow is like a relation to another person.[107]

Parfit finds this to be similar to the Branch-Line Case except for the fact that in Branch-Line his life overlaps with that of the replica's. He admits there is no analogue to this in

[106] Parfit, *Reasons and Persons*, 287.
[107] Ibid., 287-288.

The Sleeping Pill. However he claims the analogue can be found in a previous example he gave called My Physics Exam:

> I am taking an exam, and have only fifteen minutes left in which to answer the last question. It occurs to me that there are two ways of tackling this question. I am unsure which is more likely to succeed. I therefore decide to divide my mind for ten minutes, to work in each half of my mind on one of the two calculations, and then to reunite my mind to write a fair copy of the best result.[108]

In connection with the case we are currently considering, Parfit explains that,

> In both of my streams of consciousness, I know that I am now having thoughts and sensations in my other stream. But in each stream I am unaware of my thoughts and sensations in my other stream. My relation to myself in my other stream is again like my relation to another person.[109]

Since he can imagine having a divided mind, Parfit claims he need not assume that his replica is someone else. So just as one half of his divided mind in My Physics Exam is unable to know what the other half is thinking, so too are the original person and the replica unable to know what each other is thinking. Parfit is suggesting that we think of the two distinct persons in the Branch-Line Case as being two separate streams of consciousness of the same person like in My Physics Exam. As he says, "This description cannot be a factual mistake. When I talk to my Replica, this is merely like the communication in the Physics Exam between myself in my two streams."[110]

Parfit summarizes his position as follows:

> The actual case of the Sleeping Pill provides a close analogy to one of the special features of the Branch-Line Case: the fact that I am on a psychological branch-line. The imagined Physics Exam provides a close analogy to the other special feature: that my life overlaps with that of my Replica. When we consider these analogies, this seems enough to defend the claim that, when I am on the Branch-Line, my relation to my Replica contains almost everything that matters. It may be slightly inconvenient that my Replica will be psychologically continuous, not with me as I am now, but with me as I was this morning when I pressed the

[108] Ibid., 246-247.
[109] Ibid., 288.
[110] Ibid., 288.

> button. But these relations are substantially the same. It makes little difference that my life briefly overlaps with that of my Replica.[111][112]

Although he believes these two analogies are enough to defend his claim that what matters is intact in the Branch-Line Case, he admits that this case is one where it is more difficult to believe his view about identity not being what matters but rather just psychological continuity and connectedness.

We can slightly modify the Branch-Line Case to make it even more difficult for Parfit.[113] Suppose the original person would survive for ten good-quality years if he did not push the button in the replicator but the replica would survive for eleven years if he does push it (and the original person would soon die because of the damage to his heart). It seems that the rational thing to do would be to not push the button and enjoy the ten good years. Yet Parfit's view forces him to deny this. What Parfit must say is that the rational thing to do would be to push the button since the replica would live those ten good-quality years plus one more.

I have a case of my own where I believe it is very difficult to hold Parfit's view. Of the examples just described, it is most similar to the Branch-Line Case. Suppose that the scanner does not destroy the original person's body nor does it damage his heart, but instead, causes him to lose consciousness immediately after pressing the button. A perfect replica is created just like in the other examples, and it too is unconscious. We now have the original person and a prefect replica. Note that the original person has persisted despite the existence of a qualitatively identical replica. The replica is a mere duplicate and does not pose a threat to the persistence of the original person. Now suppose the two

---

[111] Ibid., 288-289.
[112] Note that Parfit does grant that if the overlap was large (his example is that of forty years), then this would make a difference.
[113] The following modification was suggested to me by Bruce Russell.

persons are taken into a room while they are unconscious and they each awaken at the same time. The last memory each of them would have would be that of pressing the button in the scanner. Neither would have any way of knowing which is the original and which is the replica. According to Parfit, there would be no reason, extraneous considerations aside, of course, for either of them to hope or prefer to be the original rather than the replica. Since each is psychologically continuous and connected with the original, there would be no reason to prefer to be the original as opposed to the replica since what allegedly matters is present in each of them. For Parfit, one should be indifferent regarding who is the original and who is the replica. I find this hard to accept. We can increase the stakes by adding that the duplicate will be killed. Should the two resulting persons remain indifferent upon hearing this news? Certainly not, but Parfit's view forces him to say that they should.

In light of cases such as the ones just described, one response is to infer that there must be something wrong with Parfit's argument. Many philosophers have made this inference and have constructed their own arguments to show where Parfit's goes wrong or to otherwise show that his conclusion does not follow. Peter Unger is one such philosopher. He believes that what matters most to us is indeed our continued existence. He offers a thought experiment to show that one's prudential concern for the resulting person of some procedure who is psychologically continuous and connected, but not identical, ought to be less than it would be if the person had persisted through the procedure. Unger has us imagine a case of what he calls 'century fission' where one's brain is divided into one hundred parts rather than just two as in the normal fission case (we are to suppose that each of the one hundred parts retains the entire content of the

original brain). Each of these hundred parts is then transplanted into empty heads resulting in one hundred persons all of whom are qualitatively identical with the original person as well with each other. Essentially, this case is exactly like the normal fission case; we just end up with one hundred persons rather than only two.[114] Unger argues that it is rational to take on less of a burden prior to century fission to prevent the torture of any or all of the one hundred resulting persons than to prevent the same degree of torture of one's unfissioned self at some time in the future. Unger concludes from this that one's prudential concern is greater for oneself than for a large number of persons all of whom are psychologically continuous and connected to, but not identical with, the original person. This seems to be a reasonable claim. I believe I would have less prudential concern for the one hundred resulting persons than I would for myself. If this is true, then Parfit's view is false; what matters to us is our continued existence.

Those who are not fans of fission cases (due to their being too unrealistic, or just plain incomprehensible) will likely object that century fission is even more difficult to grasp. Perhaps even those who have no issue with normal cases of fission might have a hard time imagining their brain being divided into a hundred parts all of which maintain the entire content of the undivided brain. Such a reaction will remove the majority of the force from Unger's argument. This is precisely the response David Hershenov has to Unger. As he says, "This failure of imagination is what accounts for their lesser concern towards the hundred fissioned slices than they each have for our own nonfissioned future self. So we should not draw any conclusions about identity from this."[115]

---

[114] Unger, Peter. *Identity, Consciousness and Value,* Oxford: Oxford University Press, 1991, p. 268.
[115] Hershenov, David. "Countering the Appeal of the Psychological Approach to Personal Identity". *Philosophy*, 79 (309), 2004, p. 458.

This strikes me as an odd response. Philosophers are generally not short of imagination. With our considerations of logical possibility, ingenious mad scientist scenarios, and other possible worlds, I do not see why it would be so difficult to imagine one hundred qualitatively identical persons existing all of whom are psychologically continuous and connected with the original person. This just means, in Parfitian terms, that each of the resulting persons will be R-Related to the original person. I agree that splitting the brain into one hundred sections all of which possess the entire content of the mind is farfetched, indeed, and likely not physically possible, but it is not logically impossible. Moreover, the difficulty in imagining the process does not mean we ought to have difficulty imagining the end result. That is, even if we have trouble imagining the possibility of the process, this does not mean we cannot imagine the mere existence of the one hundred persons. Hershenov then says,

> Our response to century fission is analogous to what our reactions might be if we lived for a thousand years…we would probably care less about what happens to us nine hundred years into the future than a year from now. While we would take on more burdens in the immediate present to avoid torture one year from now than we would to avoid the same torture nine hundred years from now, this lack of concern shouldn't lead us to believe that we would not be around nine hundred years from now.[116]

The problem here is that there may be ample reason to reject such biases about near future and far future suffering to be irrational. Parfit himself spends some time addressing this.[117] Hershenov finishes with, "Anyway, even if my explanation is inadequate, if our reactions to normal fission and century fission diverge, since the former is more realistic,

---

[116] Ibid., 458.

[117] To discuss it here would be beyond the scope of this dissertation. See Parfit, *Reasons and Persons*, pp. 165-167.

i.e., a closer possible world, it should be understood, pace Unger, as a more reliable indicator of our deepest convictions about personal identity."[118]

This is counter to the very reason we employ thought experiments in the first place. The point of them is to weed out all the constraints we face when imagining realistic scenarios. When considering a realistic case, there are many issues that may cloud our judgment in determining our deep rooted convictions regarding whatever issue we are examining. Considering a thought experiment allows us to cast aside all of these issues to allow us to dig easier, so to speak, to find out what our deep convictions really are. In short, I see no reason why our intuitions on a more realistic thought experiment should be more reliable than another thought experiment that is less realistic (provided that each are logically possible, which, in this case, I believe they are).

Regardless, even if Hershenov's response to Unger is correct, this will not help Parfit evade an argument like Unger's. Remember that what matters, according to Parfit, is psychological continuity and connectedness by any cause. The fission case is useful for those who only allow a legitimate cause; this is because in the single transplant variant of the split-brain case it is almost unanimously agreed upon (by defenders of psychological continuity theories) that this sort of underlying material cause would be a legitimate one that allows the original person to survive. But since Parfit allows any cause, we can run Unger's argument against a case similar to that of Parfit's Teletransportation as well.

The case is as follows: A person enters the Teletransporter. When the button is pressed, the scanner destroys the person's brain and body as it records the exact state of all the person's cells. It then transmits this information to the replicator device. This machine will then create, out of new matter, a brain and body qualitatively identical to

[118] Ibid., 458.

the original person's brain and body. Most will say that the resulting person will be a mere duplicate of the original and not identical with the original. But this person will still be psychologically continuous and connected with the original person; so for Parfit, what matters will be present. According to him, it does not matter that the original person has not survived. Now suppose a case of 'century replication' occurs where the replicator creates one hundred copies rather than just one. Here is where we may run Unger's argument. As Unger claimed, one would take on less of a burden prior to century replication to prevent the torture of any or all of the one hundred resulting persons than one would take to prevent the same degree of torture of one's unreplicated self at some time in the future. Unger's conclusion would follow from this just as it would in the century fission case, namely that one's prudential concern is greater for oneself than for a large number of persons all of whom are psychologically continuous and connected to, but not identical with, the original person. If this is true, then Parfit's view is false; what matters to us is our continued existence.

In my century replication case, Unger's conclusion would follow just as it would in his own century fission case, yet it avoids Hershenov's objection that the case is "too difficult to imagine". It certainly is not at all difficult to imagine the original case where the replicator only creates one copy. So simply imagine that it creates ninety-nine more. If you can imagine it once, you can just as easily imagine it happening ninety-nine more times. I believe that Hershenov's objection to Unger's argument is much too weak to succeed. But even if it does succeed, I have demonstrated that Unger's argument can be applied to a similar case that results in the conclusion he desires yet is not subject to Hershenov's objection.

However, there is another explanation that may be given in response to Unger's example that does not lead to the conclusion he draws that identity is what matters. Unger would have one believe that the importance of identity is what drives one's intuitions in his example. But perhaps it is only the sheer numbers. Perhaps it is just that since there are one hundred 'clones' of me, there is no reason to take on much extra effort to ensure that one does not cease to exist. After all, there would still be ninety-nine more should that one die. In normal circumstances, there is only one of me so I had better do all I can and take on every possible effort to ensure that I survive. This is a competing explanation of Unger's results that leads to a different conclusion than the one he reaches. If this explanation is just as good (or better) in explaining why one would take on more of a burden to ensure their own survival over the survival of one of a hundred clones, then Unger is unable to conclude that it is the importance of identity that drives our intuitions in this example.[119]

Another argument against Parfit is presented by Frederick Doepke.[120] Doepke's argument is based upon a notion he calls "direct control". This is the notion that one is able to do things simply by deciding to do them. He notes that in ordinary circumstances, one is only able to directly control oneself. Although one may control others by means of persuasion, deception and coercion, this is through "indirect" means, which are fundamentally different than simply willing to do something as is the case with direct control.[121]

Parfit claims that if he were about to undergo fission, he could intend (he actually calls it 'quasi-intend' since one can only truly intend one's own actions) what each of the

[119] This alternate explanation was suggested to me by Bruce Russell.
[120] Doepke, 83-91.
[121] Ibid., 84.

two resulting persons would do just by forming the relevant decisions. This is explained in the following passage:

> I could quasi-intend both that one resulting person roams the world, and that the other stays at home. What I quasi-intend will be done, not by me, but by the two resulting people. Normally, if I intend that someone else should do something, I cannot get him to do it simply by forming the intention. But if I was about to divide, it would be enough simply to form quasi-intentions. Both of the resulting people would inherit these quasi-intentions, and unless they changed their inherited minds, they would carry them out. Since they might change their minds, I could not be sure that they would do what I quasi-intended. But the same is true within my own life. Since I may change my own mind, I cannot be sure that I will do what I now intend to do. But I have some ability to control my own future by forming firm intentions. If I was about to divide, I would have just as much ability, by forming quasi-intentions, to control the futures of the two resulting people.[122]

A similar claim could be made about his replica. That is, in Doepke's terms, Parfit believes he would have direct control over both of the resulting persons from fissioning or the single resulting person from replication. As Doepke explains, "This is a crucial part of his argument against the importance of identity in survival, for it is supposed to seal the claim that his 'copies' are psychologically continuous with him."[123] It is the claim that Parfit would have direct control over his "copies" that Doepke denies. He points out that Parfit's claim is much more plausible if we imagine that the resulting person is unaware that they are a replica. Doepke asks us to suppose that,

> …*unbeknown* to me, I am about to be annihilated and replicated in my sleep. I go to sleep, deciding to apologize to a colleague for my rudeness, and the next day my replica begrudgingly carries out my decision. It seems that the relation between my decision and his action is just as direct as it would have been if it had really been me who made the apology; and it is because of this…that Parfit thinks that he is fully psychologically continuous with me.[124]

---

[122] Parfit, *Reasons and Persons*, 261.
[123] Doepke, 85.
[124] Ibid., 86-87.

However, Doepke argues that things greatly change when we suppose the resulting person is made aware of the fact that they are a replica:

> What will he [the replica] think–what will be the actual content of his thought–when he has a memory of something that I experienced or did? Suppose that he has a memory[125] of the impulsive decision to insult my colleague. Although he will no doubt *experience* it as I would have, the question is what judgment will he make in recalling it? Even though he will know what it felt like and how things looked, will he now express this knowledge by thinking, "I made an insulting remark"? It seems clear that he will not.[126]

From this, Doepke argues that upon realizing that his

> …apparently remembered actions and experiences really belong to a person distinct from himself, he will disown them. And this in itself seems to be practically important, for without claiming an action as his own, it is not clear how he can be made to feel responsible for it.[127]

Doepke then says a similar point can be made regarding the expression of one's will.

> My knowledge of which thing I am is expressed not only by what I remember, but how I currently feel, how things appear to me now, and also by what I intend to do.[128] Since I must regard these various "I" thoughts as directed toward a single thing[129], I cannot make decisions for my replica once I believe that he is distinct from me…But this means that I will not be fully psychologically continuous with my replica, which undermines Parfit's argument for the unimportance of identity in survival.[130]

Doepke has argued that the power of direct control is an intricate part of psychological continuity and connectedness. To which Parfit readily agrees, as demonstrated from the above passage. Doepke then shows that if the resulting person (or *persons* if we are imagining the dual transplant version of the split-brain case) were to be

---

[125] Although Doepke does not mention this here, strictly speaking this would only be an apparent memory. Since the replica is not identical with the person who had this memory, he can not really remember it. A similar remark can be made later in this passage when Doepke speaks of 'recalling' the memory. He does speak of 'apparently remembering' in the next quoted passage however.
[126] Ibid., 87.
[127] Ibid., 87.
[128] Doepke's footnote: John Perry has explained how "I" thoughts are necessary to exercise our agency in the objective world in "The Problem of the Essential Indexical", *Nous* 8 (March 1979): 3-21.
[129] Doepke argues against the denial that we are a single thing as presented by David Lewis and Harold Noonan. See pages 87-90.
[130] Ibid., 90.

made aware that he is not numerically identical with the original person, he would lose all sense of ownership over those memories–distancing himself from them. Doepke then argues that this abandonment would go beyond memories to encompass his thoughts, desires, and, most importantly for Doepke, his *intentions* as well. As Doepke says, "Once I believe that someone is numerically distinct from me, even if he is as similar to me as a perfect replica, I can no longer try to get him to do things just by deciding to do them."[131] That is, the relation of direct control is not present between the original person and the resulting person. But since this relation is necessary for complete psychological continuity and connectedness, then it follows that the resulting person is not fully psychologically continuous and connected to the original person. Hence, Parfit's argument that identity is not what matters fails because there is an important relation, that of direct control, that is present when one persists but is not present in the perfect replica. Where Parfit is considering the relation he has with his, as Doepke calls them, "copies", he asks, "Does this relation fail to contain some vital element that is contained in ordinary survival? It seems clear that it does not."[132] It is to this position that Doepke's argument is a response. According to Doepke, the relation of direct control is the vital element that *is* contained in ordinary survival but *not* contained in replication or fission. It is because of this that identity matters, according to Doepke.

It may be argued that despite what Doepke says, the relation of direct control is held between the original person and the fission descendants. Suppose that one intends (or quasi-intends) to have one's fission descendants do action X if conditions C obtain and do action Y if conditions C* obtain. If they do not change their minds sometime after

[131] Ibid., 90.
[132] Parfit, *Reasons and Persons*, 261.

fissioning, one's intending for them to do X in the event of C and do Y in the event of C* will lead them to do those respective actions. Is this not precisely how intentions work for one's own future? Suppose one intends that one will do X if C obtains and do action Y if C* obtains. If one does not change one's mind, will this intention not lead to one doing X in the event of C and Y in the event of C*? Of course there is always the chance that one's fission descendants will change their minds (perhaps as a result of what Doepke argues) but does this sever the relation of direct control? It does not seem that it would. The reason for this is that one may change one's own mind and thus, one may not do what one now intends to do. But this does not mean we do not have direct control over ourselves. Hence, it should not mean that one does not have direct control over one's fission descendants just because they may change their minds and not do what the original person (quasi-)intends them to do.[133]

As I explained in the beginning of this chapter, Parfit's positive thesis is to show what does matter, if not identity. I have given some cases where it is difficult to accept his view of what matters (Parfit's own Branch-Line case and my similar case). His negative thesis argues against the view that PI *in itself* gives us a reason to prefer identity over replication or fission. I have just finished giving some philosopher's arguments that attempt to show there is something wrong with this thesis and that identity gives us something that Parfit's positive thesis does not. However, we have also seen that there are problems with each of these arguments against Parfit.

At this point, it may have become apparent that one's intuition about what matters plays a large role. To illustrate this, I want to consider a series of cases all of which involve some difference of varying degrees of importance. I will first give all the cases

[133] This objection to Doepke was brought to my attention by Bruce Russell.

and then place them on a continuum of clearly mattering and clearly not mattering. The first is Parfit's twin/triplet example that was described earlier in this chapter. Let us remind ourselves of this case:

> Imagine a community of persons who are like us but with two exceptions. First, because of facts about their reproductive system, each couple has only two children, who are always twins. Second, because of the special features of their psychology, it is of great importance for the development of each child that it should not, through the death of its sibling, become an only child. Such children suffer psychological damage. It is thus believed, in this community, that it matters greatly that each child should have a twin.
>
> Now suppose that, because of some biological change, some of the children in this community start to be born as triplets. Should their parents think this is a disaster, because these children don't have twins? Clearly not. These twins don't have twins only because they each have *two* siblings. Since each child has two siblings, the trio must be called, not twins, but triplets. But none of them will suffer damage as an only child.[134]

The second case deals with robot love. Suppose there was a robot[135] that has been programmed to appear to perfectly exhibit human thoughts, emotions, reactions, etc. Upon forming a relationship with the robot, it would begin to simulate love gradually in a natural manner. Suppose, further, that the relationship one would have with the robot would be exactly the same as with some specific human person (the person cares for the robot just as much as they would a real person; this ensures the appropriate emotional attachment is there, etc.). In short, make everything in the relationship with a real person present in the relationship with the robot except for real love. The third case involves marriage.[136] Suppose one has a great relationship with one's spouse, but for some legal, or financial reason, the couple is forced to divorce. However, the couple still loves each other just as much, still has the same commitment to each other, still lives together, etc.

---

[134] Parfit, *The Unimportance of Identity*, p. 440.
[135] We can suppose it has a synthetic body with skin, hair, etc. Picture it to be physically indistinguishable from a real human being.
[136] This example was given to me by Bruce Russell.

In short, they still have the same great relationship, and it continues just as it would if the couple had remained married. The only difference is that the couple is no longer legally married. Finally, the fourth case is about happiness. Suppose one could have the option of being made happy (i.e., actually being in the psychological state of being happy) or merely being made to behave in a happy way.

Remembering to keep other things equal and putting extraneous considerations aside, let us line these cases up in order of how much the difference matters. From clearly not mattering to clearly mattering, I am inclined to order them as follows: twin/triplet, marriage/not married [insert a gap here for gray area] real love/robot love, real happiness/behaving in a happy way. It seems intuitively obvious to me that Parfit is right about twin/triplet. What the people in the community really care about (that is, what really matters), is that the child flourishes, not that the number of siblings a child has must be one rather than two. I also find it intuitively obvious that real happiness matters. It is not good enough to merely behave in a happy way. Having the psychological state of *being happy* is something that, in itself, matters. It is also apparent to me that truly being loved matters. It is not enough to have simulated love even if my life would be qualitatively identical to that where I am actually loved. Finally, the difference between being married or not does not really matter. What is really important is that the loving relationship one has with a person continue. If this continuing relationship is maintained along with every aspect about it, then it does not matter that the couple is not legally married.

But where ought Parfit's single/dual transplant (i.e., identity/no identity) case be placed? There are differences in intuition regarding where to place his example on the

line. Bruce Russell argues that his marriage case is very similar to Parfit's split-brain case. After all, in Parfit's example, both of the resulting persons maintain a continuing relationship with the original person (Relation R) although they also lose some of their properties (being identical with the original person). Similarly, the marriage example also involves a continuing relationship that loses some of its properties (the two are no longer each other's spouse) but maintains others (the couple still love each other in the same way after the divorce, still live together, etc.). Hence, since marriage does not really matter, this is an argument by analogy that we ought to place Parfit's example near marriage.

However, there is another example that is equally similar to Parfit's. Call it 'head injury': Suppose that due to a head injury, one's wife loses her ability to love, but she still remembers how to show and express love. Out of consideration for her spouse, she continues to show and express love (although she does not truly love her spouse, and is not truly expressing love; she is merely simulating it.) The couple live their lives together exactly as they would if she had never lost her ability to truly love. Just as Russell's marriage example involved a continuing relationship that loses some of its properties (the two are no longer each other's spouse) but maintains others (the couple still love each other in the same way after the divorce, still live together, etc.), so too does my example involve a continuing relationship that loses some of its properties (the presence of real love) but maintains others (the continuing relationship with the person that is qualitatively identical except for the wife's really loving her spouse). Initially, it seemed that since Russell's marriage case is very similar to Parfit's split-brain case, and since the difference in Russell's marriage case does not seem to matter, then neither does the

difference in Parfit's split-brain case. But since my head injury case is equally similar to Parfit's, and the difference in head injury *does* matter (since the presence of real love matters), then this example pulls the difference in Parfit's split-brain case in the other direction; that is, that it *does* matter whether or not identity is preserved.

What are we to conclude from this? I think we are to conclude what I suggested at the beginning: that intuition plays a large role in this issue and that since intuitions about Parfit's split-brain case differ, it is hard to know where to place it on the continuum. If this is true, then Parfit has not made his case that PI does not matter, even if I (or the other philosophers above) have not made the case that it does.[137]

Moreover, even if PI does not matter in itself, I think that one could have contingent reasons to choose single transplant over dual. That is, even if Parfit's positive thesis gives the correct answer in the Branch-Line and similar cases (despite our intuitions otherwise), and even if all the above arguments against his positive thesis do fail (which may be the case after considering the responses I gave to these arguments), there are certain considerations that I believe tip the scale and give one reason to opt for identity as opposed to mere psychological continuation and connectedness without identity.

When considering the split-brain case, I grant that there are certain considerations that can be easily discarded. For instance, concerns about which resulting person will get the original person's spouse, job, or house; such considerations are purely extraneous. We may simply stipulate that there will be no causal interaction between the two resulting persons, or that the original person did not have any of those things in the first

[137] My thanks to Bruce Russell for helping me clarify Parfit's argument and the relevant structural similarities between the cases just discussed.

place, and the concerns disappear entirely. Although some considerations are easily dealt with, not all of them are. What I shall argue is that even if we accept that there is nothing about identity *in itself* that matters, there are numerous considerations that could break the tie between the single and the dual transplant cases.

A great deal of these considerations can be found by considering that we have privileged access to our own mind. Although it is true that after a dual transplant, each resulting person would still have privileged access to their own mind after fission (that is, neither would be able to know what the other is thinking post-fission), they each would have access to the contents of their predecessor's mind pre-fission. Granted that the access each of these two resulting persons would have would be access to another's mind (since neither are the same person as their predecessor), still, this may be reason enough to want that such a procedure not occur. One may not like the idea that, after undergoing dual transplant fission, there would be two persons, neither of whom are identical with the original person, but each of whom would have complete access to the entire content of the original person's mind. Such a preference to avoid other persons from having access to one's mind provides numerous reasons why one might opt for the single as opposed to the dual transplant.

For instance, there are certain memories a person may have that they would not feel comfortable with anyone else having. I have in mind very personal memories that the person holds dear: one's first kiss, the first time one fell in love, losing one's virginity, the memory of one's wedding/honeymoon, holding one's newborn child for the first time, etc. Part of what makes these memories so special is that no one else has them; they are unique to that person. If both halves of one's brain are transplanted, they are no longer

unique to that person. One may not like the idea that two persons numerically distinct from him have access to such cherished memories.

Along similar lines, perhaps a person has memories of doing something that he is ashamed of. Suppose no one else knows that the person did this thing. It is reasonable to suppose that this person would not want anyone else to know that he did that shameful thing. It could be so shameful that the only thing keeping him from going insane thinking about it is the ability to push the memory out of his mind and "forget" about it since he is the only one aware of the act. Further, we may suppose that he finds comfort in the reassurance that no one else will ever find out about this shameful act. So when this person is faced with undergoing fission, he would have a reason to choose the single as opposed to the dual transplant. Again, this is because if this person were to undergo fission where both halves of his brain are transplanted, there would then exist two people, neither of whom is him, who know what he did.

There may be other reasons why one would not want someone else to have his or her memories. One may have memories of a terrible experience one endured and would not wish to subject or burden anyone else with those memories. Suppose one was in a concentration camp or in some other terrible kind of situation in which we think no sentient being should ever have to find themselves. If one has memories of such an experience, one would likely not want to burden any other person with those memories. It is bad enough that this one person has them and has to live with them every day. It is reasonable to presume this person would not want anyone else to have to deal with such painful memories. But, again, if this person were to go through a case of dual fission, there would be another person that has memories qualitatively identical to this person's.

As I said above, there are some considerations that are completely extraneous. Concerns regarding which resulting person will get the original person's things can be stipulated away by saying that there will be no causal interaction between the two or that the original person did not have any of those things in the first place. If we do this, these concerns disappear entirely, and hence, these considerations do not hold much weight. But simply stipulating that there would be no causal interaction between the two resulting persons does not eliminate any of the considerations I just gave. The mere existence of two other person's that would possess the entire content of one's mind is enough to make one not want the dual transplant to occur, for the various reasons I just explained. However, all of these considerations are still merely contingent. Just as we could suppose there is someone that would react in the ways I outlined above to a dual transplant fission case, we can just as easily suppose there is someone that responds in the opposite way. That is, someone may be relieved to find out there is someone else in the world that is able to understand them and share these joyous memories (or help take some of the burden of having the bad memories).

Let us examine one last consideration. Suppose one took a tremendous amount of pride in knowing that he achieved some great accomplishment. If this person were the subject of a dual transplant split-brain case, neither of the resulting persons would be able to take such pride because neither of them would be identical with the person that achieved the accomplishment. Being able to take pride in a past accomplishment is something we do not get out of a resulting person that is only psychologically continuous and connected with the original person; we only get it if the resulting person is *identical* with the original person.

I find this consideration to be the most difficult to dispose of. I admit that it could be true that neither of the two resulting persons want to take pride in anything. Or perhaps they do not even have anything they *could* take pride in. But do these responses make the problem disappear completely as it did above with the spouse, job and house considerations? No, because it is not just whether or not they want to take pride in something or whether or not they have something to take pride in. The point is *could* they take pride in some past accomplishment? Is it *possible* for them to do this? The answer is no. Regardless of whether they want to or not, or whether they even have something to take pride in or not, and even if there is no causal interaction of any kind between them, this is something that would be strictly impossible for either of them to be able to do, no matter what, in the dual transplant case. But it *is* possible in the single transplant case.

Although it is true that in the dual transplant variant, neither resulting person could take pride in *themselves* for the accomplishment, one may suggest that they could take pride in the original person in a way similar to how parents take pride in their children's accomplishments. After all, each resulting person is deeply connected to the person who actually achieved the accomplishment–they are psychologically continuous and connected to him–so I can see how it might be reasonable to agree that the resulting persons could take pride in the original person in such a way as described above.[138] But even if the resulting persons do take pride in the original person, this does not change the fact that neither resulting person *could* take pride in *themselves* achieving this accomplishment. This is still strictly impossible. Furthermore, being able to take pride in someone else's accomplishments is not the same as being able to take pride in one's own. This does not strike me as being any consolation.

[138] This was suggested to me by Bruce Russell.

Remember, none of these considerations I have just raised are intended to show that identity itself matters. In short, these are contingent or extrinsic considerations for why identity matters; they are not reasons why identity *in itself* matters. What these considerations do show, however, is that there are numerous contingent reasons why one may prefer the single transplant case over the dual one.

BIBLIOGRAPHY

Baker, Lynne Rudder. *Persons and Bodies: A Constitution View*. Cambridge University Press, 2000.

Butler, Joseph, "Of Personal Identity," in *The Analogy of Religion*, 1736. Reprinted in Perry 1975: 99–105.

Doepke, Frederick. "The Practical Importance of Personal Identity." *Logos: Philosophic Issues in Christian Perspective*, 1990: 83-91.

Frankfurt, Harry G. *The Importance Of What We Care About.* Cambridge University Press, 13th Printing 2007.

Frankfurt, Harry G. *Necessity, Volition, and Love.* Cambridge University Press, 1999.

Hershenov, David. "Countering the Appeal of the Psychological Approach to Personal Identity". *Philosophy*, 79, 2004: 447-474.

Kafka, Franz. *The Metamorphosis.* Kurt Wolff Verlag, Leipzig, 1915.

Kamm, Frances. "Jeff McMahan's Ethics of Killing: Problems at the Margins of Life." *Philosophical Review* 116, 2007: 273–280.

Locke, John. *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press (original work, 2nd ed., first published 1694); partly reprinted in Perry 1975.

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

Olson, Eric T., *The Human Animal: Personal Identity Without Psychology.* Oxford University Press, 1997.

Oyowe, Oritsegbubemi. “Surviving without a Brain: A Response to McMahan on Personal Identity”. *South African Journal Of Philosophy*, 29, 2010: 274-287.

Parfit, Derek. *Reasons And Persons*. Oxford University Press, 1986.

Parfit, Derek. “The Unimportance Of Identity”. In H. Harris (ed.), *Identity*. Oxford University Press, 1995: 13-45.

Parfit, Derek. *On What Matters*. Oxford University Press, 2011.

Peirce, Charles Sanders, “Prolegomena To An Apology For Pragmaticism”, *Monist*, vol.16, 1906: 492–546.

Perry, John. *Personal Identity*. University of California Press, 1975.

Reid, Thomas. “Of Mr. Locke’s Account of Our Personal Identity.” Perry, John. *Personal Identity*. Berkeley: University of California Press, 2008.

Russell, Bruce. *Memento*: *Memory and Personal Identity*. Unpublished, pp. 1-16.

Shoemaker David W., “Selves and Moral Units”, *Pacific Philosophical Quarterly*, 80 1999: 391–419.

Strawson, Peter. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.

Taylor, Charles. “The Concept of a Person”, *Philosophical Papers. Volume 1*. Cambridge University Press, 1985: 97-114.

Unger, Peter. “A Reply to Reviewers”. *Philosophy and Phenomenological Research* 52, 1992: 159–176.

Unger, Peter. *Identity, Consciousness and Value*, Oxford: Oxford University Press, 1991.

Warren, Mary Anne. *On the Moral and Legal Status of Abortion*. Monist 57:1 : 43-61, 1973. Reprinted in Mappes and DeGrazia 2001: 456-463.

Williams, Bernard. “The Self and the Future”. *Philosophical Review* 79, 1970: 161-180.

ABSTRACT

PERSONAL IDENTITY, SURVIVAL AND WHAT MATTERS

by

JAMES ALEXANDER GROMAK

August 2015

Advisor: Bruce Russell

Major: Philosophy

Degree: Doctor of Philosophy

Since the entire discussion of personal identity revolves around the identity of a person it is difficult to address these issues without presupposing that identity is maintained. In this dissertation, I propose an alternative approach to discussing the topic of personal identity (at least initially). This alternative approach is from the perspective of what I call '*continuance*'. I use 'continuance' to refer to some kind of 'continuing life' that is embodied in some person or persons. The term will be used as a neutral term for discussing the continuity of a person without any implications of identity. That is, in some cases, continuance will refer to a continuing life of one person but in other cases it will refer to a continuing life of some other person. In the literature on personal identity up until now, there has been no such neutral term. As a result, when considering the various cases, it is difficult to talk about the resulting person in those cases without presupposing that he is identical with the original person. The use of the term 'continuance' will allow us to talk of the resulting person without such presuppositions.

Most theories of personal identity can be grouped into two main types: physical and psychological. These theories place the determining factor of whether or not a person

persists on physical continuity and psychological continuity, respectively. We can address them in terms of continuance in the same manner. The question then becomes: what kinds of continuance are necessary and sufficient for persistence? I argue that neither a purely physical nor a purely psychological continuance theory are sufficient for persistence. Rather, persistence requires both a physical component as well as a psychological component. I argue that the physical substratum requirement is satisfied by continuance of at least part of the brain. Regarding the psychological component, I argue that memory, although sufficient, is not necessary. This is because I believe that what I call 'Frankfurtian continuance' (a continuance theory inspired by Harry Frankfurt's "self" involving higher-order desires and volitions) is also sufficient.

I then address split-brain (fission) cases. Most psychological continuity theorists take what is called a "non-branching" approach to split-brain cases. This allows them to claim that a person will survive if only one half of their brain is transplanted, but if both halves are, then neither of the resulting persons are identical to the original. My view is slightly different. I argue that branching is acceptable provided that transplanting the other half takes place after a sufficiently long period of time has passed.

Finally, I address Parfit's very influential work on personal identity. He argues that identity itself does not matter. Although I concede this point, I offer many extrinsic considerations for why identity matters. These considerations are not intended to show that identity *itself* matters, but they do show that there are numerous contingent reasons why one may prefer a scenario in which they do survive over one in which they do not.

AUTOBIOGRAPHICAL STATEMENT

James Alexander Gromak was born in Flint, Michigan in 1986. He grew up in Holly, Michigan and graduated with a B.A. in Philosophy with an Emphasis in Ethics from The University of Michigan – Flint in 2008 and with an M.A. in Philosophy from Wayne State University in 2013 and ultimately a Ph.D in Philosophy from Wayne State University in 2015.

His areas of specialization are metaphysics, specifically issues of identity (especially personal identity), ethical theory and metaethics. He is also interested in the philosophy of religion as well as the philosophy of horror.