

**Systems of Parts and Persons:  
A Functionalist Account of Personal Identity**

by

David McElhooes

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements of the degree of

Master of Arts

August 18, 2006

Department of Philosophy

University of Manitoba

Winnipeg

Copyright © 2006 by David McElhooes

**THE UNIVERSITY OF MANITOBA**  
**FACULTY OF GRADUATE STUDIES**  
\*\*\*\*\*  
**COPYRIGHT PERMISSION**

**Systems of Parts and Persons:  
A Functionalist Account of Personal Identity**

**BY**

**David McElhooes**

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of  
Manitoba in partial fulfillment of the requirement of the degree  
OF  
MASTER OF ARTS**

**David McElhooes © 2006**

**Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.**

**This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.**

## Acknowledgments

Many thanks to Tim Schroeder and Ben Caplan for sitting on my defense committee, providing extremely helpful comments, and for being the ones who originally helped coax these ideas from my head, giving me something to write about in the first place. Also, thanks to David Watt for agreeing to come to my aid at such short notice. I would also like to extend my gratitude towards Abi Klassen, Sam Cowling and Brendan Ritchie for a series of interesting, relevant, and helpful conversations. Finally, I'd like to thank the Faculty of Graduate Studies for their UMGF financial contribution, and also the faculty and support staff of the Philosophy department, for their patience and guidance, without which, I'd still be in business school. A terrifying prospect indeed.

## Abstract

The problem of personal identity generally involves answering two simple questions: what is our relation to our bodies, and how is it that we are able to persist through time? Some people think that we are identical to our bodies, and our persistence is thus a matter of our body's physical persistence. Others disagree. Instead they think that that our bodies act as mere vehicles, and we persist via the continuity of our respective psychologies. I argue here that both views get at least something right (but both also get something wrong). To account for the overlap I develop a hybrid theory which loosely relies on the notion of a system to explain how, and to what extent, both of these views are correct. Specifically, I maintain that people are consciousness producing systems and that we persist via our contribution to a particular output: consciousness.

## Table of Contents

Acknowledgments	I
Abstract	II
List of Figures	V
<b>Chapter 1: Introduction</b>	<b>1</b>
1. The Problem of Personal Identity	1
2. Pragmatic Considerations	4
3. The Solutions: A Very Brief Survey	9
3(a). Physical Persistence: Bodies and Brains	9
3(b). Psychological Persistence: Memories and More	13
4. A Different Approach	18
<b>Chapter 2: Persons and Psychological Character</b>	<b>19</b>
1. The Lockean Experiment: Princes and Cobblers	20
2. The Neo-Lockean Experiment: Doctor's Aliens	28
<b>Chapter 3: Bodies, Beings, and Physical Persistence</b>	<b>41</b>
1. An Appeal to Our Physical Existence and its Origins	42
2. Saving Substrate Persistence	49
3. Fission and Reduplication	58
3(a). The Problem	58
3(b). The Solutions: From Non-Branching to Best-Candidates	60

<b>Chapter 4: Systems, Persons, and Consciousness</b>	<b>71</b>
1. The Indeterminacy of Personal Identity	71
2. A Preliminary Reconciliation	75
3. The Systems Approach	79
4. Circularity and Reduplication Revisited	90
 Works Cited	 103

## List of Figures

Figure 1: Fission	95
Figure 2: Fusion	96

## Chapter 1

### Introduction

#### **1. The Problem of Personal Identity**

Imagine you wake up one Saturday morning in a groggy daze, likely due to the excessive socializing you enjoyed the night before. Everything seems normal at first; but, after a moment, confusion sets in. Your alarm clock, usually set to ultra-cool religious AM talk-radio is now, for whatever reason, set to the irritating buzzing noise you hate so much. Wait, it is not even your alarm clock. You do not even recognize the room. At this point, you might be thinking that you simply had “too good” a time and ended back at someone’s place for the night. But, catching a glimpse of yourself in a nearby mirror, you notice that you are not even the same sex; you’ve woken up in someone else’s body, or so it seems. In a raspy, throaty voice, you moan, “What happened last night?”

There are two possible answers. Either 1) you woke up in someone else’s body or 2) you woke up in your original body but with someone else’s memories and personality, so it very much *seems* like you woke up in someone else’s body. The choice between these two answers is the primary dilemma that lies at the heart of the problem of personal identity.

A related dilemma concerns our ability to persist through time. Clearly, you are the same person as the person sitting reading this right now. This fact is trivial. Not so trivial, but equally obvious, is the fact that your existence is not limited to this very moment. Not only do you exist now, but you also existed at some point in the past and, assuming that you do not die within the next fifteen minutes, you will also exist fifteen minutes from now. But how are you able to do this? Under what conditions are you the



same person now as you were in the past and will be in the future? One might think that resolving the first dilemma gives us an easy answer. That is, knowing the relationship between you and your body tells you how you persist. If you are your body (and just wake up with someone else's memories), you're able to persist because you have the same body.<sup>1</sup> If you are your memory or some other psychological thing (and thus really do wake up in someone else's body), you're able to persist simply because you remember being that earlier person.<sup>2</sup>

Unfortunately, things are not so simple. Bodies change through time. Can we really then say that you persist because you have the same body? Maybe, but we must first explain how bodies persist. Memories and other psychological states change through time as well. At this very moment in time, I, for example, cannot remember being an infant, my third birthday, or even my tenth birthday. However, at various points in time I was able to remember those things. Can we really say that I persist because I remember being (or stand in some other psychologically based relationship with) these earlier people? Maybe, but again it would take some explanation.

Even if we can produce a good account of bodily persistence or some account of psychological continuity to motivate the memory solution, there are further challenges. Consider the following science-fiction-style example: imagine you are fighting a life or death battle with the worst mad scientist ever. Just as you achieve victory over his horde of zombie research assistants and begin closing in for the final confrontation, he shouts, "You've defeated my zombie horde, but can you defeat... YOURSELF?" Surprised, you find yourself looking down the barrel of his infamous twinning laser.

---

<sup>1</sup> Those who give this sort of explanation include Ayer (1936), Thompson (1997), and Williams (1970).

<sup>2</sup> This is the view defended by Locke (1694), Grice (1941), and Quinton (1962).

The experience of getting hit by this beam is bizarre, to say the least. First, you feel yourself getting severed in two, right down the middle. Both halves of you hop away from each other using the remaining leg. One of the halves falls to the ground and lets out a cry of pain. Being severed, the other half feels none of this. Just before the half left hopping gets a chance to charge the scientist, both halves start to re-grow their missing side. The left half of you grows a new right side and the right half of you grows a new left side. Although the scientist obviously intended for the two halves to fight each other, presuming that one of the halves would be an “evil” twin as is so common in science fiction, he is wrong. The laser produced two perfectly (qualitatively) identical people, both with an equal distaste for mad scientists. As would be expected, the two team up and dispose of the villain in typical heroic fashion.

The camaraderie ends there, however. There is still the matter of which of the two gets to sleep in your bed alongside your spouse. Clearly *you* should be the one to sleep with your spouse, but which of the twins is you? If the mad scientist’s “evil twin axiom” held true, then it might be a simple decision: the “non-evil” version, being more psychologically continuous and equally body continuous, is you and should thus be the one to sleep in your bed. But the case described does not permit such an appeal, since both twins are equally continuous with the pre-twinning you, both in terms of psychology and body.

Tempting as it might be to suggest that they are *both* you, this hardly seems likely upon further reflection. In the midst of the climactic scene, one of the twins fell down and let out a cry of pain. If you are the twin who fell, then it is true that you felt pain upon falling. Uttering the sentence ‘I felt pain’ seems to be stating a fact. Yet, if you are also

the twin who remained sure-footed, then it is equally true that you felt no pain. Uttering the corresponding sentence 'I felt no pain' would be true at the same time as its negation. This, however, is a contradiction. Moreover, there is also an oft-cited "fact" about identity known as Leibniz's Law, which serves to prevent this sort of explanation (that you can both feel pain and not feel pain at the same time). According to this law, for any two things  $x$  and  $y$ ,  $x = y$  at a given time  $t$  if and only if  $x$  and  $y$  share all the same properties at  $t$ . Since only one twin has the property *has a new left side* while only the other twin has the property *has a new right side*, they do not share all the same properties at any time. So, following Leibniz' Law, the twins cannot be identical to each other.

These two cases thus highlight the main concerns surrounding our inquiry into the nature of personal identity. We have two main questions: first, what is our relationship to our bodies, and second, how is it that we persist through time?<sup>3</sup> If we had clear answers to these questions, we would be able to tell whether or not, in the first case, you woke up with someone else's memories or whether you woke up in someone else's body and, in the second case, which, if any, of the scientist-slayers is you.

## **2. Pragmatic Considerations**

At this point, an obstinate observer might ask, "Who cares?" Both of the scenarios outlined above are merely fictional; if it is fanciful sci-fi ponderings that bring us to question the nature of personhood to begin with, one might suggest that such questions would best be left to science-fiction writers rather than philosophers. It is important to note, however, that while some implausible science-fiction scenarios are indeed used to

---

<sup>3</sup> These questions are sometimes referred to as the synchronic question and the diachronic question, respectively.

motivate the debate about personal identity, the relevance of the debate extends far beyond the fictional realm. It actually touches on many concerns we have in the real world.

One of the most sought-after philosophical explanations in the history of our species involves the notion of life after death. What happens when we die? A great many people believe that we somehow live again and retain our original identity. For some Christians this means resurrection of the flesh: God will recreate our bodies and we will live on in these bodies eternally in a utopian society ruled over by God Himself. Conversely, for some Buddhists, Hindus, and others, the notions of rebirth and reincarnation are the popular forms of life after death. After we die, we are born again into a completely different body. After that body dies, we are born, yet again, into another form, and so on. By providing a rigorous analysis of personal identity, we will have all the evidence we need to determine rationally whether such theistic notions are really possible. If so, we can even likely explain *how* they are possible.

Another feature of the debate that has real-world relevance pertains to the notion of moral responsibility. There are those who seem to have multiple personalities, a condition medically referred to as Dissociative Identity Disorder (DID) or Multiple Personality Disorder. In some DID patients, the actions of one personality are independent of and totally unknown to the other personalities. If one personality, say, commits a murder, should the patient go to prison even if the personality responsible is dormant most of the time? What if it never appears again? Why would we punish a person who did not want to commit the crime, cannot remember committing the crime, and will likely not commit any violent crime ever again? Even if we choose not to punish

these people as we normally would if they were not DID patients, we must still ask, is this patient really responsible for the crime or is only that single personality responsible? If we had a clear view of personal identity, we would be able to tell whether or not DID patients are best considered to be a single person, or actually multiple people inhabiting the same body. If the patient is just one person then we know who is responsible: the patient. If many people inhabit the patient's body, then again we know who is responsible and it is just an unfortunate consequence that innocent personalities are forever chained to the guilty ones.

Another hot debate in contemporary ethics literature concerns the topics of abortion and euthanasia. In the former case, it is often argued that abortion is murder and it is, thus, morally reprehensible. However, the idea that abortion is murder presupposes the notion that the fetus is its own person, so it is something that *can* be murdered. Upon hearing a compelling analysis of personal identity, one can return to this question and test whether or not fetuses really are people. While there are, of course, other arguments opposing (and in favor of) abortion, the abortion-is-murder argument is, of all the arguments, the one that is articulated most often in public discourse. An inquiry into the nature of personal identity then seems to be quite relevant here.

Similarly, people often consider euthanasia or "mercy killing" to be unethical by reasoning that the life of a person is intrinsically valuable and should be preserved by virtually any means. Often, the terminally ill, unable to continue living by their own means, find themselves hooked up to life-support machines. If such a person is in a permanent vegetative state or is otherwise brain-dead, we commonly attempt to justify to ourselves the proverbial pulling of the plug, sometimes to make room for new patients,

other times to simply put the family at ease knowing their loved one is finally at rest. The most common of such justifications involves the suggestion that brain death and person death is the same thing. So, while it is true that we are terminating the life of the body, the debate is moot since the person herself is already dead. But is this really the case? Are brain death and person death really the same thing? To suggest they are the same presupposes some version of the psychological view of personal identity, since, if the body is the person, its termination would indicate the murder of that person. Though we might still be able to justify such action, it becomes much more difficult when the ethically saturated word 'murder' becomes appropriate for use in this context. A working account of personal identity would thus be a very useful tool in determining the ethical status of such a common medical practice.

Returning to the realm of science fiction, there are other issues that we can expect to crop up in the future given the rate at which the technological proficiency of our species is increasing. For years, science fiction novels and television shows have been toying with the idea of artificial intelligence and even matter teleportation – such as the famous transporter devices so often used in shows like *Star Trek*. Although it is still debatable whether or not these technologies really are possible, there seems to be some evidence that they are. This should not be surprising. It is often only a matter of time before science fiction becomes science fact. Airplanes and spaceships, for example, were once considered to be purely imaginary devices, dreamed up by the minds of hopeless visionaries and opium-addicted nineteenth-century novelists. Today, the possibility of such vessels has been realized and is thus hardly ever in dispute.

At this very moment, there are likely countless computer scientists all over the world hunched over their laptops, attempting to create the impossible: a thinking machine. But what are the ramifications of this? If we do create such a machine, should we consider it to be a person despite the fact that it is not a biological life form? If so, then it seems it should have the rights and privileges all people are entitled to. An account of personal identity would tell us whether or not, and under what conditions, a seemingly intelligent machine would be best considered a person. With such an account, computer programmers might be able to draw a line, creating an intelligent machine short of its being a person. In this way, we could enslave it, along with all of our other machines, without any loss of sleep stemming from the guilt associated with denying a person their basic rights. Robot revolutions be damned!

In contemporary physics labs, experiments involving the teleportation of photons – known as quantum teleportation – have been gaining attention. Although a long way off from its practical application as a viable means of personal transport, quantum teleportation suggests an encouraging and perplexing future for philosophers interested in personal identity. If we disassemble a person's body atom by atom, convert it to information and then use that information to rebuild an identical body out of different materials present at the destination site, have we really transported the person or have we simply created a clone (something like a three-dimensional photocopy) of the person? What if we use that same information to reproduce many clones at different destinations; does this mean that they are all the same people? What if the original disassembled materials that constituted the person are transported and then reassembled? Is this reassembled person the same person as the one who was disassembled? All of these questions hinge

on an answer to the two questions mentioned above: what is our relationship with our bodies; and given that relationship, how is it that we persist through time?

### **3. The Solutions: A Very Brief Survey**

#### **3(a) Physical Persistence: Bodies and Brains**

Philosophers have suggested many ways to tackle these two questions. The simpler solutions will sound familiar. Likely, the most basic and first solution that springs to mind is to suggest that personal identity is just a matter of bodily identity. This seems well motivated. When you introduce your new wife to your ex-wife, assuming the differences were amicable, you point to her body and politely say, “this is so-and-so.” So, in answer to the first question, the relation that holds between my body and me is the typical identity relation. Simply put, I am my body. In answer to the second question, I persist through time in the same way my body persists through time; and being a physical object, my body persists through time in the same way any physical object does.<sup>4</sup>

But what if someone removes my brain and successfully transplants it into a different body?<sup>5</sup> Considering that it is the new body that will possess all my memories and personality, it seems that I could not rightfully call my original body *me*. The new body would seem to be a more appropriate candidate. But, if in all cases I go where my brain goes, then, strictly speaking, I’m really not my body at all: I am my brain. The relation I stand in with my body is not one of identity; instead, it is a mereological relation – I am a proper part of my body. Again, since a brain is a physical object, to

---

<sup>4</sup> A slight modification to this case, where we are to be identified with some human animal instead of a body, is defended in Olson (1997). This modified view is sometimes known as animalism.

<sup>5</sup> Wiggins 1967 p. 50.



answer the second question, we can simply use the best theory that explicates the general persistence of ordinary physical objects despite changes in physical composition.

But what if someone removes my brain, divides it exactly in half and successfully transplants both hemispheres into two different bodies? If only one body wakes up after the surgery (with all my original memories and personality traits), then we would typically say that I still exist and it is the surviving half-brain that is the source of my identity. This is a problem. If having half a brain is sufficient for survival, then personal identity cannot be brain identity, since a half-brain is not strictly identical to your whole brain. Even more troublesome, if both bodies wake up (again with my original memories and personality), then both half-brains would stand in the same relation to my original self and both would then have equal claim to being me. If one half still resided in my original body (or even if one of the brain parts were larger than the other), this would be the intuitive tie-breaker. Unfortunately the scenario does not permit such a suggestion.

One response to the latter worry would be to suggest that, insofar as each of the two bodies possesses only a part of my brain, each body possesses only part of me. All that matters for my survival is that enough of my brain is sufficiently functional to act as the brain of a living human being. In this case, enough of my brain is functional to act as the brain of two living humans. So, although I am only a single person, I have two human bodies. As strange as this might sound, some might actually find it compelling.<sup>6</sup> For such people, the following analysis would likely be proposed: for any two people P1 and P2, they are the same person just in case enough of P1's brain survives in P2 to be the brain of a living human being. An added benefit of this modification to the brain view is that persistence of a functional whole brain is not required for personal identity. People who

---

<sup>6</sup> See, for example, Ehring 1986 and Parfit 1971 for a discussion.

lose half of their brains due to trauma, for example, are the same people as they were prior to their losses. Conventionally, this modified brain view is called *the physical view*.<sup>7</sup>

Bernard Williams has concocted a thought experiment that suggests that this modification is no help.<sup>8</sup> If we imagine a device that can capture and retain all the information stored in someone's brain and then transfer all of that information back into the same or even a different brain, we might start to doubt that even partial-brain identity is the proper analysis of personal identity. He reasons that, if one were to use such a device in order to house a person's mind temporarily while surgeons repair her brain, we would not question that the post-surgery person is the same as the pre-surgery person. According to the modified brain view, since the brain is the same, so is the person. Moreover, after the surgery, we cannot deny that she genuinely remembers her life prior to the surgery. But, in order to genuinely remember anything, it must be the case that the person doing the remembering is the same person as the one who formed that memory in the first place. To deny that the post-surgery person is the same as her pre-surgery counterpart is to deny that the information contained within her brain is genuine memory.

If they are not genuine memories, we must say that the post-surgery patient *learned* that information via the device rather than recalled it. This suggestion, however, carries the implication that the transference of information from the device back into the brain is akin to reading a diary or watching a home movie rather than genuine recollection; and this does not seem correct. Learning is typically a self-directed process and, furthermore, there is a distinct causal relationship between the events forming the original memory and the transferred memories. As such, it seems remarkably implausible

---

<sup>7</sup> More specifically, I am adopting the Noonan (2003) convention here.

<sup>8</sup> Williams 1970 pp. 180-1. It should be noted that Williams himself endorses a body-based view of personal identity.

that the post-surgery memory transference pre-empts the survival of the pre-surgery patient. Despite better judgment, experimental brain surgery is therefore nothing to worry about when concerned for your survival.<sup>9</sup>

A substantial problem arises, however, if we were to use the device to transfer the brain-state information into a completely different brain that had been transplanted into the person's body. Typically, we would say that this person has come to possess a new brain and is nonetheless still the same person. Moreover, if the memory transfer process is sufficient for transferring memories back into the original (yet renewed) brain, then it should do the same if it is an entirely new brain. Insofar as 'learning' is an inappropriate description of the transfer process when performed on the original brain, the same thing goes for the new brain, or so it seems. Consequently, the modified brain view cannot be correct. Here, the person persists but with a different brain. It therefore cannot be the case that even partial brain identity is personal identity.

If something as physically specific as the modified brain view fails, it would then seem that it is nothing physical that guarantees the persistence of persons. What, then, are our remaining options? In answering this question, let us first trace the apparent reasons for the failure of the aforementioned physical views. What seems to be the reason for the failure of the body view is that the brain is what is actually producing our consciousness and our awareness of our identity. The reason we must modify the brain view is that only part of the brain is sufficient to produce our consciousness and our awareness of our identity. Finally, the reason we are told to reject the modified brain view is that no part of the original brain is required for a future person to genuinely recall my past experiences.

---

<sup>9</sup> Assuming, at this point that the repairs involve only your original brain mass and that the surgery is successful.

Through the use of Williams' device, it seems that totally different brains are able to produce our consciousness and awareness of our identity.

It is not hard to notice a commonality shared by all of these reasons for rejection: they all involve the production of our conscious experiences. So, if personal identity has nothing to do with the persistence of anything physical, perhaps it has something to do with the persistence of something psychological.

### 3(b) Psychological Persistence: Memories and More

John Locke, who many would call the originator of the contemporary debate, maintained that it is only the sameness of consciousness that determines identity.<sup>10</sup> According to the orthodox interpretations of Locke's work, the way we are to account for such continuity is through experience-memory. This sort of memory is to be distinguished from factual-memory, i.e. memories pertaining to facts about the world. I can remember that  $2+2=4$ , that President Reagan was an actor prior to becoming a politician, and that I should not touch sensitive areas of my body after handling habanero peppers; but these sorts of memories are hardly constitutive of my identity. They bear no relevance to, nor do they play any role, in making me who I am.<sup>11</sup> Experience-memory – memories of my own experiences and actions – conversely, is quite relevant. Indeed it is this sort of memory that is "crucially involved in our awareness, from the first-person viewpoint, of our own identity over time."<sup>12</sup>

The most obvious problem with such an account is that there are gaps in our memories. As mentioned before, I cannot remember my tenth birthday, yet it would be an

---

<sup>10</sup> Locke 1694 p. 40. A detailed examination of Locke's position will appear in the next chapter.

<sup>11</sup> There are some, however, who might disagree. See Parfit 1971.

<sup>12</sup> Noonan 2003 p. 9.

offence to common sense to suggest that I was not ten years old at some point. Moreover, I can remember my eighteenth birthday, a day when I was able to recall my tenth birthday. So, I (A) am identical to my eighteen-year-old self (B), who is, in turn, identical to my ten-year-old self (C); however, since I cannot remember my tenth birthday, I am not identical to my ten-year-old self. Since identity is transitive, this conclusion leads to a contradiction: since the memory criterion tells us that  $A = B$  and  $B = C$ , it must be the case that  $A = C$ ; but the memory criterion also tells us that  $A \neq C$ .<sup>13</sup>

We must then reconcile Locke's analysis to account for these gaps so the contemporary version of me will properly turn out to be the same person as the ten-year-old version of me. To do this, we can simply say that two people P1 and P2 are the same person if and only if P1 and P2 are linked by *continuity* of experience-memory.<sup>14</sup> The word 'continuity' makes all the difference. It ensures that, as long as there is an overlapping chain of recalled experiences, I am the same person as all past instances of me. So, even if I can no longer remember my tenth birthday, I can still remember my eighteenth birthday. Since my eighteenth birthday was a day when I was able to remember my tenth birthday, it follows that the person who I am now and the person blowing out the candles on my tenth birthday cake are one and the same.

Experience-memory alone, however, might be thought to be inadequate. First, one cannot remember anything while asleep; and second, while memories are one type of psychological connection holding between a person and their temporally prior counterpart, there are other apparently relevant types as well.

---

<sup>13</sup> This is essentially a version of the Brave Officer argument found in Reid 1785 p. 114.

<sup>14</sup> This is the general strategy employed by both Grice (1941) and Quinton (1962) in their respective appeals to the notions of "total temporary state" and "soul-phase."

Beliefs, desires, and intentions are all able to persist through time. Yet, just as with memories, these can also change throughout time. It is nevertheless the case that we can trace a current belief, desire, or intention currently held by an adult back through time until its formation. In many cases, childhood experiences are what shape these psychological attributes. Since there is a causal relation between a childhood experience and its corresponding adult belief, desire, or intention that is analogous to the relation between childhood experiences and adult memories, perhaps one should include such continuities in the analysis of personal identity. As crucial as they may be, memories are not the only psychological factor involved in our awareness of our own identities over time. If, for no good reason, all of my most central beliefs and desires suddenly and drastically changed, then many would be inclined to say that I would no longer be the same person. If you are sympathetic to this inclination, then a slightly different analysis of persistence would be more favorable: P1 and P2 are the same person just in case P2 is *psychologically continuous* with P1. So, even if I have recently forgotten the events making up my tenth birthday, since I can still recall a day when I could recall my tenth birthday and since the beliefs, desires, and other psychological attributes I have now were partially shaped by the psychological attributes that I possessed on my tenth birthday, I am psychologically continuous with my ten-year-old self, and I am thus the same person.

Psychological continuity is not, however, without its fair share of problems. In fact, there are two main lines of objection to the psychological view of personal identity: reduplication and circularity. Though both of these issues have been touched upon in earlier sections of this chapter, more detail might be required.

If you recall the mad scientist case given at the start of the chapter, you recall a version of the reduplication problem. According to the psychological view, two people are the same if and only if they are psychologically continuous with each other, in which case every person standing in the *psychologically continuous* relation with my earlier self must be me. However, it is easy to imagine cases where we split my brain into two equally functioning halves and then implant each half into separate bodies (or we can simply imagine that I am shot with the mad scientist's twinning laser). Since it is my brain that produces all my psychological states and each of the separated halves is, hypothetically speaking, equally functional in terms of its respective similarity to the functioning of my brain as a whole, each half would be psychologically continuous with me prior to the split. Therefore, both halves, with consciousnesses and self-awarenesses that are now distinct from one another, are the same person. Yet, as we already know, there are problems with such a claim. If one brain-half feels hunger, the other does not, and they are the same person, then we have a contradiction. It is abhorrent for philosophers to claim that a single person can be both hungry and not hungry (or in pain and not in pain) at the same time without a good explanation.<sup>15</sup>

The circularity objection, first wielded by Joseph Butler when criticizing Locke's memory criterion, equally applies to the more sophisticated psychological views insofar

---

<sup>15</sup> Peter Geach (1967), in an attempt to avoid this explanation, instead suggests that personal identity is not numerical identity. With such a denial, both Leibniz's Law and the axiom of non-contradiction do not apply. While this might be an attractive solution to some, it seems to imply that our inquiry into the nature of personal identity is actually something like an inquiry into the nature of type identity. Two people, i.e. person tokens, are the same person if and only if they instantiate the same person type. Such a suggestion, however, does not make things any easier. We must now ask what is it that makes something a token of a person type to begin with? Is it a body or is it something psychological? More importantly, how is it that this token continues to instantiate this type throughout time? Answering these questions comes no easier than the originals; the problem has simply shifted from one set of tough questions to a nearly analogous set.

as memory still plays a crucial role in the persistence of any given person. Butler states his objection as follows:

And one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes.<sup>16</sup>

Basically, criticisms of this sort involve the charge that the psychological analysis is circular. That is, appealing to memory presupposes personal identity. Recall the experience of a first kiss. In order for a person, say, Sally, to recall her first kiss genuinely, she must have been the same person as she who partook in that kiss. Without this requirement, some other person, say Mary, who *seems* to remember that kiss (in a way that would make her believe it happened to her) would have an equally genuine memory, which is quite obviously not the case. Thus, genuinely remembering something presupposes the fact that you are the same person as the one who formed the memory in the first place. So, if the psychological view analyzes the sentence 'P1 is the same person as P2' as P2 being psychologically continuous with P1, since psychological continuity partially involves memory and memory presupposes identity, the analysis ultimately reduces to the following: P1 is the same person as P2 if and only if P1 is the same person as P2. The psychological explanation thus seems quite impotent. Insofar as the best physical view seems rather questionable, considering the potency of these two objections, the psychological view is hardly any better.

---

<sup>16</sup> Butler 1736 p. 100.



#### **4. A Different Approach**

Though one may still have intuitions that coax her to one of the two views despite these problems, the playing field is, at this point, somewhat level. So where do we go from here? While we could simply try to tweak our favored account in the face of these objections or maybe just bludgeon the opponent, I choose a different route. Instead of ardently supporting one or the other, in the chapters to come I suggest something much more radical: I suggest that, to a certain extent, both views are partially correct. In other words, I ultimately argue for a hybrid view. In the next two chapters I attempt to motivate this view by respectively arguing that both consciousness and physical persistence have something to do with the correct account of personal identity. Specifically, in Chapter Two I argue that a persisting consciousness is sufficient for personal identity; and in Chapter Three I argue that, notwithstanding the claim from Chapter Two, psychological persistence is not necessary for personal identity; and also that some sort of substrate persistence is necessary for personal identity. In Chapter Four, I appeal to general systems theory (theories about systems in general, as opposed to specific types of physical systems such as ecosystems, computers, cars, digestive systems etc.) in an attempt to reconcile these three claims in the face of a skeptical argument provided by Theodore Sider. The result is the hybrid analysis of personal identity. By phrasing the hybrid analysis in systemic terms, one would, I think, be giving the simplest and most intuitive analysis one could give of personal identity.

## Chapter 2

### Persons and Psychological Character

The previous chapter presented us with the basic description of and motivation for the two major views of the persistence of personal identity: a series of physical views and the psychological view. In this chapter I focus strictly on the psychological view.

Many of us find it pleasing to believe that the essence of our identity is to be found in our psychological character. Phrases like ‘beauty is only skin deep’ and ‘personality is what counts’ try to draw attention away from our physical existence, telling us that what is really important is the “person” inside. Admittedly, such talk cannot be plausibly used to come to some profound conclusion about our nature as persons. Yet it is difficult to deny that these phrases touch upon some strong, deep-rooted intuitions we have about personhood. When someone dies, most would be inclined to say that the person no longer exists despite the fact that the body remains. Such a belief is typically driven by a concept of personhood that ignores or rejects philosophical arguments for the importance of body. Insofar as this concept motivates the belief that dead people no longer exist, it also inspires many other beliefs. For example, people who sustain certain forms of brain damage can undergo such radical changes in personality that family and friends often report, “S/he’s not the same person anymore.” Referring to colloquial language use, however, is generally not a rigorous way to conduct a philosophical analysis of our various concepts.<sup>17</sup> Even though we often speak in a way that favors one of the views, such speech might very well be using concepts that are, in fact, mistaken. Philosophical argument is required when we want to clarify these

---

<sup>17</sup> Some people would disagree with this claim, as we will see in Chapter Four when discussing an argument given by Theodore Sider that appeals to the common conflicts in the usage of the word ‘person’.

concepts, that is, when we really want to get to the heart of the matter of personal identity.

In getting to the heart of the matter, at this stage I will be primarily discussing the original argument in favor of the psychological view: the prince and the cobbler thought experiment first formulated by John Locke. After critically discussing this thought experiment, I modify and (I think) strengthen it with an appeal to a continuous stream of consciousness. Drawing from the strength of this modification, I argue that continuity of consciousness is sufficient for personal identity. Within the context of this discussion, it will turn out that there is a distinction between psychological continuity and the continuity of consciousness. Since continuity of consciousness alone is sufficient for personal identity, psychological continuity is thus not a necessary condition for our persistence.

### **1. The Lockean Experiment: Princes and Cobblers**

As we already know, Locke's view of personal identity is basically a primitive version of the psychological view. Rather than being on psychological continuity, his emphasis is simply on recollection. That is, P1 and P2 are the same person if and only if one of them remembers being the other. Although this analysis is the most common interpretation of Locke, curiously, Locke's original phrasing says nothing of memory, but rather only of consciousness:

For, it being the *same consciousness* that makes a man be himself to himself, personal identity depends on that only, whether it be annexed only to one individual substance, or can be continued in a succession of several.<sup>18</sup>

---

<sup>18</sup> Locke 1694 p. 40. Emphasis added.

While Locke commits himself to memory in his appeal to consciousness as being the identity-preserving aspect of persons, there is no need to do so. You can analyze *sameness of person as sameness of consciousness* and avoid analyzing the latter with an appeal to memory. Locke is sympathetic to the inclusion of memory because of the definition of 'person' he gives at the beginning of his account:

to find wherein personal identity consists, we must consider what person stands for; which, I think, is a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places;<sup>19</sup>

In short, Locke believes that a person is a thinking substance with the capacity for reason and the ability to reflect upon itself as itself over time. He then continues by linking this definition with memory (or consciousness) to get his final analysis of personal identity:

which it does only by that consciousness which is inseparable from thinking and, as it seems to me, essential to it... for since consciousness always accompanies thinking, and it is that that makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things: in this alone consists personal identity... and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person.<sup>20</sup>

It seems quite plausible that memory, i.e. the aforementioned backwards extension of consciousness, would be required to be able to reflect upon oneself as oneself and would thus be required in an analysis of the persistence of personhood. It is important to note, however, that this definition applies not only to humans, but also to *any* substance that has these capacities. An extremely intelligent parrot, to use a classic example, would thus be a person according to this definition. As well, countless imaginable alien species and mythical gods, fictional or not, would meet these conditions and thus would either be

---

<sup>19</sup> *Ibid.* p. 39.

<sup>20</sup> *Ibid.*

fictional, or possibly real, persons.<sup>21</sup> Considering this wide definition of ‘person’ and its applicability to strictly possible entities, appealing to fictional, purely imaginative cases in discussions seems quite appropriate and perhaps even necessary. With this in mind, we are prepared to consider Locke’s famous argument, or at least its contemporary counterpart.

Suppose a prince goes to an evil alchemist one day after hearing rumors that the alchemist can transfer people’s souls into other bodies. Importing contemporary scientific knowledge to explain this “soul transfer,” we can say it occurs when the alchemist rearranges the neurons in a person’s brain so that they are functionally equivalent to the neural configuration of a different person, thereby producing qualitatively identical mental states. To prove to the prince that the rumors are true, the alchemist rearranges the neurons in a cobbler’s brain so that they are functionally equivalent to the neural configuration of the prince and similarly changes the neural configuration of the prince’s brain to be the equivalent of the cobbler’s. After the procedure, the person in the cobbler’s body believes himself to be the prince and the person in the prince’s body believes himself to be the lowly cobbler. The question relevant to the problem of personal identity is whether or not they are correct in their beliefs.

If one holds the traditional physical view, then they are not correct. The physical view – P1 and P2 are the same person just in case enough of P1’s brain resides in P2 to be the brain of a living person – dictates that there is no body-swapping going on in this case, since both the prince and the cobbler retain their original brains. Locke, however,

---

<sup>21</sup> While it is true that fictional persons cannot *really* remember anything, it is nevertheless true that they can *fictionally* remember events.

believes their beliefs are not mistaken; to demonstrate this, he gives what is taken to be one of the great arguments for this claim. Theodore Sider sums it up best:

Suppose the prince had previously committed a horrible crime, knew that the [body]-swap would occur, and hoped to use it to escape prosecution. After the swap, the crime is discovered, and the guards come to take the guilty one away. They know nothing of the swap, and so they haul off to jail the person in the prince's body, ignoring his protestations of innocence. The person in the cobbler's body (who considers himself the prince) remembers committing the crime and gloats over his narrow escape. This is a miscarriage of justice! The gloating person in the cobbler's body ought to be punished. If so, then the person in the cobbler's body is the prince, not the cobbler, for a person ought to be punished only for what he himself did.<sup>22</sup>

To evaluate this argument, we really need to consider only the last few sentences and the implicit conclusion. Let us then schematize it as follows:

- (P1) We would say that the cobbler's body deserved to be dragged away.
- (P2) If the cobbler's body deserves to be dragged away, then the cobbler's body houses the prince.
- (P3) If the cobbler's body houses the prince, then one of the psychological views of personal identity is correct.
- (C) Thus, one of the psychological views of personal identity is correct.

To respond, a critic might simply deny the intuition motivating (P1). According to this line of response, the guards made the correct choice. Merely seeming to remember committing a crime and then gloating over it does not make one deserving of punishment.<sup>23</sup> If the alchemist had been more subtle in his neural manipulation, he could have made it so that only one memory was changed. Suppose the alchemist changes only a memory of some crime the cobbler had once witnessed such that this cobbler now instead "remembers" himself committing that crime and then gloats over it. This sort of

---

<sup>22</sup> Sider 2004 pp. 6-7.

<sup>23</sup> We can only appeal to *seeming to remember* in our motivation: if the intuition is motivated by a belief that the cobbler *genuinely* remembers the crime, then one would be presupposing that the cobbler is the prince. But this is exactly what we need to establish in order to show the psychological view to be correct and is indeed what supporters of the physical view would deny.

thing could even perhaps occur naturally. If I, through a brain injury, were to come to believe that I am Jack the Ripper and take noticeable delight in the fact that I have been getting away with murder all these many years, it would seem rather unjust to condemn me to prison for a crime I only think I committed but actually did not.

One might suggest that, if there is a risk that I would pick up where the real Jack the Ripper left off, then I would indeed deserve to be dragged away. This suggestion thus highlights a distinction in our notion of desert. That is, there could be two different ways of deserving such action: by being guilty or for the prevention of future crimes. If the cobbler's body now houses the prince, then he deserves to be dragged away both because he is guilty as well as for preventative measures. And, if we read the scenario as the cobbler merely acting like a raving lunatic, he still deserves to be dragged away, not because he is guilty, but to prevent others from coming to harm, since he now has the personality of a sociopathic prince. So, even if the cobbler's body retains the cobbler's identity after the alchemist's meddling, there is a sense in which he still deserves to be taken away: as a preventative measure. Thus, even though the cobbler deserves to be dragged away, it does not follow that the cobbler's body houses the prince. This assertion, however, is in direct contention with (P2).

With a simple modification, this turns out to be only a minor problem. Considering the very last sentence in the quoted passage – “for a person ought to be punished only for what he himself did” – we can change (P2) to read, ‘if the cobbler's body deserves to be dragged away because it houses the guilty party, then that body houses the prince.’ The argument will, however, remain valid only if (P1) is similarly changed, that is, if (P1) is modified to read ‘we would say that the cobbler's body deserved to be dragged away

because it houses the guilty party.' But we now must ask whether or not the cobbler's body does in fact house the guilty party. Given the initial criticism of the argument, it seems that there is good reason to deny this intuition if the only motivation for it is that the cobbler *seems* to remember committing the crime and gloats about it; gloating and apparent recollection of criminal activity are insufficient as proof of guilt.

Things become even stranger if we consider further modifications to this scenario. Suppose the alchemist never changes the configuration of the prince's brain. Now there are two people gloating. But which body houses the guilty party? Since there was no change in the prince's neural configuration, it seems rather obvious that the prince's body continues to house the prince. Strictly speaking, that body then houses a person who is indeed guilty of the crime: namely, the prince. However, since only one person committed the prince's crime, only one person is guilty. Thus, only the prince's body houses the guilty party. But this presents a problem for the psychological view. No longer would we say that the cobbler's body houses the guilty person, yet, strangely, nothing has changed about that body. The only modification to the scenario is that the prince's brain lacks reconfiguration. However, guilt should not depend on brain surgery performed on someone else.<sup>24</sup> It would be supremely bizarre to claim otherwise. But this is exactly what one would be doing if one absolves the gloating cobbler body of guilt when the prince's brain retains its original configuration.

Taking a step back, let us suppose that the prince's brain is changed, except instead of simultaneous reconfigurations, the cobbler's brain is reconfigured prior to the prince's. After the alchemist finishes his work on the cobbler's brain, that person wakes

---

<sup>24</sup> This is basically what Harold Noonan is getting at when discussing the only x and y principle. See Noonan 2003 p. 129.



up with visions of the prince's crime. Until the prince's brain is reconfigured, according to intuition that only one person is guilty of the crime, the person inhabiting the cobbler's body cannot be considered to be guilty; the prince still exists in his body and is thus still the only one guilty of the crime. So, while the alchemist is working on the prince's brain, we would imagine that the cobbler-body would be pacing back and forth wondering whether or not the alchemist's procedure is going to be successful that second time. He knows that if it is not successful, and the prince's brain retains its original configuration, not only will he avoid prison, but, being religious, he will avoid an eternity in hell, for only the original prince would be truly guilty of the crime. If the alchemist is once again successful in practicing his art, however, then the person inhabiting the cobbler's body might worry that the only person who can stand guilty for that crime in the eyes of God is him. Basically, he would believe that the damnation of his eternal soul is dependent on whether or not the alchemist's "soul transfer" procedure is successful. Rejoicing at the alchemist's failure seems to be an odd way to celebrate your passage into heaven. Access to such a paradise should depend only on the actions performed by you, not on the actions performed by some evil alchemist.

But perhaps after the reconfiguration of the cobbler's brain there are then two guilty parties. To adopt such a response requires us to do one of two things: first, we can deny the intuition that, if only one person commits a crime, then only one person can be guilty of that crime; or second, we can claim that both guilty parties are in fact the same person, and thus both, being a single person, committed the crime.<sup>25</sup> Either way, the

---

<sup>25</sup> A third, rather counterintuitive, option would likely come from those who believe that a single body can house many different people when there is a split similar to that described by the mad scientist case at the beginning of the preceding chapter. For example, if you, at some point in the future, are split into a

cobbler's rejoicing at the alchemist's failure would then be unwarranted. Both the original prince and the person inhabiting the cobbler's body would be equally guilty and thus equally condemned to hell, if not prison.

However, it seems that the only way one could adopt this strategy would be to choose the latter option; it is hard to deny that someone is guilty of a crime if and only if he or she committed it. After all, when we say someone is guilty of a crime, it is generally intended to mean that someone is guilty of *committing* a crime. It is thus more or less analytically true that, if only one person commits a crime, then that same person, and only that person, must be guilty of committing it.

Claiming that the person in the cobbler's body is the same person as the person residing in the prince's body is just as problematic. Assuming that these two bodies are actually housing the same person, if we stipulate that the cobbler, being impoverished, had not eaten in a week, we would be compelled to claim that the prince feels hungry and does not feel hungry at the same time.<sup>26</sup> This, as we know, is a contradiction and we should thus be compelled to accept that the prince and the cobbler are distinct people despite the unsuccessful alchemistic practice.<sup>27</sup> If this is the case, however, there seems to be no reason to suppose that the cobbler is guilty. Locke's "soul-transfer" argument,

---

thousand different people with a thousand different bodies, then the lone body you have now is currently housing a thousand different people. See Lewis 1976 for more details.

<sup>26</sup> Presuming that the more affluent prince had a chance to feast after the perpetration of his misdeed, we merely have to ask the different bodies how they feel. One will report, "I feel hungry"; the other, "I don't feel hungry."

<sup>27</sup> One might suggest that this procedure instead results in one person with two brains (Parfit 1971). He feels hunger with one of his brains and does not feel hunger with the other. This response is not available to Locke, however. Each brain has the capacity for self-reflection and everything else that makes something a person (according to Locke). So there really are two people in the alchemist's shop. These people might be said to compose a third, but such an entity would not be able to reflect upon itself as itself since its two halves would be inaccessible to each other. Such an entity is not a person, by Locke's standard, or even common sense.

endowed with the explanation given from neurological science, then seems much less convincing.

## **2. The Neo-Lockean Experiment: Surgeons and Aliens**

One thing that must be acknowledged is that my description of Locke's thought experiment touches upon contemporary beliefs about neurobiology. In Locke's original description there is no mention of neural reconfiguration. Instead, he explains the body swap in terms much more familiar and acceptable to philosophers of his day: the transference of the soul. For Locke, this means that the prince's consciousness is transferred simpliciter, without neurological explanation. By ignoring the input of contemporary science, we might be able to strengthen Locke's argument with an explicit appeal to a continuous conscious experience. Doing so also allows us to avoid Locke's appeal to those deserving of punishment, which, as seen above, can get quite messy.

The easiest way for one to imagine that the body-swap occurs with an unbroken conscious experience is to refer to our common notion of out-of-body experiences. Although we might question the plausibility of such experiences in the actual world, many would admit that they are not mere illusions: they do occur or are at least possible.<sup>28</sup> According to popular description, these experiences make one feel like a disembodied consciousness, that is, a consciousness that is somehow aware of its surroundings without having a physical body. If such experiences are possible, then it should also seem possible for people from a different planet (or universe) to disembody

---

<sup>28</sup> If out-of-body experiences are not nomologically possible, then they still seem to be metaphysically possible. I will be assuming their metaphysical possibility throughout the discussion of the thought experiment.

themselves at will, retaining the continuous conscious experience. With such an alien in mind, we can tell a story similar to Locke's.

Suppose this alien is involved in a serious car accident, injuring herself to an extent requiring surgery. When the paramedics drop her on the operating table, she disembodies her consciousness. After floating around the room for a bit, she notices (through some sort of sensory modality possessed only by disembodied consciousnesses) that her body is damaged beyond repair, and since her species' ability to exist disembodied is limited to but a few minutes, she decides to invade the body of the head surgeon. As she lowers herself into the head of the doctor, she reconfigures his cognitive system – via an integration ability that comes naturally to members of her species – in such a way so that, instead of producing the doctor's consciousness, it now produces the patient's consciousness, thus re-embodying it.

As this integration process occurs, the alien begins to gain the usage of the doctor's natural sensory modalities: she begins seeing through his eyes, a moment later she hears sounds through his ears, and suddenly begins feeling the tactile sensation of a bead of sweat rolling down his brow. After gaining the remainder of the doctor's senses, she finally gains control of his muscles and is hence wholly present in the doctor's body. After the patient's consciousness is fully integrated into the doctor's body, she exclaims in the most professional manner possible, "I give up, the patient is dead!"

This exclamation, it seems, is a lie; the patient's body might be dead (because the brain is inactive, i.e., brain-dead), but we would intuitively say that the person lives on in the doctor's body. The strength of this intuition, or so I maintain, comes from the unbroken stream of consciousness that has moved between bodies. If the disembodied

consciousness had simply appeared out of nowhere and did not have the experience of leaving the patient's body, then some might question whether it really was the same consciousness. In such a case, the doctor's brain might have the memories and all other psychological attributes of the patient, but some might be tempted to say he is still the doctor. With a continuous stream of consciousness guiding the transition into the new body, it is guaranteed that it is the alien's consciousness, and hence the alien itself,<sup>29</sup> that resides in the doctor's body.

If, then, it were somehow argued that the patient's consciousness did not persist, despite being part of an unbroken stream, then we could not plausibly say that the alien is inhabiting the doctor's body; the consciousness now inhabiting the doctor's body would not be the alien's consciousness. In order to do this, however, one would have to argue that an unbroken stream is not sufficient for the persistence of a single consciousness.

One way to argue for this claim would be an appeal to the possibility of consciously willing yourself to change your psychological character. Suppose a person were to, through strength of will alone, change their neural configuration such that they have an entirely new set of memories, beliefs, intentions, and personality traits. Even if we assume that this person is continually conscious through this willful reconfiguration, one might assert that the change in neurons ushers in a brand new, willfully created, consciousness. Since none of the psychological and phenomenal elements that previously composed this consciousness have persisted, the consciousness itself cannot be said to have persisted. Its composing elements are completely different. So it appears that an unbroken stream of consciousness does not guarantee that the original consciousness persists. Thus, even the neo-Lockean argument is unsound.

---

<sup>29</sup> This fact is trivial since 'X's consciousness' presupposes the identity of X.

I must confess that there have been some changes to the elements that make up this hypothetical person's consciousness, but I'm not sure why this is supposed to be bad news for the neo-Lockean argument. Just because the elements that compose consciousness have taken on a different phenomenological or qualitative character does not mean that the consciousness is a different one; it only means that the phenomenological character of the consciousness is now different. But these qualitative differences are not sufficient for a claim of numerical differences. A house can be painted yet remain the same house, people's bodies can change their qualitative characteristics – for example, I look much different now than I did twenty years ago and with plastic surgery I can look completely different – but despite these changes we continue to believe that it is the same body.<sup>30</sup>

Moreover, it just seems rather odd to think that we can will ourselves to have a new consciousness. If I should have a power as the one described, I would consider it a change in my consciousness rather than the creation of a new one. I am not sure anyone could convince me otherwise, though admittedly I could just be stubborn. If I will my psychological states to change, each change is nonetheless a change in *my* psychological state. Since my psychological states must be part of my consciousness,<sup>31</sup> it follows that my consciousness persists throughout these changes. An unbroken stream therefore seems to guarantee that the consciousness persists.

---

<sup>30</sup> This line of criticism is similar to one given by Butler (1736) in reference to Locke's claim that an oak tree might survive a complete change in composition. Butler says it is only "in a loose and popular sense... the plant... [is] the same, notwithstanding its perpetual change in parts" (1736: p. 101), but in a stricter, and more important sense, it is not the same. I would deny, however, that the stricter sense is more important. Given that all people change over time, it is the looser sense that we are after when analyzing personal identity; otherwise there would be virtually no point in thinking about diachronic personal identity: people would simply not be able to survive any change in their parts, psychological or physical.

<sup>31</sup> It seems absurd to think that my psychological states can be a part of someone else's consciousness but not my own. I just cannot conceive of this being possible.

Even more relevant to the critic's position, suppose I take some sort of new super-hallucinogenic drug, call it super-LSD, and it completely alters my experience of the world in every conceivable way. Nothing is as it seems. Not even my beliefs and desires are safe from the effects of this drug. After ingesting this bizarre substance, I experience every last one of my memories, beliefs, and desires changing in reaction to this chemical; but, after the drug has taken its course, I am back to my normal self. Typically, we would say that the super-LSD has changed my consciousness, but has not created a new one, no matter how drastically different it might be (in terms of phenomenological character) while I was under its influence.

Though it has not shown the neo-Locken argument to be unsound, one thing that this line of reasoning is successful in doing is highlighting the difference between continuity of consciousness and psychological continuity: it seems a consciousness, and thus a person, can persist without there being continuity of psychological states. Psychological continuity is therefore not necessary for personal identity.

But, it might be said that this all assumes that the super-LSD user is the same person, and that without that assumption it certainly seems as if there is a different consciousness at work. I, however, am not merely assuming that the person persists. The super-LSD example serves to support both claims: that a single consciousness can persist without there being psychological continuity, and that a person can also persist without such continuity. This latter claim is not only intuitive, but also finds support in the following rationale.

Since the changes come slowly, the user is experiencing the changes in her psychological states. What is important is that each change is not sufficient to produce a

different consciousness. People's memories, beliefs, and desires change all the time; yet, this does not affect their numerical identity. If these gradual changes built up so that the new psychology bore no resemblance to the original, and the super-LSD user never changed back again, then it might not seem that there has been any persistence of a person or the consciousness. This, however, is irrelevant in the present case. In this thought experiment, the psychological landscape reverted back to the original once the drug had run its course, but not without a disruption in psychological continuity. None of the drug-induced states had anything to do with the psychological states of the person before the ingestion. However, these states did have something to do with the microphysical realization of the consciousness inside the brain. The drug produces a chemical interaction that results in a change in the chemical structure of the part of the brain that realizes the conscious experience. The consciousness has been altered by the drug but is nonetheless microphysically continuous during its effects. The consciousness is therefore continuous; it persists.

Furthermore, since the original psychology returns after the affects of the super-drug subside, it seems that the resulting person is the same as the original. Adding that the person seems to remember the experience of having all of their memories and beliefs change, but only after the affects have worn off, should make it seem clear that the person has persisted through the entire ordeal and has not simply popped out of existence for the duration of the super-LSD's effects due to being psychologically discontinuous with the original. Asking the user what the experience was like, she would probably report something similar to the following: "I suddenly started to forget everything, and then I started to believe some weird stuff, and then I started turning back to normal." To suggest



that the post-super-LSD user is a numerically different person than the psychologically discontinuous original is to suggest that this purported first-person account is false. But, intuitively, it seems like a perfectly genuine account of the ordeal.

Indeed, if one takes this criticism seriously, it only serves to support my thesis. If a critic denies my claim on the basis that, if the pre-drugged person is thought to be a different person than the post-drugged person, then it seems that they have a different consciousness, my sufficiency claim has been secured. The contraposition of this conditional amounts to the following: if it seems that they have the same consciousness, then it seems that the pre-drugged and post-drugged individuals are the same person; in other words, having the same consciousness *seems* to be sufficient for personal identity. Though this is a weaker claim since it appeals to how things seem to be, the imagined critic would nonetheless be supporting the main thesis of this chapter by challenging the aforementioned purported assumption.

I should also point out that a supporter of the physical view is in no position to object on the grounds that it is possible for a person to will herself a new consciousness. According to the physical view, sameness of brain matter guarantees sameness of person, which, in turn seems to guarantee the sameness of consciousness. This is because there are two ways in which we generally refer to our respective consciousnesses: at a moment in time – my consciousness being what I'm experiencing right now – or as something existing throughout time. When we talk about our consciousness in the latter sense, we talk about it being a single thing that extends from our first conscious experience to our last (and perhaps through periods of time in which it is dormant, such as when unconscious). For any conscious experience that I might have at any given time, it would

most certainly be a part of my consciousness; and, since 'my consciousness', in this temporally extended sense, simply refers to the series or set of experiences I have had during my time on earth, it follows that I can have only one consciousness. Indeed it seems rather silly to say things like, "I felt a sharp pain in one of my consciousnesses" or even "I am consciously experiencing having two consciousnesses right now."<sup>32</sup>

Even if there is a change in psychological character, the physicalist is then forced to admit that the unbroken stream of consciousness guarantees sameness of consciousness. To explain: if the physical view is correct, then, given the preceding paragraph, any conscious experience produced by a single brain must be part of the single consciousness we refer to as 'that person's consciousness.' However, if the critic is correct in her claim that a new, entirely distinct consciousness has been willfully created, she would be presenting a case in which there is a single brain, and thus a single person (given the physical view), producing two numerically different consciousnesses. But there is an incompatibility here. If any experience produced by a brain is part of a single consciousness, then either the critic's claim (that a new consciousness has been created) or the physical view itself must be false.

Even if it is the case that only, say, 30% of the brain is responsible for producing consciousness, it is nonetheless the case that the unbroken stream guarantees that it is the same consciousness. Recall that the condition for persistence is that enough of the brain survives to be the brain of a living person. Without the capacity for producing

---

<sup>32</sup> If one disagrees and says that a person can have two consciousnesses at the same time, one must be aware that that person would not have the ability to reflect upon itself as itself, since one consciousness would not be able to reflect on any of the psychological or identity-relevant attributes of the other consciousness. Such a person would be able to reflect only upon half of itself as itself (that consciousness might not even realize it is not the only consciousness possessed by that person). Since this "person" would not meet Locke's conditions for personhood, we would hardly call such a thing a person. Typically, we would call such a thing a two-headed monster. See Nagel 1971.

consciousness, even the enthusiast of the physical view should be compelled to admit that the person no longer exists, or at the very least that such a being ceases to be a person. In either case, transplanting the majority of the brain is insufficient for survival as a living person unless it includes enough of the consciousness-producing portion for self-reflection and rational thought. Moreover, if there is an unbroken stream, then clearly enough of the brain has survived to be the brain of a living person. If I am thinking about the various uses of plastic buckets, and someone comes along and removes the 70% of my brain not responsible for this thought process, *I* am nonetheless still thinking about buckets no matter where the majority of my brain goes. In any case, it also seems safe to assume that the mental basis for consciousness occupies the majority of the brain, preventing this objection in the first place.

Proponents of the physical view might then object to the neo-Lockean thought experiment on the grounds that it is too far removed from the real world for us to take seriously. Such a criticism seems, to me, to be rather unwarranted. Given Locke's definition of 'person', we are bound to admit that personhood is not a property restricted to humans (or entities inhabiting the actual world). Any analysis of personal identity that restricts personhood to humans is false. Hyper-intelligent parrots, for Locke, have just as much right to the label 'person' as any human. So, given this definition of person, the proper analysis of personal identity must apply equally to possible non-humans as well; yet non-humans might have a completely different biological or even spiritual structure. Our analysis of personal identity must cover, at the very least, all metaphysically possible persons.

Aliens could have radically different minds, brains, and could also very well possess souls with dramatically different properties from the properties of human souls, assuming that such things are metaphysically possible. If one feels that some sort of physical instantiation is required to support consciousness, this is not a problem for the thought experiment. We can suggest that the alien consciousness is some sort of pure energy state floating around in space. One could also suggest that the consciousness is perhaps realized by "ectoplasm" or an equally remarkable non-corporeal substance that otherwise has properties determinable by some metaphysically possible set of natural laws.

Furthermore, this thought experiment is, in many ways, quite similar to how some people view the actual world. If the possibility of physically based non-corporeal aliens is too hard to accept, there are other examples to support the "same consciousness, same person" thesis. There are some, for example, who genuinely believe in literal demonic possession. We might then attribute the disembodiment of the patient's consciousness to a supernatural demonic power as opposed to a natural ability of some alien species. The demon or alien lives on in the doctor's body and persisting consciousness is sufficient for persisting persons, whether the demon or alien kills the doctor in the possession process or not. The cases are parallel.

To explain: if such a demon (or alien) were to possess the doctor's body, then that body would house two different people, either at the same time or at different times. In other words, either the doctor survives this possession or does not. If the doctor does not survive, then she can have no further conscious experiences. If she were to continue having conscious experiences via this body, then she would consider herself very much alive, and it would seem silly to contradict her. However, if the doctor can have no more

conscious experiences, then whatever consciousness that body is producing is not her consciousness; that body is thus housing an entirely different consciousness: the demon's or the alien's.

Even if the doctor survives the possession, by either demon or alien, her body must nonetheless now be supporting two independent consciousnesses. This is because, if one is to survive anything, then one's body must still have, at least, the capacity for producing consciousness. If the body lacks this capacity, then one is generally considered to be dead.<sup>33</sup> However, if a body has the capacity for producing a given person's consciousness, then it houses that consciousness, which is what happens when we are asleep. So, if the doctor survives the possession, then both the doctor's consciousness and the demon's or the alien's are housed within the doctor's body. If this is the case, then the doctor's body houses two different consciousnesses during the time possessed; the consciousnesses are clearly distinct, since we can stipulate that the doctor is not aware of the demon or alien consciousness and vice versa.

So, if a body houses two different people either at the same time or at different times, it is nonetheless housing different consciousnesses (either at the same time or at different times). Contraposing this conditional secures our thesis. If a body does not house two different consciousnesses (neither at the same time nor at different times), then it is not housing different people (neither at the same time nor at different times). That is, if a body is housing the same single consciousness, it is housing the same single person. Sameness of consciousness is therefore sufficient for sameness of person.

While the above line of argument applies to both the alien case and its parallel demon example, consideration of demonic possession alone is a quicker way to

---

<sup>33</sup> Otherwise it would be possible for me to be identical to a corpse, a thought that many find disturbing.

draw the same result. Assuming that the doctor survives the possession, is conscious, and merely loses control over her body, I must point out that the only thing of the demon's that is persisting is its consciousness.<sup>34</sup> The demon meets Locke's conditions for personhood but has no physical presence of its own; it shares the doctor's body with another person: the doctor. Having no presence in our universe other than its consciousness, the demon example therefore shows that a persisting consciousness is sufficient for personal (in this case demonic) persistence.

If one wishes to reject the possibility of both demonic possession as well as the (perhaps more plausible) ectoplasmic alien, one would seem to be deciding on what is metaphysically possible based on her own opinions about personal identity. Even if our existing laws of physics do not support the existence of such aliens or demons, it seems possible that there is another set of laws that do support such creatures. If it is possible that there is another set of laws, then it would seem that these scenarios are indeed metaphysically possible. If the critic wishes to be stubborn on this issue, there is not much I could offer in response.

Granted that the pursuit for an analysis of personal identity must cover all metaphysically possible cases, in light of the above modifications to Locke's basic argument it seems that sameness of consciousness is sufficient for personal identity. The difficult part now is to explain what is meant by sameness of consciousness. Locke's idea is that memory of a past experience is both necessary and sufficient for a persisting

---

<sup>34</sup> If one suggests that there are other things persisting, we can simply stipulate that there are not. If we suppose that the demon either took some of the aforementioned super-LSD or willed its psychology to change, then psychological continuity would be disrupted; and if we deny a distinction between the demon's consciousness and its soul-like elements (if there must be such a thing), then we will have reduced the demon's persistence to its consciousness alone. Combining these two main thought experiments provides what I take to be a compelling case for my assertion that consciousness alone is sufficient for personal identity.

consciousness. However, as we know, there are problems with this account: circularity and reduplication, as described in this and the previous chapters. Those who take memory alone to be insufficient for personal identity, instead proposing a more robust view – namely, that of psychological continuity – must unfortunately face these same problems. Conversely, the physical view seems immune to, at least, the circularity problem. And it is to this view, as well its own version of the reduplication problem, that we shall now turn.

## Chapter 3

### Bodies, Beings, and Physical Persistence

In the previous chapter, I dealt almost exclusively with the psychological view, ultimately concluding that it gets at least one thing right: that continuity of consciousness is sufficient for the continuity of persons. In this chapter, I look at the issue primarily from the opposite perspective: the physical view.

As mentioned in the introductory chapter, proponents of the physical view traditionally maintain that the person persists just in case she has enough of her original brain to be the brain of a living person; that is, the persistence of personhood is a matter of strictly physical persistence. Despite the fact that the physical view is the topic under discussion, determining whether or not the physical view is correct is not the goal for this chapter. Instead, I argue only that an aspect of that view is true: that physical persistence is a necessary condition for diachronic personal identity. Even so, this holds only for physically based ontologies. Taking this into account, I then generalize this thesis, modifying the assertion: some sort of *substrate* persistence is necessary for sameness of persons, even if not the traditional sort of physical persistence.<sup>35</sup> So, while the psychological view gets one thing right, so does the physical view.

In the previous chapter I began by motivating the psychological view with an appeal to language use. The physical view also finds support in the way we talk about ourselves, however. If we run with scissors and sustain a nasty cut as a result, while we have simply cut our physical bodies, we generally express that we have cut *ourselves*.

---

<sup>35</sup> The possibility of demonic possession and other forms of non-corporeal conscious existence as discussed in the previous chapter seem to show that traditional body-persistence is not necessary for personal identity. I will return to this point later.



Cutting oneself and cutting one's body appear to be the same thing. Thus, colloquial language use suggests that we are our bodies. But this is unconvincing. When speaking, we generally employ very loose concepts of the things we are talking about. This is why people are prone to correction from others with greater expertise. For example, people often use the word 'ironic' in inappropriate contexts because of their loose grasp of the term. If only one of these ways of speaking – either the person-as-body or the person-as-consciousness – reflected a philosophically sound concept of the person, then one of these ways of speaking would actually be mistaken (indeed a commonly made mistake). So, being loose and ambiguous, colloquial language alone should not be used to support one view or the other. What one must do is attempt to find the correct analysis of personal identity through other means. If we are lucky, this account would preserve our colloquial usage; if not, at least it would give us a better picture of what usage is correct and what is not.

### **1. An Appeal to Our Physical Existence and its Origins**

One thing that is difficult for a dualist to deny is that we, at least in some way, exist physically. In the modern world, people not only are more sympathetic to the materialist ontology, but in many cases see it as being the only plausible ontology given what we know about the physics of our world. If everything in the world is physical, then we too must be physical, for clearly we are a part of the world. This does not entail, however, that we have strictly physical persistence conditions. Locke, for example, believed that we are thinking substances, but exactly which thinking substance we happen to be is to be determined by our consciousness.

A better argument would appeal, not just to the physical nature of the world, but rather to the type of physical thing we are. For example, with the exception of alien persons, we are human beings. Human beings, as strictly physical creatures, seem to have strictly physical persistence conditions. Considering this, it then seems to follow that the persistence of persons requires physical persistence as a necessary, if not sufficient, condition.

While this might be a better argument than the one derived from language use, many will nonetheless remain unconvinced. Just because we are human beings does not mean we are identical to our human body (or brain). Rather than meaning 'being that *is* a human body', the phrase 'human being' can simply mean 'being that *has* a human body'. Which human body we have (if any at all), at any given time, is perhaps subject to revision; Locke, for instance, would say that it depends on our consciousness. If we are not identical to our bodies, then it is possible that the persistence of our body (or brain) alone is not necessary for our personal persistence. That is, my persistence does not entail that the body I currently possess also persists. I could come to possess a different, non-human, body;<sup>36</sup> or, if I become a brain in a vat, I could come to have no body at all.

One might attempt to strengthen this argument by changing the wording. If we change 'we are human beings' to 'we are human animals', we can perhaps present a more compelling case. Since animals are biological things, they have strictly biological – that is, physical – persistence conditions.<sup>37</sup> Nascent fetuses, for example, are human animals and so are people in vegetative states. These sorts of creatures do not have mental states

---

<sup>36</sup> As we have seen from the previous chapter, it is metaphysically possible for some types of people, namely, aliens with the power of disembodiment, to possess a different, non-alien, body. Since we want an analysis of personal identity that applies to all people, not merely humans, this alien factor must be taken into account.

<sup>37</sup> This is the view of Olson (1997). In the next chapter, we will consider one of his arguments.

and thus cannot have psychological persistence conditions. Since these entities do not possess a consciousness, the only way they can persist is physically. Thus, if I was ever a non-conscious fetus, which it seems that I was at some point in time, then I must have persisted physically. In the absence of consciousness, then, physical persistence must be necessary for my persistence.

If one were to suggest that being an animal entails that we persist strictly biologically, one would be guilty of begging the question; one would be assuming that the 'are' in 'we are human animals' is used as a synonym for identity rather than as a copula. If 'are' is taken in the sense of identity, then one would simply be assuming the physical view. In effect, one would be saying that we are identical to some biologically persisting thing. This is, however, just a version of the body view: the most basic of all physically based theories of personal identity. Taking 'are' as a copula would allow for the claim that some human animals can persist via a persisting consciousness, since it would be merely contingent that we, as people, instantiate the property *being a human animal*. As described above, we have a human body, which is a biologically persisting thing, but again this does not imply that we are identical to some human body.

Furthermore, we might currently be human animals, but we are not necessarily so. Indeed, biological persistence itself seems too strong to stand as a necessary condition for our persistence as people. Suppose that when we were fetuses each of our biological cells was gradually replaced by a tiny functionally equivalent nano-robot. As the replacement takes place, we are gradually losing our statuses as biological creatures. At first we would

be hybrids (some kind of nano-human cyborg); but, once the replacement was completed, our existence would become strictly non-biological.<sup>38</sup>

However, dropping the necessity of biological persistence does not damage my main thesis in this chapter. This argument might move us away from the necessity of biological persistence but nonetheless continues to support a sort of physical view. Gradually replacing biological cells with nano-robots can be considered to be a sort of physical persistence. So, even if biological persistence is not necessary, some sort of physical persistence, provided that there is a lack of consciousness, is necessary.

Up until now, we have been able to conclude only that a sort of physical persistence is necessary provided that there has been no consciousness. This poses a deeper problem for the above arguments. All human animals and cyborgs might be physical, but this does not entail that in all cases people persist physically. After all, it is not necessarily the case that all people are animals or even cyborgs. Gods, angels, and demons might all be people, but none is an animal, nor is any of them composed of nano-robots.

Moreover, this animalist conclusion holds only in worlds with a strictly physical ontology. Within the realm of metaphysical possibility there could be non-corporeal animals, aliens, angels, demons, or gods. In the previous chapter, I dealt with a thought experiment that had an alien disembody her consciousness and replant it in the body of the doctor. Since the alien persisted without physical persistence, it seems that it must be the case that physical persistence is unnecessary for personal identity.

---

<sup>38</sup> If being non-biological entails being non-animal, then it is indeed the case that we are not strictly identical to some human animal. Rather *being a human animal* is a property that we just happen to instantiate. And if it is contingent that we instantiate *being a human animal*, then it cannot plausibly be the sole factor that governs our persistence. It is possible for us to simply stop instantiating that property.

However, even if the animal, demon, or alien is non-corporeal, it is nevertheless tethered to its universe somehow. In the previous chapter's main example I appealed to "ectoplasm" or some sort of mystical substance to explain the disembodiment. Though such substances are non-corporeal, they are nonetheless some sort of substrate, since they are the medium that enables the consciousness to exist. Even if it is metaphysically possible for a consciousness to exist without such a medium, this possibility does not discount the presence of a substrate. A substrate does not need to be considered as medium-like in its conception. These disembodied consciousnesses themselves facilitate and produce their own existences and as such are to be considered substrates in and of themselves.

From this, we can take a substrate to be simply that which facilitates the existence and production of a consciousness. To account for the various types of purported people listed above, we simply need to generalize this conclusion: some sort of substrate persistence is necessary for personal persistence. Biological and physical persistence might both be unnecessary, but substrate persistence is, on the other hand, quite necessary.

This is not only a claim that a physical view supporter *can* hold; it is one that they *must* hold. Given the definition of 'substrate' given above, anything that facilitates consciousness is a substrate. Since our physical bodies, especially our brains, facilitate consciousness, these physical things are, by definition, substrates.

A body view supporter might object by saying that, even if a body or brain stops facilitating consciousness, then you can still persist, but without a substrate. However, when a brain is no longer capable of producing conscious experience, we generally think

the person stops persisting. When consciousness is no longer possible, we take the person to be dead; and, when we are dead, we cease to exist. One could hold that substrate persistence is unnecessary on this version of a physical view; however, this would entail that we could be identical to our corpse and this is not something that I think many could stomach. So, if, as I think most would agree, we are to avoid being identified with our corpse, it seems that substrate persistence is necessary for personal identity.

It should also be noted that according to what I have been calling *the physical view* – the view that maintains that persistence of enough brain matter to be the brain of a living person is both necessary and sufficient for personal identity – substrate persistence must be necessary for personal identity. In order for a hunk of brain matter to act as the brain of a living person, it must have the capacity for producing consciousness. If it had no such capacity, then it would not be possible for the individual to reflect upon itself as itself or to have any conscious thoughts whatsoever; it would thus not be the brain of a person. As stated above, if a brain has such a capacity, then, by definition, it is a substrate. Therefore, if having enough of one's original brain to be the brain of a living person is a necessary condition, then substrate persistence is, by default, also a necessary condition for personal identity.

While the physical view proponent should be compelled to admit the necessity of substrate persistence in an analysis of personal identity, the psychological viewist might still try to deny this claim. Even if a disembodied consciousness must have a substrate (or must itself be a substrate), it is not necessarily the case that this substrate persists. For example, in the previous chapter I appealed to demonic and alien possession. As I had described it, it is possible that an alien or demon can disembody itself. This

disembodiment entails a change in the substrate. It might be said that while the alien once had a body as its substrate, it now has "ectoplasm" as its substrate. This ectoplasm, however, is not identical to the original body. Hence, it might be concluded that the original substrate has not persisted, although the person has.

Following this line of reasoning, the supporter of the consciousness-is-sufficient claim from the previous chapter (or a supporter of the psychological view) would likely jump in to tell the following story. At some point before you were born, you were not yet a person and, if you were to later fall into a vegetative state, you would then cease to be a person. So we could then say that, when you were a non-conscious fetus, you were a non-person animal with strictly substrative persistence conditions. Then, at the time when you became conscious and thus became a person, it was the persistence of your consciousness that determined your persistence. If you ever become vegetative, you would then again persist strictly physically. So, while your existence was once strictly physical, it could be said that it developed beyond the physical along with the progressive sophistication of your brain in your early nascent stages. You began as a non-person animal with strictly physical substrative persistence conditions, then moved to a person-animal with strictly consciousness-based persistence conditions, and could perhaps move back again if, as a person, you somehow sustain enough brain damage. Since this explanation accounts for persistence-as-people as persistence-of-consciousness, such an explanation seems, *prima facie*, to be completely compatible with the sameness-of-consciousness account of personal identity hinted at in the previous chapter.<sup>39</sup>

---

<sup>39</sup> This is all, of course, notwithstanding the conclusion from the previous chapter: namely, that psychological continuity is not necessary for personal identity.

But this *prima facie* compatibility would be neglecting one important thing. The question of personal identity, while concerned for our persistence *as* people, is more a question of our persistence *as us*. That is, our actual concern when inquiring into the problem of personal identity is the persistence of the thing I refer to by 'I' or 'me', not merely the thing I refer to by 'me' when it just so happens that I am also a person. Thus, the above explanation works against both the psychological and sameness-of-consciousness views in the following way: At some point in time I was a fetus. The word 'I', then, not only refers to me right now, but also to myself in the past; so, 'I' also refers to some fetus existing long ago. Since fetuses do not have conscious states, it follows that 'I' does not refer to something that persists by having a persisting consciousness (or by psychological continuity). In fact, 'I' refers to something (*viz.* a fetus) that had no psychological or even conscious states. By this line of reasoning, it in fact seems that both continuity of consciousness and psychological continuity are not necessary for our persistence, in the latter case adding support to the supplementary conclusion of the previous chapter. But, then again, substrate persistence equally seems unnecessary for our persistence, at least when there is continuity of consciousness to pick up the slack. Is this then the end of substrate persistence as an analysis of personhood?

## **2. Saving Substrate Persistence**

At this point it seems that substrate persistence is not necessary for the persistence of persons. After all, there is that strong intuition cited by followers of the psychological view that we can come to possess other bodies. These other bodies, it seems, are different



substrates. Being discontinuous with our original substrate, we are thus told to conclude that substrate continuity is not necessary for personal identity.

Stepping back from the notion of substrate persistence for a moment, recall the intuition that we were once fetuses. If we were non-conscious entities at one time and can possibly lack psychological features in the future (if we were to, say, become vegetative), then the only way it seems that we can persist in such scenarios is physically. So it appears that there is some tension here: in some cases of persistence, physical continuity does not seem necessary; but, in others, it does seem necessary. While it is in the next chapter that I seek to fully resolve this tension, some preliminary remarks can be made to support the central claim of this chapter until the full explanation can be provided.

First, given the conclusion of the last chapter (that sameness of consciousness is sufficient for sameness of person) we have a way to account for personal identity through time without an appeal to physical persistence. However, if there is no consciousness in one of the two entities to be analyzed, we tend to assert the sameness of those two entities nonetheless (for example, identifying my present self with the zygote it seems I once was). The only way to account for such sameness, however, is through some sort of physical continuity.<sup>40</sup> The same thing goes when accounting for the sameness of two non-conscious entities. So, if two entities existing at different times are the same entity, if they do not have the same consciousness, then they must be physically continuous in some sort of way. Thus, if there is no continuity of consciousness, then a sort of physical persistence is necessary for personal identity, at least in between body swaps.

---

<sup>40</sup> Though possibly not wholly through physical continuity, since I could have swapped bodies since becoming conscious.

However, even if there is continuity of consciousness, since we are presumably inhabiting a physicalist world, such continuity might also require physical continuity. If a world has a physicalist ontology, then consciousness (i.e. phenomenal experiences) must be realized by something physical. There is simply nothing else through which it can be realized. Even something as abstract as computer software is physically realized on some sort of electronic medium, whether it be a hard drive, floppy disc, or CD/DVD. If consciousness is something physical, then the persistence of consciousness is just the persistence of some physical thing. Thus, if there is sameness of consciousness, then a sort of physical persistence is necessary for personal identity. Therefore, whether or not there is a persisting consciousness changes nothing: under both conditions, a sort of physical persistence is necessary for personal identity.

One might object to this on the basis that a physical consciousness does not persist physically. However, if you consider some of the arguments from the previous chapter, it seems that physical persistence is a must. The only option besides a sort of physical persistence is psychological continuity. However, as we had seen from the super-LSD case in the previous chapter and in the case where one wills her psychological character to change, psychological continuity is not necessary for sameness of consciousness.<sup>41</sup>

Others might object on the grounds that *a sort of* physical persistence is too vague. At this point, I am inclined to agree. But this does not bother me, since I have no

---

<sup>41</sup> And considering the possibility of amoeba-like splits, such continuity does not seem sufficient for sameness of consciousness; otherwise, the people who result from such a split would have the same consciousness. But, since they have no access to one another's thoughts after the split, it seems counterintuitive to say they have the same consciousness.

intention of giving my full analysis here. A description of the sort of physical persistence I take to be necessary is therefore reserved for the next chapter.

Other criticisms might come from the claim that, even in a physical world, there are physical things that do not have physical persistence conditions. An object that exists for a single moment in time, for example, is something physical, yet it does not persist, so it cannot have any persistence conditions. Such an example, however, should not provoke too much worry. Temporally unextended objects could be said to have persistence conditions; they just do not meet them.

To reply, one might attempt to dream up an object that, by definition, does not persist. Since it is impossible for something that, by definition, lacks persistence to nevertheless persist, there can be no conditions under which that thing persists. We can therefore concoct a class of physical objects that do not persist, not because they do not meet their persistence conditions, but because there are no such conditions. However, this objection applies only to objects that, by definition, do not persist. Consciousness is a temporally extended thing. It is thus the sort of thing that does have persistence conditions. This reply has no effect on the persistence of consciousness. What is more, it could very well be the case that the concocted class of objects simply has conditions that are impossible to satisfy.

The most poignant objection to the argument from the physicalist ontology is that this assumption precludes a universal analysis. While in the actual world, if it is indeed strictly material (physical), an analysis of personal identity might have something to do with physical persistence, that analysis would not hold in dualist or idealist worlds. So the

analysis that the physicalist is proposing can be correct only if physicalism is necessarily true. But we want our analysis to be neutral on this claim.

This, nonetheless, poses a problem only if we fail to give an analysis in generalized terms: that is, an analysis in terms that are neutral on how consciousness is realized, whether physically or dualistically. I see, however, no reason why we cannot give such a neutral analysis. To do so, one would merely need to reintroduce the notion of substrate persistence. In physical worlds this substrate is physical; in non-physical worlds, the consciousness realizing substrate is non-physical; and, in dualist worlds the substrate can be both. So, even if a disembodied consciousness is possible, such a consciousness is merely self-realizing; it is itself the consciousness-realizing substrate. Thus, we merely modify the claim: some sort of substrate persistence is necessary for personal identity.

Now we are prepared to confront this issue of the ectoplasmic alien. It might seem that, by disembodiment, the alien patient has defeated the continuity of the original body. This, I maintain, is true, but it does not matter. What was problematic in the previous section was that the claim pertaining to substrate persistence was too strong. That claim amounted to the claim that the persistence of a single substrate is necessary for personal persistence. However, in this section the line of argument has resulted in a somewhat relaxed condition, namely, *some sort of* substrate persistence is necessary for the persistence of persons.

How does this solve the problem of disembodied alien? The answer is that you have to stop looking at substrate persistence *as* physical persistence. Just because the alien's substrate has changed from having a corporeal nature to having a non-corporeal

nature does not entail that it is a numerically different substrate. Substrate persistence is not simply the persistence of a specific type of substrate, such as a body or ectoplasm. Substrates themselves could be given different persistence conditions than specific instances of that substrate. One could thus reply that, in the case of the disembodied alien, the substrate had changed from having a physical nature to having a non-physical nature, but it is nonetheless the same substrate.

After all, the ectoplasm must have come from somewhere. The way the case was described in the previous chapter, the disembodiment and reimbodiment process is gradual. Indeed the alien's body had generated the ectoplasm that later serves as the substrate. If this is the case, then there is a sort of continuity going on here. The original substrate (i.e., the body) caused the alien's consciousness to be "squeezed" out of her brain via the release of this mysterious ectoplasm into the atmosphere. Such a causal relation seems to be sufficient for the persistence of a person's consciousness facilitating-substrate. Insofar as gradually replacing every molecule in my body with a tiny, functionally equivalent nano-robot is sufficient for physical continuity, gradually transporting my consciousness over to a non-corporeal substance should be sufficient for substrative continuity.

To clarify: while generating this ectoplasm, since it begins to facilitate the alien's consciousness, it seems that it is part of her while her body is also a part of her. If this mysterious substance is part of her, partially facilitating her consciousness, then her substrate during this disembodiment process, by definition, is part ectoplasm. Slowly, the ectoplasmic portion increases until the strictly physical portion of her substrate is lost. Thus, there is a transition of the substrate into a new form, rather than the creation of a

new substrate upon disembodiment. Just because the alien's consciousness becomes disembodied does not entail that it is realized by a different substrate.

To respond, one might attempt to present a case where consciousness seems to persist but where there is no gradual change to support continuity. If the ectoplasm were not generated, but instead just appeared out of nowhere or was produced by something else, then it might seem that there is no continuity of the substrate. This, I admit, is true. However, if this is the case, then there is no continuity of consciousness either. The consciousness might seem to be continuous, but that does not entail that it is continuous.

Suppose you are walking across a bridge. Once you are halfway across, the evil scientist from Chapter One shoots you with his infamous disruption-laser. The effect of this weapon is to disintegrate your body, brain, and very soul and then to instantly replace these things with functionally identical replacements by reorganizing particles taken from the surrounding area. Because this effect is instantaneous, the person who is now more than half-way across the bridge is completely unaware that you had been shot. From that person's perspective, he or she walked across the bridge without incident. It seems that here we have a case of a seemingly continuous consciousness but no substrate continuity.

However, suppose that there is a substantial delay in the reorganization process such that, from the perspective of the new substrate, day has instantly given rise to night. Now it seems less likely that the person, and the consciousness itself, have actually persisted. It seems that both a new person and a new consciousness have been created from elements taken from the surrounding area. Moreover, if your body had not been destroyed but had been thrust forward in time by five years, or had merely been transported to a hilltop a short distance away, in either case retaining consciousness

throughout the time travel or transportation, what seemed like a continuous consciousness to the newly formed thinking animal is not continuous. The continuous consciousness, intuitively, has traveled through time or has been transported to a grassy hilltop in the distance. The reason for this, it seems, is because that is where the substrate went.

In any case, even if the original body had been destroyed, neither the physical view, nor the psychological view would say that the person persists. The newly constructed body is not physically continuous in any sense, ruling out support from the physical view. Similarly, we can construct a case where the newly constructed body is clearly not psychologically continuous. If the rematerialization process depended on the configuration of the original brain in order to produce the seemingly continuous psychology, then one might suggest that the newly constructed person is indeed psychologically continuous and is thus the same person. However, we can remove this dependence by supposing that the rematerialization process produced a seemingly continuous consciousness purely by accident. If the consciousness seems continuous only because of some great coincidence, then the psychological states of the reconstructed being do not depend on your psychological states and they are thus not really continuous. Therefore, even if there is an apparent continuous consciousness, since there is no continuity of the substrate, we have reason to believe that it is an entirely new consciousness.

The burden of proof, then, is on the person who claims that the same consciousness can be transferred to a numerically different substrate without any substrative continuity. We have already encountered one argument that might purport to

do this: Williams's memory transfer machine from Chapter One.<sup>42</sup> Suppose your brain is to be repaired and its contents must be transferred to some sort of storage device. Insofar as transferring the contents back into the original brain seems more like genuine memory than learning, the same thing goes when transferring the contents into an entire new substrate. Sameness of substrate thus seems, according to this interpretation of Williams's thought experiment, to be unnecessary for survival.

In response, I suggest that there is an option that Williams does not mention. It could be that the contents are *programmed* into the respective brains rather than recalled or learned. If this programming is performed on or by the original brain, then it strikes me that this is sufficient for genuine memory. If it is an entirely new substrate, however, then this programming does not suffice for the apparent recollection to be a genuine one. Nor, however, does it suffice for the recollection to be a learned one; it is merely programmed.<sup>43</sup> Williams has presented us with a false dichotomy. Nothing is different about the device or transfer process if someone with a full set of memories were to become programmed with some of my own. However, it is hardly the case that such a person could be considered to genuinely remember events from my life. More importantly, it is not necessarily the case that she has learned about my experiences through the device. What has happened is simply that she has been programmed to seemingly recall events that had originally been processed and stored by my own brain. It is because it is the same substrate that originally processed and stored those seemingly recalled events that those recollections are genuine, not merely, as Williams seems to

---

<sup>42</sup> See Williams 1970 pp. 180-1.

<sup>43</sup> Programming is clearly not learning. When I program my robot, computer, or calculator, no matter how sentient they might seem to be, this programming has nothing to do with their sentience, and it seems to have nothing to do with their ability to learn.



maintain, because the learning alternative seems incorrect; otherwise, other people with their own memories could come to genuinely remember aspects of my life, since the learning alternative seems equally inapplicable. But, as we know, genuine memory presupposes personal identity, and other people who intentionally use Williams's device to gain access to my memories are clearly not the same person as I am. After all, I would never think to invade someone's privacy in such an intimate way. It seems then that substrate persistence is indeed a necessary condition for personal identity; otherwise various people could steal my memories and truly claim that they are genuine. However, since these other people are clearly not me, those memories cannot be genuine. With the necessity claim defended, one might now ask, "what about sufficiency?"

### **3. Fission and Reduplication**

#### **3(a). The Problem**

Provided that the result of the previous section is correct and substrate persistence is necessary for personal identity, one might be tempted to assert its sufficiency as well. One of the main tactics of the physical viewist has been to undermine the more popular psychological position by employing two powerful objections: the circularity objection and the fission (also known as the reduplication) objection. Since the circularity objection is generated by the psychological view's appeal to memory, it does not apply to the physical view and so cannot be used to attack its sufficiency. The latter objection, however, is a different matter. It is thus this objection that will be considered in detail here.<sup>44</sup>

---

<sup>44</sup> I also avoid a discussion of the circularity objection, since I consider the main response, which utilizes something known as quasi-memory in the place of genuine memory, to be adequate. Briefly, quasi-memory

The idea behind the fission objection is to present a case where a single person splits into two or more psychologically continuous autonomous people. There are many ways of doing this. You can posit something as basic as an amoeba-like split, the “twinning-laser” scenario presented in the first chapter, or even a case where a single brain is removed, split in two, and then the two separated hemispheres are transplanted into two different bodies.<sup>45</sup> While the earlier thought experiments involve a healthy imagination, the latter involves what we know is possible of the actual world. For example, we know from experience that it is possible to survive with one brain hemisphere. Tragic accidents sometimes destroy the functionality of one of the brain-halves, but the affected person is somehow able to compensate for that loss and continues to survive. Given this knowledge, it then seems possible for one of your brain hemispheres to be transplanted into another body and for it to be functional. And, if this is possible, then it should also be possible for both of your hemispheres to be transplanted into two different brainless bodies where each remains functional.

In this latter case, we would generally say that the transplant results in there being two people. However, each of these people possesses enough of your brain for it to be the brain of a living person. This is bad news for the physical view; for, if this is sufficient for personal identity, then these two people, occupying different spaces at the same time, are both, quite literally, you. Except, perhaps, for extreme cases where you travel back

---

can be defined as follows. A person P1 quasi-remembers some event e iff 1) P1 *seems* to remember e; 2) e actually occurred; and, 3) P1's seeming to remember e is caused in the right sort of way by someone's witnessing or doing e. It is the reference to 'someone' that removes the circularity since it no longer matters whether or not the original person really was the one who witnessed or did e. This, of course, changes the psychological account: P1 and P2 are the same person iff they are psychologically continuous in a way that utilizes quasi-memory instead of memory. See Shoemaker 1970, Parfit 1971, and Noonan 2003 pp. 144-62, for a more detailed analysis and evaluation. I will return to this psychology based account in the next chapter.

<sup>45</sup> Wiggins 1967.

through time and encounter a past version of yourself,<sup>46</sup> people generally occupy different spaces only at different times, thus experiencing different things only at different times. If this maxim did not hold, then, as mentioned in the previous chapter, one of the transplant survivors could feel hungry, while the other does not. Since such contradictions are impossible, it follows that the physical view (and its substrate counterpart) is not sufficient for personal identity.

One might suggest, in response, that both of these people are a mere part of you; that is to say, you are composed of two autonomous people.<sup>47</sup> The substrate view, however, is not equipped for this sort of response for the same reason Locke's view is not either. Both of the resulting half-brained bodies are indeed people (they can engage in rational thought and can reflect upon themselves as themselves). Moreover, they are distinct people; they cannot access each other's thoughts and go about their respective lives, perhaps never encountering each other.<sup>48</sup> However, if substrative continuity is sufficient for a later person to be identified with you, since both of these individuals are continuous with you, both are identical with you. There is no need to repeat the problems with this assertion. Another solution is required.

### 3(b). The Solutions: From Non-branching to Best Candidates

One way for the substrate account to avoid the absurd conclusions of one person being the same as two people who seem, quite clearly, to be distinct is to add a caveat to the account that explicitly prevents this sort of thing. This way, the obviously problematic

---

<sup>46</sup> For a more detailed argument from time travel cases, see Ehring 1986.

<sup>47</sup> Parfit 1971 p. 201.

<sup>48</sup> Parfit (1971) gives other reasons to reject this as an appropriate solution as well. See, for example, his footnote 8.

contradictions cannot arise. Such a move might *prima facie* seem ad hoc, but it, in fact, might be well motivated. Moreover, this is exactly what proponents of the psychological view tend to do.

Take what is known as the no-competitor modification to the psychological account, for example. As I write this, I am eating a Hot-Rod™: a pre-packaged stick of spicy pepperoni. Let us call this piece of pepperoni 'Rod'. As I eat pieces of Rod, depending on the size of the bites I take, it gradually gets smaller and smaller. Suppose I start feeling full after I consume exactly half of it and stop eating. What remains is a much smaller piece of pepperoni, call it 'Od'. If I began this paragraph eating Rod and end this paragraph holding Od, we would typically think that I have been eating the same piece of pepperoni for the duration of my writing. In other words, we would hardly deny that Od and Rod are the same piece of pepperoni. All that has happened is that Rod has developed into Od as a result of my gluttonous behavior.

With this in mind, consider what would have happened if I were not hungry and merely broke the original stick of pepperoni exactly in half. In my left hand, I would be holding a piece half the size of Rod, call it ' $R_1$ ', and, in my right, I would be holding the other half,  $R_2$ . While in the actual world we have no trouble asserting that Od is the same piece of pepperoni as Rod, we would not make the same identification in the world where I break Rod in half rather than eating it. Even when  $R_2$  is composed of exactly the same particles as Od (which is identical to Rod in the actual world), we would nevertheless be uncomfortable asserting that  $R_2$  is the same piece of pepperoni as Rod. The reason for this is the existence of a competitor that has an equal claim to being Rod:  $R_1$ . If  $R_1$ , as it was broken off, merely disappeared from existence, thus eliminating the competition, we

might say that  $R_2$  is Rod. But this is not the case. When we break the pepperoni stick into two pieces, we would not allow either piece to be individually identical to the original unbroken stick. So, while in the actual world it seems correct to say that Od is Rod when there are no competitors,<sup>49</sup> it does not seem correct to say that  $R_2$  is Rod when there are competitors, despite  $R_2$  and Od's being composed of the same molecules. Therefore, the existence of competitors is what is relevant when we decide whether or not to count the smaller piece as being the same object as the original stick of pepperoni.

If the existence of competitors is relevant to the identification and persistence of everyday objects, then it does not seem so ad hoc to impose this condition upon our analysis of personal identity. The result is the typically called the non-branching psychological view of personal identity,<sup>50</sup> but it can easily be modified to meet the needs of the substrate theorist.

P1 at  $t_1$  = P2 at  $t_2$  if and only if

- 1) Enough of P1's substrate resides in P2 to produce consciousness;  
and
- 2) there is no person P3 at  $t_2$  such that enough of P1's substrate also resides in P3 to produce consciousness.

But one might ask, what happens to P1 if there are competitors? Just as how neither  $R_1$  nor  $R_2$  is identical to the original piece of pepperoni, if there are two continuous people – that is, if there are competitors – then neither P2 nor P3 can be identical to P1. Since there is nobody else that can be identified as P1 at that time, it must be the case that P1 simply no longer exists. This occurs in all cases of physical reduplication, whether it is the result of half-brain transplants, amoeba-like splits, or being struck with the beam of a twinning-

---

<sup>49</sup> If, say, I took a really large bite of pepperoni, equivalent to exactly half of the original, then regurgitated it, then there would be a competitor in the actual world as well. In such a case, neither the regurgitated piece nor Od would be identified as being the same piece of pepperoni as Rod.

<sup>50</sup> See Wiggins (1967) p. 55 or Shoemaker (1984) p. 85.

laser. In any of these cases, if a person branches off into two continuous people, the original person no longer exists. This is why this modification is traditionally known as the non-branching view.

Though reflection on my pepperoni snack example might lead us to believe that the non-branching view is not ad hoc, it is still extremely problematic. It breaks a principle we all intuitively hold in regard to people. This principle, called the *only x and y principle*, can be explained as follows:

[the only x and y principle] is the principle that whether a later individual y is identical with an earlier individual x can depend only on facts about x and y and the relationships between them: it cannot depend upon facts about any individuals other than x or y. Otherwise put, what the principle asserts is that whether x is identical with y can only depend upon the *intrinsic* relationship between them, it cannot be determined *extrinsically*.<sup>51</sup>

If the non-branching substrate view is correct, then P1 is the same person as P2 only if there are no other, in this case, physically continuous competitors. However, if the lack of competitors is explicitly included in the analysis, then that analysis must maintain that P1's being the same person as P2 is to be determined extrinsically. That is, since the fact that there are no competitors is what determines P2's identity with P1, and since this fact has nothing to do with P1 or P2, nor has anything to do with the relationships that hold between them, it is a purely extrinsic fact that determines their identity. This seems to be a clear breach of the only x and y principle. Given the intuitive support for this principle, we simply cannot accept this modification to either the psychological view or the substrate view. While the non-branching modification itself might not be ad hoc, an outright rejection of the only x and y principle would be extremely ad hoc unless one

---

<sup>51</sup> Noonan 2003 p. 129.

were to convincingly argue against the principle. Without such an argument, the non-branching view seems clearly mistaken.

To demonstrate the intuitive support for the only x and y principle, we need only consider the criminal prince and innocent cobbler argument found in the preceding chapter. If the non-branching view is correct, then neither person resulting from the reduplication procedure is guilty of the crime. This is because, if there is a competitor, then the prince – the only one who committed the crime – no longer exists; so no person is guilty of the crime, even if the person in the prince's body thinks he is. However, if the reduplication had been a failure, then that same body (the prince's) would indeed house the criminal. Without a competitor, that person is the prince. Thus, if the non-branching view is correct, then being guilty of a crime, and indeed your very existence, depends on the failure of a medical procedure that has nothing to do with you (as it is being performed on someone else). But this is absurd. Reduplicating yourself should not end your existence, nor your guilt, any more than removing and transplanting the hand that was used to commit the crime in the first place. Indeed, the only x and y principle is difficult to deny.

In the face of such a bizarre result, one might suggest an alternative solution to the reduplication problem. Robert Nozick formulates two versions of this solution when presenting his *closest-continuer* theory:<sup>52</sup> the local version and the global version. According to the simpler local version, what makes a potential candidate the closest continuer, and thus the best candidate, is having

the highest degree of spatio-temporal and qualitative continuity with the original item, where qualitative continuity includes, for Nozick, the

---

<sup>52</sup> It should be noted that Nozick makes his modification for the benefit of the psychological view, whereas I am using it in an attempt to guard the sufficiency of the substrate, or (pseudo-physical) view.

existence of causal links between the qualities possessed by successive temporal stages of the item, i.e. the qualities of a succeeding stage must have developed out of those of a preceding stage in such a way that the succeeding stage would not have been the way it was (for the same reason) if the preceding stage had been different.<sup>53</sup>

So, according to Nozick's local theory, in half-brain transplant cases, if one of your brain hemispheres remains in your original body, then that entity is more qualitatively continuous with you than the transplant recipient. The person with your original body and half of your original brain is therefore you (according to Nozick's local version of the closest-continuer theory). If there are no closest continuers, in that nothing is qualitatively continuous with the original person or in the case of a tie between two competitors, the person then ceases to exist.

But what if the qualitatively closer candidate (i.e. the half-brained person with your original body) meets an untimely demise immediately after the transplant takes place? Many people might be inclined to say that the longer-lived candidate, though not the closest continuer on the local version, is nonetheless the original person. Nozick, one of those who believes it to be "so unfair for a person to be doomed by an echo of his former self,"<sup>54</sup> then suggests an additional complication, thereby forming the global version of his closest-continuer theory. He emphasizes longevity as a primary factor that governs persistence. According to Nozick's global version, the closest candidate is the item that is the

longest-lived item which, as a whole, is a sufficiently close continuer of the original item, and significantly closer than any comparable long-lived entity – even if some initial temporal part of that item is a significantly less close continuer of the original item than some other contemporaneous item of comparable duration.<sup>55</sup>

---

<sup>53</sup> Noonan 2003 p. 130.

<sup>54</sup> Nozick 1981 p. 43.

<sup>55</sup> Noonan 2003 p. 130.



Simply put, the global version is claiming that, for any competing entity that would be identified as the original person had the other competitor or competitors not existed, that entity is the closest continuer if and only if it exists for a longer duration than its qualitatively similar competitors.

Supplemented with Nozick's closest-continuer theory, the substrate view might now seem to be fully equipped to confront the only x and y principle. In any case where we want to identify two people P1 and P2, the supplemented analysis tells us that P2's being the same person as P1 has to do only with P2's being the competitor that is closest to P1. What makes this an excellent move on Nozick's part is that being the closest continuer seems to be a purely intrinsic fact;<sup>56</sup> P1's being the closest competitor to P2 seems to have to do only with the qualities possessed by P1 and P2. Since identifying a later person with an earlier one now seems to be determined intrinsically, it therefore appears that the physical view can borrow from the competition to avoid denying the only x and y principle.

In most cases, Nozick's modification does avoid a conflict with the principle; however, there are still some cases that pose a serious problem for his beefed-up psychological view (and also a substrate view that adopts the global formulation). Suppose the body of a person, call him 'Rob', is damaged beyond repair. Rob's brain is removed from his head, split exactly in half, and both hemispheres are simultaneously implanted into two new cloned bodies. According to the closest-continuer modification,

---

<sup>56</sup> Though 'closest' like 'tallest', 'biggest', and 'laziest' might seem extrinsic in that being closest depends on who else there is and what they are like, it is nonetheless the case that being identified with an earlier person is merely a matter of what properties you possess. Though some might still consider this to be a breach of the only x and y principle, it is one that seems less problematic. The absurdities derived from denying the principle are those such as where your existence depends on brain surgery performed on someone else.

whichever of these two people lives longer is the original person, namely, Rob. Now suppose that moments before Rob's body sustained the irreparable damage, he won a \$10 million lottery jackpot. After the transplant, neither person knows which of the two will live longer, and thus neither will know who is in fact the rightful heir to Rob's fortune (provided that, for whatever reason, the prize money cannot be split between the two of them). Unbeknownst to the two Rob candidates, both cloned bodies possess a severe genetic flaw. Because of this flaw, both bodies will simultaneously expire at midnight if they do not receive last minute surgery that, coincidentally, has a price tag of exactly \$10 million. In such a scenario, it would seem quite rational for either of them to attempt to murder the other. If one of the Rob candidates is successful in assassinating the other, then the candidate left standing, being the longest living candidate (and hence, the closest-continuer), is then to be identified as the original Rob according to the modified substrate (as well as psychological) view. He thus earns the \$10 million, saving himself from his body's genetic time bomb.

In this case, Rob survives only if he happens to successfully murder the other candidate. If he had not, since both bodies would expire at the exact same time, neither would qualify as the closest continuer – but this poses a problem. Survival of the fittest, in the Darwinian sense, is one thing, but this is something else. It is supremely abnormal to think that your survival depends on murdering a person who could very well be you if he or she only thought to kill you first. This is exactly the sort of thing the only x and y principle is meant to rule out.

Even more bizarre would be if both Robs thought to kill each other and their respective attempts occurred simultaneously. If we imagine a Russian-roulette scenario,

the survivor would have been so only by random chance. Therefore, which competitor is to be identified with you, is also decided by chance. While considering this by itself might prompt some skepticism about the closest-continuer theory, the major problem is this: the fact that determines Rob's identity in the Russian-roulette scenario is the position of the bullet in the chamber. This, however, is an extrinsic fact; it has nothing to do with facts about the original Rob, either of his competing offshoots, or the relationships that hold between them. The only x and y principle has therefore been breached. Furthermore, if the Rob competitors are unaware of the fatal defect of their cloned bodies, they would cease to exist at the exact same time. This, of course, entails (according to the closest-continuer theory) that neither competitor is Rob. As soon as the transplant occurred, Rob ceased to exist, and two entirely new distinct individuals came into being. The closest-continuer theory now seems remarkably similar to the simple non-branching view. As with the non-branching view, either of the two competitors would be Rob if the other had not existed. This again demonstrates a breach of the only x and y principle. The existence of the other competitor and the properties he instantiates is what prevents each from being identified with Rob, and this is an extrinsic fact. Despite the intuitive power of Nozick's theory, there is nonetheless a case where the closest-continuer account of personal identity breaks the only x and y principle in a way that cannot be ignored.

However, just because Nozick's psychological view breaks the only x and y principle in a problematic way does not mean that a modified substrate view will succumb to the same problem. The difference is that Nozick's account is a fully reduced analysis, while I have not yet given any persistence conditions for the substrate. All I have done so far is hinted that it has something to do with gradual change and the

potential for continuous consciousness. If I can give an analysis of persisting substrates that eliminates references to properties held by the various competitors (such that they seem to remember event *e*, or otherwise possess some psychological property), then it seems less likely that I would be breaking this principle in an intuitively acceptable way. There being no closest continuer would have nothing to do with any psychological property instantiated by anyone; the lack of a closest competitor would have something to do only with what happened to P1's substrate. It is not so much a dependence on which substrates there are as what they are like, but rather, where each substrate came from and if its origin is sufficient for it to be considered identical to the identical person. This seems to be an intrinsic fact about P1. A more detailed explanation, however, will have to wait until some persistence conditions for this substrate have been provided. This will be one of the goals in the next and final chapter.<sup>57</sup>

At this point, we seem to have strong support for three claims: 1) sameness of consciousness is sufficient for personal identity, considering the disembodied alien or demon example; 2) psychological continuity is not necessary for personal identity, considering that we tend to identify ourselves with some past fetus and that we can survive a disruption of psychological continuity either through the influence of some hypothetical drug or by willing such psychology to change; and 3) some sort of substrate persistence is necessary for personal identity, considering that we rely on a sort of substrate persistence to explain both the identities of non-conscious entities and

---

<sup>57</sup> It should be noted that the persistence conditions for the substrate must be different from those of typical physical objects if one wants to endorse a sufficiency claim. Reflecting back on the alien or demon case from the previous chapter, it seems that the traditional type of physical persistence is not sufficient for personal persistence either. If it were sufficient, then the doctor would continue to exist in virtue of having the same brain. Intuition tells us, however, that the doctor does not exist; rather, some demon or alien wholly occupies her body.

conscious entities. Accounting for all of these claims within a single lucid analysis is the primary goal of the next chapter.

## Chapter 4

### Systems, Persons, and Consciousness

In the previous two chapters, I presented conclusions that seem to be at odds: that both continuity of consciousness and some sort of physical continuity are essentially involved in the correct analysis of personal identity. Here I attempt to reconcile the results of these two chapters. I maintain that a systemic analysis of persons allows one to retain both these theses and is less problematic than either the psychological view or the physical view alone. Ultimately, I hold that two people P1 and P2 are the same person if and only if they are the same consciousness-producing system. A further analysis of this claim will follow from an examination of the classic objections to the psychological and physical views.

Prior to the explication of this account, I first present an argument given by Ted Sider for the claim that there is no correct account of personal identity. Specifically, he argues that the meaning of 'personal identity' is indeterminate and there is thus no fact of the matter. I use this argument as a stepping stone to motivate the systemic account I later defend. If such an account fails, then Sider might very well be correct, though there are still obstacles to overcome.

#### **1. The Indeterminacy of Personal Identity**

In his argument, Sider uses a certain analysis of 'meaning' in order to motivate his claim of indeterminacy.<sup>58</sup> Sider's argument can be schematized as follows:

---

<sup>58</sup> This analysis came about in response to an argument Hilary Putnam had given against realism. See Putnam 1981, Chapter Two.

- (P1) There are two competing theories of 'personal persistence'.
- (P2) These two theories fit use equally well.
- (P3) These two theories are equally eligible.
- (P4) Meaning is determined by use plus eligibility.
- (P5) No other theory fits use and eligibility as well as those two theories.
- (C1)  $\therefore$  The meaning of 'personal persistence' is indeterminate.
- (C2)  $\therefore$  There is no fact of the matter which view is the correct view of personal identity.<sup>59</sup>

At first glance, the argument certainly seems valid; its soundness is, of course, another question and we will return to it later. To briefly explain what Sider is getting at, let us examine the argument premise by premise. The two competing theories he refers to in (P1) are, of course, the psychological view and the physical of body view. In (P2) he is appealing to our everyday usage of the word 'person' as I had done at the beginning of the previous two chapters; sometimes we refer to people as psychologically persisting things, sometimes as bodily persisting things. Eligibility in (P3) deals with facts about the theory that are external to us. He suggests two ways in which a theory may be more eligible than another, though there might be more. "The winning candidate" might be one that is either

causally related (in the right way) to language users ... [or] the natural kind, or the most natural kind, that fits our meaning-determining behavior. Either way, the determination of meaning is not accomplished solely by us. Meaning is jointly determined by our meaning-determining behavior and facts external to us, whether causal relations between us and meanings or the intrinsic features of the meanings themselves.<sup>60</sup>

This is simply another way of stating (P4): that meaning is determined by use plus eligibility. What I take to be the crux of the argument is found in (P5): that there is no other theory besides the psychological view and physical view that stands as a better

---

<sup>59</sup> Sider 2001 p. 197. Note that the argument schematization here is slightly different from the original.

<sup>60</sup> Sider 2001 p. 191.

candidate than these two main competitors. This of course entails that there is no determinate meaning. Both meet the conditions equally well, and there is thus no fact of the matter. This is not, however, what we typically think about personal identity. We think there is, and ought to be, a fact of the matter.

To briefly comment on this argument without critically evaluating any of the premises, I wonder whether Sider's insights are made to reflect what the meaning of personal persistence is or what it ought to be. Consider an example I have used before: 'irony'. Listening to people speak, I have noticed that many use this word incorrectly. The sense in which they use the word is usually consistent between them, as it is a commonly made error. If both usages are equally consistent, 'irony' could just as well be associated with the "mistaken" usage. It might be less eligible in the aforementioned causal sense, but it is easy to imagine that one day the usage of 'irony' in the mistaken sense will be substantially more prevalent than the contemporary "correct" usage. Language use does evolve after all. If we imagine that increased incorrect usage cancels out the lack of eligibility, the fit between usage and eligibility would be equal between the two competing meanings. This would, of course, entail that the meaning of 'irony' is indeterminate. However, there is a set meaning. We merely have to look it up in the dictionary. A vast number of people are just getting it wrong. They are either unaware of the other definition or mistakenly believe that theirs is more eligible. For perhaps various reasons, one of which is to maintain maximum consistency, people *ought* to use the



traditional meaning; however, the way people *are* using the word renders its meaning indeterminate. Can the same not be said for one of the views of personal identity?<sup>61</sup>

I think it can. What we are trying to determine is the correct conceptualization: how we *ought* to analyze personal persistence. Sider's argument tells us only the *is* of the matter; specifically, how we currently view personal persistence. But this tells us nothing we do not already know: that we view the persistence of persons in two seemingly incompatible ways. Sider concludes from this that there is no fact of the matter, but I do not think this is necessarily the case. We might simply be ignorant as to what the fact of the matter is.

This then indicates one of two ways one might attack Sider's argument. First, one might reject (P3) on the basis that it only seems like both are equally eligible. However, this is because we have no dictionary-like source that we can appeal to for 'personal persistence' as there is for 'irony'. One view might be more eligible than the other; we simply lack the conceptual resources to discover which. If we knew what meaning we ought to use, then the meaning would not be indeterminate.

The other way to reject Sider's claim of indeterminacy is to reject (P5). This is the strategy I favor. I suggest that the two theories of persistence can be combined and are, in fact, co-dependent. The result is a hybrid theory. Since such a theory would take elements from both the physical and psychological views, it seems to meet the usage requirement quite well. Provided that it enjoys the benefits of both and suffers the problems of neither (or even just one), such a view would then also be more eligible than either of the two theories independently. The remainder of the chapter will provide an explication of such

---

<sup>61</sup> Similarly, some might claim that the increased usage entails that both meanings are equally correct. However, this would mean that there is a fact of the matter. Rather, there are two facts of the matter. Again, this seems to be an available option when thinking about personal identity.

a view, rendering Sider's indeterminacy thesis false. This is a desirable outcome. Intuitively, there is a fact of the matter of personal identity.

## **2. A Preliminary Reconciliation**

The final argument of the last chapter hinted at the sort of explanation I will ultimately be getting at. Recall that, at least in worlds with an ontology like ours (presumably physical), for any person's consciousness there is a physical structure or substrate that realizes it. We can call this physical structure a *c-realizer*. Though this c-realizer is often a brain, or at least a part of one, I call it this to reflect the contingency of its being a brain. If a person comes to have a different brain in the future, she nonetheless has the same c-realizer provided she retains the potential for conscious experiences; otherwise she would have no c-realizer at all. My c-realizer can thus be functionally defined as the thing that realizes my consciousness. This definition is functional in that it merely describes the c-realizer as the thing, whatever thing it might be at that time, that functions to produce a given person's consciousness.

The c-realizer is not simply a made-up, ad hoc entity. Right now, my c-realizer is my brain or is at least a part of my brain. My brain and its parts are neither made up nor ad hoc in regard to their relevance to my consciousness. What is more, just because the definition is functional does not directly entail a possibility that at one moment my c-realizer is my brain, and that at the next moment it can be some other person's brain. Persistence conditions for the c-realizer must first be given before one can claim what is possible and what is not.

Pertaining to these persistence conditions, it strikes me as being analytically true that, in a physical world, a person can continue to have conscious experiences if and only if her c-realizer persists. If my c-realizer did not persist, then I could not have any more conscious experiences since the c-realizer is functionally defined as that part of *me* that realizes *my* consciousness. The persistence of a c-realizer is thus a necessary condition for the persistence of consciousness. Additionally, if a c-realizer does persist, then, given its functional definition, so does the consciousness.<sup>62</sup>

The persistence of a c-realizer then seems to be a necessary and sufficient condition for a consciousness at one time to be the same consciousness as at another time. Granted that in physical worlds, the c-realizer is a physical thing, sameness of consciousness is therefore governed by the sameness of something physical.

One might interject here to point out that, as I describe it, the c-realizer is simply the human brain. Even if the c-realizer is defined functionally, in humans it is the human brain that performs this function. So at this point it might seem that I am merely leading up to an over-complicated version of the physical view.

I, however, maintain that, while the c-realizer is a single thing over time, it is not necessarily the brain. To say that different things can realize a person's consciousness at different times is merely to say that the composition of the c-realizer can change. That is, it can be composed of different things over time rather than being a numerically different thing from moment to moment. It is through gradual changes in composition that it might eventually come to seem to be a numerically different entity; and such changes are indeed

---

<sup>62</sup> Note that the functional definition does not necessitate that the c-realizer is active. That is, one does not have to be having a conscious experience in order for the c-realizer (and thus their consciousness) to persist. The consciousness simply becomes dormant.

persistence-preserving, as we see with everyday objects such as lamps, computers, or even something more abstract, such as a family.<sup>63</sup>

Even if it turns out that the c-realizer is a seemingly different thing at different times, this should not discount its persistence. Psychological view supporters believe that a person – a single entity – persists even though it can change bodies through mind-swaps and brain state transfers. Such persistence is, of course, governed by psychological continuity and thus not a sort of physical persistence as I am claiming for the c-realizer. However, we can modify an argument presented by Eric Olson that seems to show that a person can persist physically but come to have a completely different physical form.

Imagine you are diagnosed with a genetically engineered brain-wasting virus. Your doctors tell you that this virus will eventually cause your synaptic connections to break down and turn your brain into mush. There is one last recourse, or so they tell you: you can undergo a procedure utilizing nano-technology in which every biological cell (or even the atomic components of these cells) is replaced with a functionally equivalent inorganic counterpart. Nervous about such a dangerous-sounding procedure, you look hesitant. To put you at ease, the doctors tell you that you will, and indeed must, be conscious through the entire procedure so you can periodically tell them what you are experiencing.<sup>64</sup>

Later, the doctors tell you that the procedure was a success; however, there is a downside. The virus had been developed to respond to this procedure, evolving itself to

---

<sup>63</sup> It can be said that a family is a physical thing (or set of physical things). It is composed of you, your parents, your grandparents, and so on. A family, however, is not composed of the same physical things over time: people have children and grandparents die. 100 years ago, the physical composition of my family was very different from what it is today. It is thus, in a compositional sense, a very different physical thing. Babies, however, do not come from storks. Families persist through sexual reproduction and this is most certainly physical. So there are physical things that persist physically even though they are composed of different physical objects at different times.

<sup>64</sup> Olson 1997 p. 141.

affect the rest of the body as well. Since the inorganic entity that now functions as your central nervous system is immune to the effects of the virus, they transplant it into another unaffected body. Again you are conscious through this entire procedure. Afterwards, you stand from the operating table with a new body and a new inorganic brain.

Olson, an animalist,<sup>65</sup> denies the truth of this last sentence. He writes

The result would be a rational conscious being with your mind ... Despite appearances, the resulting being would not be you... Either that being is one of those people that aren't living organisms, along with gods and angels and rational electronic computers, to whom the [animalistic] view does not apply; or there is no thinking being there at all, but only thoughts and sensations that are not the thoughts of anyone.<sup>66</sup>

Such a response strikes me as being incredibly counter-intuitive. Given the result of Chapter Two, possessing the same mind (i.e. consciousness) is, at least, a sufficient condition for personal identity. Moreover, since you are conscious throughout the entire process it certainly seems as if the being that results from these two procedures is you.<sup>67</sup> Provided that this intuition is a good one, we can see that a single physically persisting entity – namely, your central nervous system – was once a biological entity but is now an inorganic entity.

---

<sup>65</sup> Animalism is a type of physical approach to personal identity that states that you are identical to some animal. This account is much closer to the bodily view than the physical view I have been promoting (whereby two people are identical if enough of the earlier person's brain exists in the later person's) in that it denies the intuition that we survive brain transplants, waking up in the new body.

<sup>66</sup> *Ibid.* pp 141-2.

<sup>67</sup> One might also ask, at which point is it not you? If after each replacement we destroy the replaced organic bit, eventually the remaining organic bits will not be sufficient to produce conscious experiences on their own. That is, they would not have the capacity for self-reflection unaided by the inorganic parts. At this point, then, it seems that according to Olson you would not be a person, but you would still be an animal; but there would be an inorganic person that shares some of your parts. I admit I am not sure what to make of this. In any case, I would be willing to go along with such a procedure if necessary, and I would not be concerned for my survival. This is, of course, just speaking for myself.

If c-realizers and central nervous systems are persisting functionally defined objects, then perhaps we should investigate the possibility of a person being defined functionally as well: as the entity responsible for the production of a particular stream or streams of consciousness.<sup>68</sup> But, if consciousness is physically realized, then part of the c-realizer, it seems, just *is* the consciousness; it is what is being produced rather than what is doing the producing. In this sense, consciousness is an effect of our bodies; it is a final product or even output. This is interesting. When talking about outputs, we are generally talking about systems. Perhaps, then, the answer to the question of personal identity can be stated in systemic terms.

### **3. The Systems Approach**

Systems theorist Oskar Lange defines a system as “a set of elements together with the set of relations between those elements.”<sup>69</sup> Prior to giving this definition, he suggests that, in physical systems, “the elements ... are linked by a set of cause-and-effect relations.”<sup>70</sup> Two or more elements that stand in these relations are said to be “coupled.”<sup>71</sup> Though not much is said about this (at least, in a clear way) in the systems literature, I take this coupling to be a type of regulated causal interaction between the systemic elements that generally produces some sort of output or joint effect.

Automotive assembly lines are systems whose outputs include cars, photocopiers are systems that produce a replica of an original document as an output, and stereo

---

<sup>68</sup> I say *streams* to reflect the intuition one might have that a single person can have multiple streams of consciousness. Split-brain cases are what motivate this intuition. See Tye 2003 pp. 109-33 for an explanation. We will return to this point shortly.

<sup>69</sup> Lange 1965 p. 17.

<sup>70</sup> *Ibid.* p. 1.

<sup>71</sup> *Ibid.* p. 11.

systems produce sound as their output. Even if a system does not produce what we would typically recognize as an output, the interaction of the elements nonetheless produces some sort of effect that is determined by that interaction. The solar system, for example, is a system that might seem to have no output. The interactions between the parts merely result in a consistent set of planetary orbits. However, whatever the result of this interaction, whether or not it is an output, we can call it the *system effect*.

Though Lange and other systems theorists<sup>72</sup> usually analyze systems in a strictly mathematical way, we can give a definition of 'system' that is more suitable for our purposes:

A **system** is a group of things coupled together in such a way that those things are poised to causally contribute to the production of a certain jointly produced effect.<sup>73</sup>

The nature of this effect is, of course, determined by the way the parts are coupled. If the parts were coupled in a different sort of causal relation, then we would expect that the sort of effect would be different in some way.

I use the word 'poised' here in the same sort of way Michael Tye does when giving his PANIC theory of phenomenal consciousness: that the elements "stand ready and in position to make a direct impact"<sup>74</sup> on the systemic effect. This allows that something's being a system does not depend on its actually producing the effect. That is, the cause-and-effect relation that allows for the coupling is not required to be active. If you turn your computer off it is not producing any output; however, it is still a system. If one removes all the parts from their case and disconnects them from one another, only

---

<sup>72</sup> Bertalanffy (1962) or Ackoff (1971), for example.

<sup>73</sup> Not all system theorists would be happy with this definition, since they tend to give much more general definitions such that this one may reflect only a certain type of system. Nevertheless, when I refer to a system, this is what I mean.

<sup>74</sup> Tye 1995 p. 138.

then do they seem to cease composing a system. Only if the parts can continue to interact with one another (say, via a wireless connection) would we say a computer system exists while disassembled.

This definition then also informs us that something, *x*, is a part of a system, *s*, if and only if *x* is poised to contribute to *s*'s system effect. If I install a new processor in my computer, it becomes part of the system as soon as it is capable of regularly interacting with the other parts in a way that is appropriate to producing a computational output.

So a system gains a part *x* iff *x* becomes poised to contribute to the system effect, and a part *x* is lost iff *x* is no longer so poised. If I add a welding robot to an automotive assembly line, or a step in a cookie-making recipe,<sup>75</sup> those things become part of the system only if the welding robot, or additional cookie-making step, is poised to contribute to the production of the output. If the robot is positioned in a way so that it welds nothing, that is, it has no effect on the car, then it is not part of the system. Likewise, if step ten of the cookie-making recipe is to jump up and down four times, it is a superfluous action that in no way contributes to the production of the cookie. It is not really part of the system.

One might be hesitant to accept any analysis of personhood based on systems, since it is terribly unclear what sort of causal relation is appropriate for two parts to be coupled. However, specifying the exact type of relation is simply not possible. The sort of relation depends on the system: it is system-specific. For a solar system, the causal relation would have something to do with gravity; and, for a computer, the relation would

---

<sup>75</sup> By 'recipe' I mean the process of baking cookies from scratch, rather than the piece of paper that process is written on.



involve an electric current. Furthermore, it is possible that, in order to produce an effect, a system might require more than one sort of causal relation.

In order to make things more comfortable, one might appeal to a general description of the relation similar to what Shoemaker does with the M-type causal chain in his analysis of what he calls q-memory.<sup>76</sup>

M-type causal chains should resemble as much as possible the causal chains that are responsible for actually remembering, i.e., should resemble them as much as is compatible with their sometimes linking mental states belonging to different persons.<sup>77</sup>

For systems, we can call the general relation an S-type relation. In a similar vein, S-type causal chains should resemble as much as possible those chains that are responsible for producing system effects of the specified type, i.e., should resemble them as much as is compatible with the production of effects of type-similar systems. For example, an S-type chain that is acceptable for solar systems would not be an acceptable S-type chain for automotive assembly lines. Furthermore, if there is more than one way to produce this sort of effect, then all of those ways are acceptable S-type causal chains for that type of system. Most importantly, however, the coupled elements are repeatable or regulated. That is, the objects relate “in a sufficiently regular way to justify attention.”<sup>78</sup>

That said, one might ask, what do systems have to do with people? The answer should be obvious. People, I suggest, are a type of system. I, for example, am the mereological sum of all my parts that stand in certain relations. This is not to endorse a

---

<sup>76</sup> If you recall from footnote 43, that a person P1 q-remembers some event e iff 1) P1 seems to remember e; 2) e actually occurred; and, 3) P1's seeming to remember e is caused in the right sort of way by someone's witnessing or doing e. An M-type causal chain, then, is simply what Shoemaker is getting at by 'caused in the right sort of way.'

<sup>77</sup> Shoemaker 2003 p. 36.

<sup>78</sup> This is how Kuhn, A. (1974) p. 21 explains the coupling.

strong composition as identity thesis;<sup>79</sup> that is, I am not claiming that I am identical to my parts simpliciter, but rather my parts, whatever they might be, taken as a whole.

Nevertheless, even if I am not identical to my parts taken as a whole, I still have parts and those parts interact with one another in a causal way, producing various effects. For example, my digestive system, circulatory system, cardio-vascular system, and central nervous system are all composed of different types of tissues. Moreover, they are each themselves systems, and they also compose a greater living system. Such systems are called subsystems for this reason. Given the definition of 'system' we are working with, I am therefore a type of system. Even if we are to identify ourselves with our consciousnesses, a consciousness is nonetheless composed of elements that interact to form a single experience.<sup>80</sup>

One might then be compelled to ask, what type of system are we? What is our effect? Clearly, we produce many effects. We sweat, blow our noses, and excrete other waste products. Many other systems produce multiple effects – nuclear power plants produce both nuclear energy and radioactive waste, for example – but it is the contribution to an essential effect or effects that governs the composition of a system. So what is the effect that we are after?

The interaction of some of those parts, namely, neurons and certain other things, directly results in consciousness. Thus, some of my parts form a system where consciousness – in physical worlds, a part of the c-realizer – is the system effect. This system is typically called, or found in, the central nervous system, though it also might have additional effects. Given the result of Chapter Two, that sameness of consciousness

---

<sup>79</sup> See, for example, the works of Baxter (1988), Lewis (1991), Merricks (1999), and Sider (forthcoming).

<sup>80</sup> See Tye 1995.

is sufficient for sameness of persons, I go where my consciousness goes. This gives us our answer. If I am a system, and the persistence of this system can be governed by the production of a single consciousness (that is, the system goes where my consciousness goes), then I am essentially a consciousness-producing system.

Recall that a functional definition allows the denial of the claim that I am just my brain; rather, I would simply be whatever happens to be producing my consciousness at that particular time. In order to give persistence conditions for consciousness-producing systems then, one must bring in the persistence of the c-realizer. Given the above definition of 'system', and given that my consciousness is simply realized by some part or state of my c-realizer, anything that is poised to contribute to the state of that c-realizer is a part of me. Since this contribution is causal and takes place on a temporally extended playing field, it was an interaction between elements that had occurred in the past that affected my current c-realizer. In other words, the exact state of my c-realizer right now causally depends on an interaction between things (in this case neurons and other physical subsystems) that had occurred in the past. If those past neurons had been configured in a slightly different way, then my current c-realizer would have been different, even if the phenomenal content of my consciousness did not differ. If something in the past has contributed to my current state of consciousness via an S-type relation, then it must have been poised to produce my consciousness. And, if this is so, then it must have been a part of me.

So, as the c-realizer changes throughout time, whether it is a brain or a computer or whatever, it develops a causal history. Each realization of a conscious state is structured the way it is because it is caused to be that way, at least in part, by the previous

realization. In the same way Shoemaker's M-type causal chains can provide the basis for psychological continuity, this sort of S-type causal chain can equally provide a basis for continuity of consciousness (recall from Chapter Two that continuity of consciousness and continuity of psychology is not the same thing). Two realizations of consciousness at different times are part of the same stream of consciousness only if they are connected by an S-type causal chain.

But there are gaps in consciousness. When you go to sleep or find yourself in a coma, there is no neural configuration that realizes consciousness, for you are not having any conscious experiences. One might ask what happens to you during such moments. Do you simply cease to exist? The answer is no. I am not claiming that you are your consciousness. Rather, I am claiming that you are your consciousness-producing system. As long as the thing that functions to produce your consciousness persists, you do as well. Moreover, the consciousness itself persists because the realizations we have at different times are connected via an S-type causal chain. If there is a gap in consciousness, such as one from being asleep or being in a coma, the S-type causal chain nonetheless continues, via the c-realizer, connecting the previous conscious experience to the next one.

Not only does the c-realizer, the thing that has the function of realizing your consciousness (in this case a brain), persist but, given the definition of 'system' stated above, and given that the realization of your consciousness has parts (namely, neural connections) that interact in a way that produces an effect, the realization of your consciousness is itself a mere subsystem of you. This is simply because there are many other things that contribute to the c-realizer that exist outside of it.

My current experience is also due, in part, to a series of overlapping causal chains that originate with my sensory faculties. My eyes take in light; this information runs down my optic nerve, into the back of my brain, and eventually leads to the realization of a visual experience. This sort of causal chain seems to be of the S-type (for consciousness-producing systems) as mentioned above. My eyes, since they are poised to contribute to a future S-connected c-realizer, are thus a part of me. Eardrums, tastebuds, and olfactory elements are likewise contributing and are thus likewise a part of me. In a looser sense, but still a seemingly appropriate sense, my circulatory system affects my c-realizer (given that it contributes to the functioning of my brain), as does my digestive system and respiratory system. Most of the parts of my body, the appendix notwithstanding,<sup>81</sup> are thus parts of me. Again, given the temporally extended nature of the causal process, those same parts, in the past, were a part of me since their contribution to my consciousness occurred at that past time.

In the same way, if my brain was removed from my body, placed in a vat, and hooked up to a brain-support machine, that machine would also become part of me, whereas my body would cease being part of me. Currently, my body is what is, and what has contributed to my consciousness. So, if I am a consciousness-producing system, right now (and in the past) I am my body, but this fact is modally and temporally contingent; I could very well be something else. This is the direct result of the brain-transplant intuition and the intuition that we would survive inorganic replacement (pace Olson) as mentioned above. It then seems that a systemic account is a physical view of personal

---

<sup>81</sup> Since the appendix seems to have no effect on the operation of the rest of your system, it might not be a subsystem of you. Since, however, blood moves through it, it still seems to be a non-systemic part of you.

identity that is compatible with at least some of the intuitions that drive the psychological view.

One worry some might have is that, since things outside us causally contribute to every conscious experience we have, it might turn out some odd results. A tree in the distance causes a certain impression on my consciousness; so does the sun. The gravity of the earth seems to as well; and so does getting hit on the head with a metal pipe. While these things nonetheless, I admit, contribute to our various experiences, they do not contribute in the right sort of way. We must then limit the sorts inputs that we want to turn out to be relevant to our consciousness-producing systems. This, however, is the point of positing the notion of an S-type causal chain. The sorts of contribution mentioned above do not strike me as being sufficient to be poised to contribute, in a noticeably regular way, to the production of conscious experiences. If one needed to be repeatedly hit on the head in order to experience things, only then would I think to entertain the idea that such violence is literally a part of you. To clarify what I mean by an S-type relation, I can only supplement the above definition with some examples. Biological processes are obviously of the sort that result in consciousness; these sorts of relations are then appropriately considered to be S-type. If consciousness can be produced non-biologically, for example, if computers can become conscious, then the sorts of causal relations that occur along circuit boards are also sufficiently S-type. If consciousness can be produced by the interaction of non-physical substances such as ectoplasm, then such relations are also to be considered S-type. Very general relations, such as gravity, light bouncing off a tree and hitting the back of your retina, and so on, do

not seem to be the sorts of relations that are sufficient.<sup>82</sup> By placing such a limitation on the relation, the boundaries of the system can be defined in a much clearer sort of way.

All of this might be hard to digest; however, I can find no other words that explain and motivate the sort of causal persistence I am advocating. Simply put, it is a view of personal identity that spawns directly from a perfectly sensible sounding definition of systems (or at least a specific type of system) and a common-sense explanation of how causal chains work over time. To summarize the account thus far.

For any two people P1 and P2,  $P1 = P2$  iff P1 and P2 are the same consciousness-producing system.<sup>83</sup>

This analysis is compatible with the Chapter Three claim that a sort of substrate persistence is necessary for personal persistence. Systems are physical things, or at least the types of systems we are restricting our talk to are physical things. Given that the persistence of systems is governed through systemic (in our case physical) causation, it seems to follow that systemic persistence *is* a sort of physical, or substrate persistence.

Systemic persistence is then to be accounted for as follows:

Two consciousness-producing systems S1 and S2 are the same system iff it is the interaction between S1's (and only S1's) parts that, via S-type causation, ultimately leads to the c-realizer of S2 being the way it is.

Since an essential subsystem of S1, namely, her c-realizer, is causally linked to the realization of S2's conscious experience in this way, it also follows that S1 and S2 have the same consciousness. This is thus a sort of continuity of consciousness similar to Shoemaker's M-type causation based account, without the memory requirement. The

---

<sup>82</sup> One might also attempt to restrict the causal chain through the systemic notion of feedback. Since light hitting the back of the eye cannot change its state in virtue of its relation to the eye (or the rest of the system), it cannot receive feedback from the rest of the system, and is thus not part of the system.

<sup>83</sup> While all people are consciousness-producing systems, not all consciousness-producing systems are people. Fetuses, for example, might be consciousness-producing systems, but they are not really people.

systemic account is thus compatible with Chapter Two's thesis that continuity of consciousness is sufficient for personal identity.

In Chapter Three I told a story that reflected common conceptions of our origins and development, and described how the physical view was more compatible with such a conception than the psychological view. I will now retell this story from the systemic perspective, showing how the system view also retains these intuitions. This time, however, I think it would be clearer if we first worked from the present and worked backwards.

Right now, I am having a conscious experience. This experience is realized by the configuration of certain neurons that are part of the c-realizer. This configuration is causally dependent on the state of a system that existed in the past: my body moments ago. Likewise, the conscious experience that the "moments ago" body was having was causally dependent on the state of a system that existed moments before that, and so on. All systems that are so related are the same system as me. At some point, I had my first conscious experience. That experience came about as the result of the interaction between things that once composed a pre-conscious body: a fetus. That fetus, a mere consciousness-producing system in development, is thus me, though I was not yet a person.<sup>84</sup>

---

<sup>84</sup> For those who dislike the idea that we were once non-conscious fetuses, one merely needs to modify what they consider to be the appropriate S-type relation, making it a bit more restrictive. I myself am unwilling to make such a restriction, since I think it to be intuitive that we were once non-conscious fetuses. Just because something is not actively producing a consciousness does not preclude it from being a consciousness-producing system. Since the interaction of my fetal-parts eventually resulted in consciousness, it very much seems like a consciousness-producing system, albeit a very slow one. If there were a "seed" that you could water which would, shortly after, autonomously grow into a massive stereo system, it strikes me that the seed itself would be a stereo system. After all, the seed-system does play music eventually; you just have to wait a short time after you activate it.



Going in the other direction, the state of my consciousness-producing system right now is going to influence a future c-realizer and, through it, another realization of my consciousness: that is, the realization that is part of a future consciousness-producing system. This causal connection entails that this future system is me, as is the case with all future systems whose realization subsystem is causally dependent on the compositional and causal descendents of that parent system. As soon as it becomes impossible for any more conscious experiences to be produced, then that system is dead. According to the aforementioned definition of 'system', those parts are no longer coupled to produce an effect, and the system therefore does not exist. So you are dead. Your parts exist, some of them might even be functional, but you no longer exist. Brain death, therefore, is person death even if your heart is still beating and your blood is still pumping. If, however, the brain merely appeared dead and it was possible for doctors to get it working again, you would still exist. This seems to be a sensible story of how life goes and is quite compatible with the account presented above. However, the story is not over yet.

#### **4. Circularity and Reduplication Revisited**

As we know from Chapter One, circularity is traditionally a problem that plagues only the psychological view while reduplication is a problem for both the psychological view and the physical view. The reason the psychological view succumbs to the circularity objection, however, is that having a genuine memory presupposes that you are the same person who formed the remembered experience. However, the account I have given for the persistence of persons has nothing to do with memory.<sup>85</sup> What is more, the corresponding analysis of the persistence of consciousness also drops the memory

---

<sup>85</sup> or even q-memory, the common reply to circularity. See Shoemaker 1970 or Nozick 1981.

requirement. The fact that a given realization at a given time is S-type causally linked to the state of a past c-realizer is all that necessary (and sufficient) for sameness of consciousness and sameness of person.

This might strike some as odd. In his essay, Locke makes it seem that memory is extremely important to both consciousness and personal identity. But considering real-world examples, and some fictional ones, I think Locke has been misleading us; in the face of these examples, it seems appropriate to downplay the role memory has. In patients with late stage Alzheimer's disease, and physically induced brain damage, memories are completely lost. If memory is necessary for personal identity, then such people are no longer persisting people. In this sense, the memory requirement makes it impossible to sustain such brain damage or to progress into the late stage of Alzheimer's. The patient is dead; the doctors are merely treating the body.

Likewise with patients suffering from certain cases of multiple personality disorder where the personalities are completely unaware of each other's actions. If memory were a necessary condition, then such patients are indeed numerically different people in one body. This might strike some as being an acceptable result; however, I am not one of them.

Turning to fictional cases, suppose there is a woman who has been asleep for her entire life, but otherwise has the capacity for speech (the ability having been genetically inherited), reason, and self-reflection; she merely has to wake up. Though she very well might never wake up, this woman seems to be a person, and indeed the same person as some past sleeping baby. However, without any experiences, she remembers no events and is thus not psychologically continuous with anything. She is, however, suitably

poised to produce experiences and is thus an S-type causally continuous consciousness-producing system.

Given that circularity is not a problem for the proposed systemic account, we must then turn our attention to the reduplication problem. There are many way of formulating this problem. First, if one were to disassemble all my molecules, analyze the relationships between them, and use that information to reproduce an exact copy of my body on the moon, that person would not stand in an S-type causal chain with the “pre-transported” person. That person would not be me. The same thing goes for brain-state transfers where an evil alchemist rearranges your neurons to a type-identical neurological state of another person: the alchemist-to-brain causation is not at all appropriate for being S-type. Having a computer record your neural configuration and reproduce it in another host is not any more person-preserving, even if your original brain is destroyed in the process.

One problem might come in the form of a temporary disassembly and reassembly of our brain. If the alchemist rearranged our neural connections such that we come to believe we are someone else, then arranged them back again so the consciousness is phenomenally (but not micro-physically) equivalent to the way it was before, it would seem that we are, after the re-arrangement, the same person as the original untouched person. But how is this possible on the systemic account? It seems that the causal chain is broken.<sup>86</sup>

---

<sup>86</sup> If such meddling is such that we can remain conscious throughout, then the alchemist’s instruments would indeed be part of an S-type causal chain. This, however, would entail that the instruments – or at least the part of them that facilitates consciousness – become part of you. But I think this seems correct. If I were to install a device into my brain that allows me to perceive infrared light (or to communicate telepathically), it strikes me that this cybernetic implant would become part of me. I see the alchemist’s tools as being no different.

In response, I am willing to admit that it is not the same person. In systemic terms, if I rearrange the parts of my computer so that the output of my monitor changes, then rearrange them again so that the output of my monitor changes back to what it was before, but with its parts interfaced in a totally different way, it would not be the same system. For systems, the relationships between the parts are just as essential to its identity as the parts themselves.

If the alchemist rearranged the neurons to their original state – that is, one that is not merely phenomenally identical, but also micro-physically – then that final configuration would depend on the pre-reconfigured brain state. This dependence might be sufficient for S-type causation, but how to argue for this in a convincing way, I confess, I do not know.<sup>87</sup>

In any case, these are not simply problems for my account; they can also be directed at Shoemaker's account. His definition of M-connectedness requires that the causal chain is explicitly memory-like: that is, of the sort typically involved in memory. The alchemist's meddling is certainly not the sort of causation typically involved in memory. Both Shoemaker and I must admit that the double reconfigured brain is an entirely new person.

---

<sup>87</sup> One way I could go is to suggest that the alchemist's actions are a type of causal process that is appropriate for the coupling of parts in consciousness-producing systems. This, however, would entail that, while the alchemist is operating on you, she is part of you. This is too bizarre for most to take seriously. In support of this peculiar claim, however, one can point out that we are composed of cells: independent living entities. If we imagine that each cell has its own brain and is aware of itself in such a way that it fits Locke's description of persons, then all that really seems to be driving the bizarreness intuition would be her size and the fact that she is located outside her head. Her status as an independent person would have nothing to do with it. However, we can imagine that some sort of pseudo-cancerous growth causes a single functional neuron to grow to the size of a human, forcing us to cut open the skull allowing the neuron to exist outside of it. Although it takes a stretch of the imagination to picture, we should have no trouble thinking that the huge neuron is a part of us. But really, what is the relevant difference between a giant independently intelligent neuron and an alchemist with the ability to make an equal impact on your neural processes?

Finally, we must turn to split brains. When the brain is split, both consciousness-producing halves stand in an S-type relation to the original brain. Here we might encounter a problem, for if we implant each half into two different bodies we have two new consciousness-producing systems the parts of which are not causally related to each other. However, nowhere in my analysis do I say that such a system must produce a *unified* conscious experience. Michael Tye, for example, has been known to argue that it is possible for a single subject to have two disjoint consciousnesses, albeit only “under certain experimental conditions.”<sup>88</sup>

One response, then, is to say that both brain halves remain part of the same c-realizer; they have just been spatially separated, and the consciousness itself is no longer unified. Whatever systems they are a part of are then mere subsystems of a greater consciousness-producing system.<sup>89</sup> This would of course entail that we are some sort of multiconscious entity (who is not a person anymore, since the non-unified consciousness cannot reflect upon itself as itself). Though we are only contingently persons – since we were once non-conscious fetuses and could one day find ourselves living in vegetative, albeit conscious, states – this change in status to multiconscious might strike some as being supremely bizarre and difficult to believe. But, with the landscape being as bizarre as it is, we should not be surprised.

There is another option, however, one that is perhaps more plausible. If one modifies the analysis in way that requires the system effect to be a unified conscious experience, we can say that the resulting two people are different from one another, as well as the original whole-brained person. This works through a certain conception of

---

<sup>88</sup> Tye 2003 p. 126.

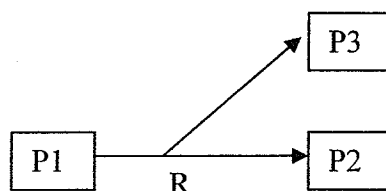
<sup>89</sup> Parfit 1971.

conscious unity: a consciousness is unified if and only if its c-realizer stands in an S-type causal relation with a previous c-realizer and is itself the only c-realizer that stands in such a relation. Similar to the “non-branching” view mentioned in Chapter Three, the modified analysis would then appear like this:

Two people P1 and P2 are the same person iff it is the interaction between P1’s (and only P1’s) parts that, via S-type causation, ultimately leads to P2’s c-realizer (and only P2’s) being the way it is.

One might accuse this modification of flying in the face of the only x and y principle also described in Chapter Three. That is, it might seem that a later person’s being identical with an earlier person depends on facts external to those people or the relationship between them. Consider the following diagram:

Figure 1: Fission

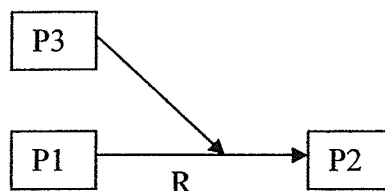


The sameness of two people has to do with only the effects caused, via the S-type causal relation denoted by ‘R’, by the parts and subsystems of P1 and whether those effects are wholly contained within P2. However, determining whether those effects are wholly contained is an extrinsic matter.

The same can be said about gestalt cases. If one half of my brain is systemically coupled with one half of your brain, intuitively the resulting fusion should not turn out to be either you or me. A restriction on P1 being the only consciousness-producing system whose parts stand in the S-type causal relation at that time would allow us to retain this

intuition.<sup>90</sup> If there were another such system, then the causal relation between P1 and P2 would branch yet again, albeit in another direction. This can be depicted as follows:

Figure 2: Fusion



Again, the fact that P1 and P2 (also P3 and P2) are different people is determined by the fact that P2 has a mixed origin. This fact is extrinsic, and the only x and y principle has been violated, despite the intuitive result.

There are other problems for the aforementioned account. What if we are to retain one of the brain halves in the original body when considering a case of brain transplant reduplication? Intuitively you would be identical to the original system: the original body and remaining brain half. The systemic account might be able to account for this; however, a move from a “non-branching” to a “best-candidate” structure (such as in Nozick’s “local”<sup>91</sup> account as mentioned in the previous chapter) is required. In fact, in light of certain fusion cases, we would have to modify the corresponding non-branching element as well. It would not do well to have an account that entailed your demise if a small part of someone else’s brain were fused with your own. These alterations then give us our final analysis:

<sup>90</sup> It also allows us to avoid the implausible conclusion that we are to be identified as both a sperm and an egg that existed in the past. Sexual reproduction is a type of fusion, in this sense, that would otherwise be problematic.

<sup>91</sup> I use the “local” version mostly due to the problems his “global” version seems to have considering the Russian roulette scenario outlined in the previous chapter. However, I also take the local version to be more intuitive: if a (locally) closest continuer were to die immediately after an instance of fission, I would think myself to be dead. If a scientist’s twinning laser were to make an imperfect split, such that one of the “twins” had purple skin and the other were to die almost immediately after, I would hardly think it to be intuitive to consider the purple twin to be the original person, however psychologically similar she might be.

Two people P1 and P2 are the same person iff

- 1) it is the interaction between P1's parts that, via an S-type causal chain, ultimately leads to P2's c-realizer being the way it is;
- 2) P2 is the closest continuer of P1; and,
- 3) P1 is the closest continued by P2.

Taking from Nozick, we can define the "closest continuer" and "closest continued" as the person who has "the highest degree of spatiotemporal and qualitative continuity with the original,"<sup>92</sup> or in the latter case, with the resulting entity. This is not getting too far away from the original application of systems concepts, interestingly enough. Recall that Nozick's notion of qualitative continuity rests on causal links between the qualities possessed by both people. The possession of such qualities in the latter system is thus an additional system effect of the earlier. Such additions should thus not be considered to be ad hoc in the light of system theory.

Moreover, by moving to a best-candidate approach, it becomes more obvious that the only x and y principle is respected, at least in regards to counterintuitive cases where the determining factor of your continued existence involves some property, quality, or action possessed or being performed on some other individual. Like Nozick's global account however, the systems approach must worry about cases where two continuers are equally close.<sup>93</sup> A violation of the only x and y principle looms. However, this is a problem for every account of personal identity that has been mentioned. No account, whether it is physical, psychological, nor even hybrid can avoid such a violation.

Another challenge might be made to the systemic account on the basis that I am not really giving an account of personal identity. By giving an analysis that identifies me

---

<sup>92</sup> Nozick 1981 p. 43.

<sup>93</sup> It should be noted that this approach need to worry about longevity. If the Rob twins from Chapter Three fight, the winner is not necessarily the original person. If they are truly equally causally continuous, then they are fighting over nothing. Neither of them is, nor will ever be the original person. No matter what they believe.



with some fetus or zygote, if it were truly an analysis of personal identity, then this would entail that fetuses and zygotes are people. However, I am not just giving an account of personal identity, I am giving an account personal identity *and more*. I am giving an account of your persistence as a single entity, whether it is a person or not. An account of personal identity is included in your persistence. If one wants an account of strictly personal identity – i.e. as persons – then one need only to restrict the analysis to people as follows:

P1 and P2 are the same person iff P1 and P2 are both people and are the same consciousness-producing system.

The analysis of 'same consciousness-producing system' remains the same. A person is merely a type of consciousness-producing system. If you want an analysis of only people, then you merely have to restrict which type of consciousness-producing system you want an analysis of. We are no more essentially people than we are human beings or even biological entities.

This then concludes the explication of the systemic account. Some might take certain results as being counterintuitive: a body with a doubly reconfigured brain is not the original person, despite what some people might claim; and cybernetic implants such as pacemakers and microchips that allow you to perceive infrared light are literally a part of you. I, however, am not bothered by this. In fact, in the latter case, these things strike me as indeed being part of you; and, in the former, it seems that contemporary accounts of psychological connectedness are equally unable to account for that intuition.

The one thing that many philosophers will likely disapprove of is the fact that I appeal to systems. Systems, it might be said, are not well understood. Moreover, some might think my definition, or even the notion of a system in general, is too arbitrary for

use in such an important type of analysis. But am I really that committed to the existence and metaphysics of systems? I do not think so. All I am talking about are groups of parts, a certain function of those parts, and a type of causal history that facilitates this function; talk of all these things is common discourse in contemporary philosophy. I can thus rephrase this analysis without any appeal whatsoever to a system. While it is true that systems concepts are the motivation for the account, the account itself can stand without such motivation. All it needs to do is hold up against counterexamples. And indeed it does so against even those that have traditionally plagued the psychological view and the physical views independently: fission cases and even fusion cases. Though the only x and y principle has been violated, no seemingly absurd results, such as your survival depending on brain surgery performed on someone else, must be acknowledged. No other theory has done any better.

To sum it up, the structure of the theory is this: persistence of personhood can be defined as two sets of parts, grouped by their relations in producing consciousness, existing at different times, that are linked by a causal chain: the sort of chain that is sufficient to regularly producing consciousness. I then add two conditions that aim to end the problems with physical fission and fusion cases<sup>94</sup> by eliminating lesser candidates or the person herself; that is, by adding the condition: if the fission produced or fusion inducing candidates are equally causally connected (spatiotemporally and qualitatively similar), then the original person ceases to exist; but if one is more connected (similar) than the other then the more connected (similar) is to be identified with the original person.

---

<sup>94</sup> These problems come from having a physical – and psychological qua physical – causal chain that is guided by its potential to produce consciousness, applied on the above mentioned non-arbitrary grouping of parts

The mechanics of the theory is this: it is essentially a slightly modified case of Shoemaker q-memory. The modification is that I drop the memory requirement from both his analyses of q-memory and M-type causation. I strip his analysis to simply its functionally directed causal chain, where I replace the function of memory-producing-type-of-connections with consciousness-producing-type-of-connections. I combine this causal chain with, not a single object as in traditional physical views, but rather a similarly functional method of grouping together a group of parts;<sup>95</sup> that is, grouping parts together by the functional relations between them, namely their consciousness-producing or realizing-type-of-relations. By maintaining only the physical nature<sup>96</sup> of an M-type causal chain, I reduce the psychological component of the view to a purely functional role in both the diachronic answer to the question of personal identity – the persistence conditions – and the synchronic answer: our relation to our bodies.

Even more pleasant, talk of people as systems (or merely groups of parts) accords with our everyday talk of people in a more robust way than either the psychological or physical views. It allows that we can refer to people both as physically persisting things and as entities with persisting consciousnesses. The results of both Chapters Two and Three are compatible, and this is a feature not found in any strictly psychological or physical view. When someone dies, their parts can no longer perform the consciousness-producing function, and without the potential for such a function the person no longer exists, though we can still refer to that group of parts by name when, say, seeing them in a casket. When someone is conceived through sexual reproduction, the interaction of

---

<sup>95</sup> Not necessarily physical parts, to apply to disembodied consciousness worlds.

<sup>96</sup> But not necessarily physical since if disembodied consciousnesses are possible, then the metaphysics of such a world would be much more plausible if that disembodied consciousness was produced by a complex consistent, physics-like, directed causal process. That process would be of the consciousness-realizing type, thus making it S-type.

those parts eventually lead to a first conscious experience, and the zygote is then the same thing (though not yet a person) as the later person.

When talking about meaning as being determined by use plus eligibility, as Sider does, it then seems that the systemic or functionalist account of persons has the upper hand on both the psychological view and physical views. It has the upper hand on the physical and body views since it can account for body-swapping and it has the advantage over the psychological view in that it can account for the persistence of a person despite drastic overnight changes in personality, for the possibility of acquiring amnesia, and for the possibility of suffering from late stage Alzheimer's disease. These are all things that the psychological view cannot account for; if such things were to happen to a person, then the psychological account would entail that the person ceases to exist. Furthermore, the system view goes beyond an analysis of personal identity qua persons. It enjoys the primary benefit of animalism which gives us an identity analysis of the relation between me and some past fetus.

(P5) of Sider's argument is thus false. The systems view is compatible with both the conclusions of Chapter Two and Chapter Three, gaining benefits that neither the psychological nor physical views have individually. Moreover, even if the only x and y principle has been breached, this is a fault shared by both the competing views (and even then, it is not an overly startling violation of the principle considering the pepperoni example given in Chapter Three, and the fact that the account does lead to the absurd result that our survival can depend on brain surgery being performed on someone else). Since the systemic view has an additional benefit, and seems to suffer from no problems that do not plague both of the competing theories, the systemic view seems to be the most

eligible candidate. There is thus a fact of the matter of personal identity: we are persisting systems.

## Works Cited

- Ackoff, R. L. 1971. "Towards a System of Systems Concepts", *Management Science*, 17(11).
- Ayer, A. J. 1936. *Language, Truth, and Logic*. London: Gollancz
- Baxter, D. 1988. "Identity in the Loose and Popular Sense", *Mind*, 97: 575-582.
- Bertalanffy, L. 1968. *General Systems Theory*. New York: Braziller.
- Butler, J. 1736. "Of Personal Identity", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Ehring, D. 1987. "Personal Identity and Time Travel", *Philosophical Studies* 52: 427-433.
- Geach, P. 1967. "Identity", *Review of Metaphysics* 21: 3-13.
- Grice, H. P. 1941. "Personal Identity", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Kuhn, A. 1974. *The Logic of Social Systems*, San Francisco: Jossey-Bass.
- Lange, O. 1965. *Wholes and Parts: A General Theory of System Behavior*, Oxford: Pergamon Press.
- Locke, J. 1694. "Of Identity and Diversity", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Lewis, D. 1976. "Survival and Identity", in *The Identities of Persons*, A. Rorty (ed.), Berkeley: California, and reprinted in his *Philosophical Papers* vol. I, Oxford University Press, 1983.
- Lewis, D. 1991. *Parts of Classes*. Cambridge: Blackwell.
- Merricks, T. 1999. "Composition as Identity, Mereological Essentialism, and Counterpart Theory", *Australasian Journal of Philosophy* 77, no. 2: 192-195.
- Nagel, T. 1971. "Brain Bisection and the Unity of Consciousness", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Noonan, H. 2003. *Personal Identity*, 2<sup>nd</sup> ed., New York: Routledge.
- Nozick, R. 1981. *Philosophical Explanations*, Oxford: Clarendon Press.

- Olson, E. 1997. *The Human Animal: Personal Identity without Psychology*, Oxford: Oxford University Press.
- Parfit, D. 1971. "Personal Identity", in J. Perry (ed.), *Personal Identity*, Berkeley: University of California Press, 1975.
- Putnam, H. 1981. *Reason, Truth, and History*, Cambridge: Cambridge University Press.
- Quinton, A. 1962. "The Soul", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Reid, T. 1785. "Of Mr. Locke's Account of Our Personal Identity", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.
- Shoemaker, S. 1970. "Persons and Their Pasts", in *Identity, Cause, and Mind*, 2<sup>nd</sup> ed, S. Shoemaker (ed.), Oxford: Clarendon Press, 2003.
- Sider, T. 2001. "Criteria of Personal Identity and the Limits of Conceptual Analysis", *Philosophical Perspectives* 15: 189-209.
- Sider, T. 2004. "Personal Identity over Time", in *Riddles of Existence: A Guided Tour of Metaphysics*, E. Conee and T. Sider (eds.), Oxford: Oxford University Press, 2005.
- Sider, T. "Parthood" (forthcoming).
- Thompson, J. J. 1997. "People and Their Bodies", in *Reading Parfit*, J. Dancy (ed.), Oxford: Blackwell.
- Tye, M. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, Cambridge: MIT Press.
- Tye, M. 2003. *Consciousness and Persons*, Cambridge: MIT Press.
- Wiggins, D. 1967. *Identity and Spatio-Temporal Continuity*, Oxford: Blackwell.
- Williams, B. 1970. "The Self and the Future", in *Personal Identity*, J. Perry (ed.), Berkeley: University of California Press, 1975.