

2004

Concerning theories of personal identity

Patrick, Bailey

University of South Florida

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

Scholar Commons Citation

Bailey, Patrick,, "Concerning theories of personal identity" (2004). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/945>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Concerning Theories of Personal Identity

by

Patrick Bailey

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Department of Philosophy
College of Arts and Sciences
University of South Florida

Major Professor: Charles B. Guignon, Ph.D.
Stephen P. Turner, Ph.D
Roy Weatherford, Ph.D

Date of Approval:
March 31, 2004

Keywords: personhood, memory, consciousness, mind, self

© Copyright 2004 , Patrick Bailey

Table of Contents

ABSTRACT	ii
INTRODUCTION	1
CHAPTER 1 HISTORICAL ACCOUNTS OF PERSONAL IDENTITY	6
Historical Significance	6
The Cartesian Mind-Body Problem	6
Locke and Consciousness	18
Hume's Denial and the Bundle Theory	28
CHAPTER 2 THE MENTAL PHENOMENA	36
Mental Phenomena and Personal Identity	36
The Memory Criterion	37
When Memory Fails	43
The Psychological Criterion	47
Division, Replication and other Problems	52
Persons Through Time	57
CHAPTER 3 THE BODILY CRITERION AND REDUCTIONISM	68
Persons and Bodies	68
Soul Searching	68
Bodies, Consciousness and Reduction	82
CHAPTER 4 THE MEANING OF IT ALL	99
Drawing Conclusions	99
Contingency and Arbitrary Decision	99
Language and Meaning	107
REFERENCES	117

Concerning Theories of Personal Identity

Patrick Bailey

ABSTRACT

The purpose of this thesis is to provide a brief examination of the historical accounts of philosophical theories of personal identity and show the influence that each has had on the development of contemporary theories. In doing so, the thesis explores the problems associated with these theories, attempting to establish a meta-theory (i.e. a theory about theories) of personal identity. What is demonstrated is that the fundamental problems of personal identity arise from issues related to the use of language, as well as assumptions involving the concept of personhood.

By demonstrating that our understanding of personhood is relative to frameworks of understanding based on assumption, the meta-theory states that propositions made about persons are not factual statements, but are, rather, matters of contingency. As such, propositions about persons contain truth-value only within a particular frame of reference that is based on these assumptions. Therefore, the problems that traditionally arise in theories of personal identity – problems with dualism, the mental criterion, and bodily criterion – result from a flawed approach to the problem altogether. The conclusion is that it is possible to construct *a* theory of personal identity (a relative theory), but not *the* theory of personal identity (one which is definitive and strictly conclusive).

INTRODUCTION

Examining the philosophical problem of personal identity requires considering several inter-related concepts, all of which help answer questions pertaining to different aspects about this problem. What is the problem of personal identity? It is a problem that arises when considering what it is to be the same person from moment to moment. That is, what makes us inclined to say that we are the same person now as we were five years ago, as well as that we will be the same person five years from today? At the level of experience, nothing could be more obvious than the fact that we are the same persons now as we are at any other given point in our lives. Yet, what does it mean to be the *same person*? To attempt to answer this question, we must consider concepts such as identity and sameness, personhood, mind and the self, bodily continuity, and memory and psychological continuity.

To begin, identity is the relationship that a thing bears to itself, as compared to its relationship to other objects. In other words, *identity* is what makes a thing what it is, which separates it and makes it distinguishable from all other things. Our understanding of identity is what gives rise to our concept of *sameness*. For example, if we claim that X and Y are *the same*, then what we are asserting is that both X and Y are, in fact, identical.

Personhood is the concept of what it means to be classified or qualified as a person. Personhood, then, is the sum total of all criteria that a thing must possess *to be a person*. Typically, these criteria are cited as being properties such as consciousness and,

more specifically, self-consciousness, as well as freedom of will, being a moral agent, and the ability to use language, among others. Some of these criteria imply certain perplexities that are contrary to our phenomenal experiences about personal identity. For example, at a phenomenal level (i.e. the level of experience), it seems counterintuitive to suggest that there was a time when we were not persons. Yet, claiming that moral agency and the ability to use language are two of the criteria of personhood implies that some human beings do not qualify as persons. One such example is babies. Babies lack both the ability to act morally and the ability to use language. The implication, then, is that persons are not things we simply *are*; rather, what is implied is that persons are things we *become* as we acquire the appropriate characteristics of personhood. Yet another implication is that, if we were to lose these criteria, we would lose our status of personhood. We see, then, that there are many perplexities that arise when we consider the concepts of personal identity, which are not apparent at the simple level of experience. The examination throughout this thesis will focus on the problems that are involved in attempting to develop a definitive theory of personal identity. By a *definitive* theory, I mean one that is conclusive – one that resolves the issues concerning the concept of personal identity. From our understanding of the concepts of identity and personhood, we see that we are essentially asking three fundamental questions, when inquiring about personal identity: 1) what is identity?; 2) what is a person?; and 3) what makes a person the same from one moment to another?

Another aspect of addressing the problem of personal identity involves examining ideas regarding our first-person perspective. That is, not only do we address issues about personal identity as it applies to others, but we also address personal identity as it relates

to ourselves. The first-person perspective poses problems that are not necessarily present when we examine the identity of others. For example, if we consider the identity of other persons, we may not believe it necessary to think that questions about their identities *must* have determinate answers. We may feel inclined to say that there are instances when we could not determine whether or not someone was the same person from one instance to another. Yet, when we consider our own identity this assertion appears to be an absurdity. It seems we should always be able to give a definitive answer to the question, “Am I the same person as I was or will be at any other given moment?” Furthermore, how do we *know*, in the strict sense, that we are the same from moment to moment?

In turning questions about personal identity towards ourselves, we then begin to examine concepts such as *mind* and the *self*. These concepts, often conflated, refer to a kind of internalized representation of who or what we are. This internalized representation comes from the realization that we are individuals, separated not only from all other people, but also separated in a unique way from the entire universe – there is no other thing that exists that is identical to our individual being. While sometimes used interchangeably, there are subtle differences that arise in our use of the concepts of mind and self. For instance, there are times when we describe *mind* as being thought or the process of thought and brain functioning, whereas *self* is often described as something entirely different. *Self*, in such instances, appears to take on the description of a kind of psychological *core* or *center*, in that it is described as being the essence of what we are. The concept of this sort of psychological centrality or unification comes from our representations of ourselves at the phenomenal level. There is a sense of being *in our*

head, so to speak, which gives us not only the feeling of being separated from all other things, but also gives rise to the belief that we are something more than the collective parts of our bodies.

When we consider the idea of losing various parts of our body, we realize that such a loss does not affect what we say regarding our personal identity. The loss of our limbs and replacements of internal organs do not, we say, make us different persons than we were before these changes occur. Yet, there is also a sense in which our bodies do play a part in answering questions about our personal identity. For example, some of the criteria we regard when answering questions about personal identity are bodily identity, and mental criteria, such as memory and psychological continuity. Bodily identity allows us to determine whether or not a body at one time is the *same* body at another time, because we can trace a body's spatio-temporal continuity from one moment to the next in a series of causally connected moments.

We understand memory as the ability to recall events about our past. Memory is an important concept regarding the investigation of personal identity, because it is our recollection that helps establish our sense of being continuant individuals. That is to say, memory allows us to recall whether or not we *are* the same person that did X at a previous time. Our ability to recall our past actions connects us to those actions *as* the person who performed them. They are actions that are uniquely ours. No other person, we believe, can share in our self-history in the way that it relates to our own first-person perspective. Yet, it is obvious that our memory is fallible. It is in instances where memory claims become dubious that we often consider the other criteria, such as bodily identity or psychological continuity, to support our theories of personal identity.

Psychological continuity goes beyond the scope of memory in that it includes other sorts of mental phenomena and psychological states, such as a person's beliefs, intentions, desires, and character. By including these phenomena, in addition to memory, a theory of personal identity can be posited even if the memory criterion proves to be invalidated. We see, then, that psychological continuity (not unlike bodily identity) is a concept that involves a causal relationship – an overlapping of various psychological states that connect a person's mental history into a series of such states, which spans from one time to another. These psychological states and various mental phenomena are also posited as unique to each individual person.

It is with an understanding of the above concepts that we will examine the problem of personal identity. Our examination will investigate historical and contemporary theories and will demonstrate the strengths and weaknesses of each. The result will be to explain why these theories of personal identity have failed to provide the sort of conclusive, *definitive* theory that we hope to establish.

CHAPTER 1

HISTORICAL ACCOUNTS OF PERSONAL IDENTITY

Historical Significance

In this opening chapter, it is my intent to introduce three of the primary historical accounts of the problem of personal identity. The historical accounts we will examine are those presented by Rene Descartes, John Locke, and David Hume. The purpose of including these historical accounts is to demonstrate the influence each has had on our contemporary discussions about personal identity. As such, it is not my intent here to present a painstakingly scholastic examination of each philosopher's position. Instead, I present a less specialized (i.e. general or introductory) approach, merely because I want the focus of this thesis to rest on the contemporary discussions. While important in their own right, these historical accounts are used herein as a tool for laying the foundations of the contemporary views examined in the chapters ahead.

The Cartesian Mind-Body Problem

In investigating the nature of the human mind, Rene Descartes (1596-1650) creates what is called the mind-body problem. In short, Descartes' position affirms that there is a fundamental difference between mind and body. The mind-body problem, as we shall see, derives from the belief that mind is a substance that is not extended in space, unlike body. Describing mind as a non-extended (immaterial) substance raises the

question of whether mental phenomena are equal to physical phenomena or, if not, how mental and physical phenomena relate to each other. The mind-body *problem*, then, is an attempt to reconcile conflicts in the concepts of the interactions of mental phenomena and physical brain processes. While it is often the case that we examine the mind-body problem separately from that of personal identity, I believe the two are not mutually exclusive of each other. Sydney Shoemaker echoes this idea when he states, “The problem of personal identity can be viewed as an aspect of the mind-body problem.”¹ We might say that Descartes was, in a sense, exploring personal identity inadvertently when he examined the mind-body problem. Reasons supporting this idea will become apparent as we proceed with our investigation.

Descartes’ query into the nature of the human mind begins when he asks himself what he can know with certainty – those beliefs he might have which are beyond all doubt. He proposes to set aside anything, “which admits of the slightest doubt,”² in order to find what can be known with unyielding certainty. What this means is that Descartes will hold as false any belief he has where doubt can be raised regarding its truth-value. In doing so, he aims to uncover propositions of certainty or come to the realization that there is no certainty. From this beginning, Descartes determines that the one thing he cannot doubt is his own existence, because if he can put forth a thought regarding his existence, then he necessarily exists (16). The one thing inseparable from him, he believed, was thought (Descartes, 18). It is with this foundation that Descartes begins to address notions pivotal for the concept of personal identity.

In his attempt to discover the nature of his existence and, furthermore, what can be known (in the strict sense), Descartes examines the concept of ‘I’. His realization that

he cannot remove himself from thought brought him to describe 'I' as a thinking thing, which is, essentially, "a mind, or intelligence, or intellect, or reason..." (Descartes, 18) and also a "thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions." (Descartes, 19). It is easy to see why I (like Shoemaker) believe the mind-body problem is an aspect of the problem of personal identity. "What am I?" and "Am I the same, today, as I was ten years ago?" are ways that questions about personal identity are often phrased.

Descartes' mind-body problem results from the assertion that his mind is not identical with his body. That is, he states that self-movement is foreign to the nature of bodies (Descartes, 17) and claims that, "I am not that structure of limbs which is called a human body." (Descartes, 18). This basic assertion about the nature of minds and bodies provides the foundation he needs for a mind-body distinction and it is precisely this sort of assertion that some people, such as John Searle, think allows the mind-body problem to persist as it does in our contemporary philosophical discussions. In fact, Searle states, "I am convinced that part of the difficulty is that we persist in talking about a twentieth-century problem in an outmoded seventeenth-century vocabulary."³

Clearly, we can understand the importance of Searle's assertion by comparing a Cartesian description to one from contemporary sources. For example, in the *Meditations*, Descartes describes 'body' as, "whatever has a determinable shape and a definable location and can occupy a space in such a way as to exclude any other body."

(17). In contrast, Paul Churchland states:

It is now neither useful nor accurate to characterize ordinary matter as that-which-has-extension-in-space. Electrons, for example, are bits of matter, but our best current theories describe the electrons as a point-

particle with no extension whatever (it even lacks a determinate spatial position).⁴

Noting this distinction between previous and present descriptions is important because it is Descartes' understanding of bodies that is essential for his belief of mind-body separation. Once we blur or erase the Cartesian line between mind and body, those like Searle believe we can finally put the issue to rest. Part of the problem with the Cartesian position is the need to explain why mere matter cannot produce a phenomenon such as thought. I believe that the idea that there must be something *extra* required for intelligence, thought, intention, consciousness, and the like, comes from our observations of inanimate and animate objects. We observe various objects, both natural and artificial, some of which display intelligence and consciousness, while others do not. For example, human beings behave with intelligence and consciousness – behavior not observed in things such as liquids, gases, solids and all things typically categorized as inanimate. Therefore, presumably, from the Cartesian position, there must be some fundamental difference between animate and inanimate objects. For Descartes the immaterial substance of mind is what accounts for this difference, which is not possessed by those things we observe to be inanimate. However, nowadays, computers (especially as they relate to artificial intelligence) are a peculiar kind of example, in that they are man-made objects that can perhaps be described as acting intelligently. We will consider the implications of intelligently behaving machines in the chapters ahead.

Searle asserts that the mind-problem has less to do with immaterial substances than it does with a need for a better understanding of causation (20). Searle argues that there are essentially four things that have caused us to say such “strange and implausible things” about the mind: 1) consciousness, 2) intentionality, 3) subjectivity, and 4) mental

causation (15-17). We can appreciate Searle's perspective a bit more by understanding that he believes the mind is nothing more (or less) than a result of the simple biological functioning of the brain. Hence, he asks, "Why do we still have in philosophy and psychology after all these centuries a 'mind-body problem' in a way that we do not have, say, a 'digestion-stomach' problem?" (Searle, 14). For Searle the mind and the brain are separate only in our descriptions, not in the substantial sense of the Cartesian position.

Descartes' sense of 'I' is not one that is identical to mental states. In other words, this 'I' is a *thing* that *has* mental states and, curiously enough, exists apart from the body. He states, "...it is certain that I am really distinct from my body, and can exist without it." (Descartes, 54). We understand, then, that his position champions mind as the necessary substance of one's existence. Bodies, he believes, in contrast, are not essential to the existence of thinking things – allowing that we could all very well go on existing should our bodies vaporize at any given moment.

All of these ideas raise questions regarding personal identity. For example, if we grant the separation of mind and body, then which of these, if either, account for personal identity? The Cartesian account, resting on the mind-body distinction, allows us to formulate the following as our options for a theory of personal identity: Are persons identical with minds, with bodies, or with the union of the two? Or does personal identity consist of something else entirely? Descartes never directly answers this question. That is, he leaves no doubt as to what he believes is the essential nature of his being (mind, as opposed to body), yet at no time does he explicitly state that he equates minds and persons.

What we do get from Descartes, however, is that minds are neither identical to mental states, themselves, nor to bodies. Accordingly, if we argue that persons are identical to Cartesian minds, then persons are not identical to mental states or bodies. As we have seen, Descartes defines a mind as a thinking thing that *has* mental states and is separate from his body. It appears, I think, that there are contradictions in Descartes' argument. For example, thus far we have seen that he argues he is essentially a mind, and that this mind is not simply identical to thought, but is, rather a thing that *has* thoughts – a thing that thinks. Yet, he also claims, “For it could be that were I totally to cease from thinking, I should totally cease to exist.” (Descartes, 18). The contradiction becomes clear through the following questioning: If minds are thinking things, which are not equal to thought but things that *have* thoughts, then how could the cessation of thought cause a mind to cease its existence? That is, only if minds are identical with thought should Descartes' assertion logically follow. If minds are things that *have* thoughts, then a mind should still exist even when all thinking ceases, because the *thing that thinks* should remain even when thinking (the action performed by this thinking thing) has stopped.

What then is a person – an immaterial substance? Let us consider each of the above questions, briefly, to understand their implications. Greater detail will be given when we look at the mental criterion and the bodily criterion.

Are persons identical with minds? By equating persons with Cartesian minds, we are then claiming that persons are immaterial substances, which are thinking things that do not rely on a body for their existence. Therefore, wherever this mind goes, so goes the person. Speaking in this way, we can say that persons *have* bodies. It may be argued that minds are not extended and, as such, minds do not *go* anywhere, because they are not in

space. Yet, if minds are not in space, then in what way do they exist separately from the body? How do they exist when not embodied? This does not entail that, since minds cannot be observed in any way apart from bodies, therefore minds do not exist apart from bodies. Clearly, they *can* exist apart from bodies without our being able to observe them directly, but how would we determine the truth or falsity of such a claim? Claiming to know that minds can exist apart from bodies, without any directly observable phenomena, is essentially like stating, “I cannot see, hear, or touch a mind when it exists separately from a body, yet I am certain that it exists in such a way.” I think an appropriate response to this assertion is, “OK, so, how do you *know*?” The problem is not only that we cannot verify the claim through direct observation – we cannot even describe a conceivable way to verify the truth of our assertion. Again, the need for verification here does not in any way alter the truth or falsity of the proposition. Rather, what it changes is our right to claim we have genuine knowledge about the truth or falsity of our proposition.

Similarly, how can we determine if an immaterial substance is the same from one moment to the next? That is, if we cannot in any way examine the immaterial substance claimed to inhabit a body, then how can we verify whether or not the immaterial substance inhabiting Michael Ellis’ body is, in fact, that same immaterial substance that inhabited his body two years ago? When discussing the bodily criterion, we will understand why it is difficult to defend a theory of personal identity based on the premise of immaterial substances.

Are persons identical with bodies? If we maintain the logic of the Cartesian position, then we agree that minds and bodies are separate; therefore, minds and persons

are separate (if persons are bodies); therefore, minds are not persons. This approach eliminates the problem of verifying claims about immaterial substances, since it asserts that persons simply *are* bodies. However, we will see that similarly difficult consequences arise for a theory of personal identity founded on the premise of the bodily criterion, when we consider several problem cases.

Are persons a result of the union of minds and bodies? This question may seem a bit peculiar, since it is not often considered, so I will restate it in another manner. Do persons exist only when we have the union of both mind and body, in the Cartesian sense? Here is what this description implies: minds, apart from bodies, are not persons. Furthermore, bodies, apart from minds, are not persons. Only when we have a union of a particular mind with a particular body can we have personal identity, by this account.

The problem with the above position is that neither minds nor bodies, themselves, constitute persons. This implies that a person *is a unified* mind and body, but is not identical with a mind or a body. Therefore, it further implies that having both a mind and a body is a necessary condition of being a person. Wherever the mind *and* the body goes, then, so goes the person. Like the claim that persons are minds, this position faces the same problems regarding the verification of immaterial substances. That is, if we claim persons are a result of the union of minds and bodies, then how do we verify the existence of the immaterial substance (mind), which accounts for part of this unification? So far as we can tell, all that exists is the material body, which does not allow for persons, according to this position. The result is that we could never tell whether or not a genuine person exists, since we cannot verify the existence of fifty percent of this union – namely, the immaterial mind.

Perhaps the objection will be raised that verification of our claims about immaterial substances is not necessary in order for these claims to be meaningful. I will agree with this claim to the extent that beliefs are in fact meaningful, to some degree, without needing verification. However, what I argue is that propositions made without any sort of demonstrable verification (whether it be direct observation or premises that follow from self-evident truths) bear no legitimate claim to genuine knowledge. To suggest otherwise dissolves the distinction between knowledge, in the strict sense, and belief.

Returning to the concept of causation, we find other accounts that compound the difficulties of the Cartesian mind-body distinction. How does something entirely immaterial (mind) interact with and influence something entirely material (body)? Answering this question is important for our understanding of personal identity, because the answer we get will help shape the contemporary arguments for and against the mind-body problem. The root of this problem, from the Cartesian account, is explaining causation between two fundamentally different substances. Descartes was well aware of this problem and made attempts to reconcile it, since clarifying this point is pivotal to the strength of the Cartesian position. That is, in order to solidify the logical foundations of his argument, Descartes must account for the mind's ability to interact with or upon a body. He elaborates upon his previous descriptions about mind and body, stating, "...I am not merely present in my body as a sailor is present in a ship, but I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit." (Descartes, 56).

It is counterintuitive, I think, to describe minds as being both separate from and, at the same time, closely joined with the body. This is why those such as Searle believe it is so important to rethink the way we conceive the mind-body problem. In Searle's case, if minds and bodies do not separately exist, then we do not face the problems of causal interaction present in the Cartesian position. This might lead us to the hasty assumption that personal identity consists solely in the physical or bodily criterion. Yet, the problem cases ahead will demonstrate why we must also carefully consider this conclusion.

As we saw, the Cartesian mind-body distinction allows us to formulate four possible accounts of personal identity: 1) personal identity goes where the mind goes; 2) personal identity goes where the body goes; 3) personal identity is a result of the union of both mind and body; 4) personal identity is a result of something entirely apart from the mind and the body. To this we can add a fifth account – that there is no such thing as personal identity. These accounts, in many ways, form the summation of the various theories of personal identity we find in contemporary discussions. We might argue that these accounts, as described here, are counterintuitive, because identity does not *move*, as it were. Yet, we will see in the following chapter that the mental criterion and psychological criterion both suggest that personal identity follows the brain.

The first of the above accounts regards the content associated with the memory and psychological criteria. The second and third accounts address the bodily criterion and the notion that mind and body are essential to personal identity, respectively. The fourth account affirms that what matters is something such as survival, rather than identity, which we will discuss later. Finally, the fifth account addresses the idea that either there really are no criteria we use when we talk about personal identity, or that

personal identity is an illusion of sorts. Before addressing these issues in a contemporary forum, we will discuss how the ideas of both Locke and Hume influenced the way we think about personal identity.

The concept of sameness is a concept closely related to the concept of personal identity, as well as our understanding of identity in general. This is obvious in such questions as, “Will I be the *same* person if I suffer from amnesia?” In fact, without the idea of sameness we would not have the notion of identity. It is sameness that allows us to recognize an object at time T_1 as the object we see also at time T_2 . This applies to Descartes’ argument in that he claims he is not only a thinking thing, but is also the *same* thinking thing from one moment to the next (19). Yet, this idea merely begs the question, “How or what is it that is the same from one moment to the next?” It is not enough for us simply to state we are the same persons (or minds, for Descartes). To establish a viable theory about personal identity we must also demonstrate how or why we are the same persons.

It is arguable that we could claim, from the Cartesian perspective, that being the same mind from moment to moment demonstrates the separation of mind and body. That is, if we lose any number of limbs (or other body parts), we would not be inclined to claim that we were different persons. The Cartesian could argue, then, that the mind is genuinely separate from the body and, therefore, no amount of bodily loss will change our identity. For Descartes, what follows from this is the demonstration that 1) minds are “single and complete,” and 2) the mind is entirely different from the body (59). In Chapter 3, we will see how our understanding of the brain weakens the impact of such claims.

Another claim Descartes makes about mind-body causality is that the only part of the body that immediately affects the mind is the brain – specifically, the pineal gland.⁵ This claim is important because it demonstrates upward and downward causality in Descartes' argument, for not only do we have the mind causally affecting the body, but we also have the brain affecting the mind in a causal relation. The idea of two-way causality is a notion that Searle champions, which he believes helps dispel the mind-body problem. Searle states, "Nothing is more common in nature than for surface features of a phenomenon to be both caused by and realized in a micro-structure, and those are exactly the relationships that are exhibited by the relation of mind and brain."(Searle, 22). Again, the idea here is that understanding the process is what is important in order to dispel the mystery.

What Searle's statement brings to the discussion of mind-body causality is that we do not need to appeal to immaterial substances for our explanations. Accordingly, our theory of personal identity will be based on our understanding of the micro-level functions of the brain and their relationship to the behavior of the higher-level features of the system. This kind of explanation is very similar to the notion of the emergent properties of systems, in that the brain, by virtue of its organization and simple functions, produces such phenomena as consciousness, intentionality, thought and the like. We can clearly see the rejection of immaterial substances and their role in causal relations, when Searle asks, "How, for example, could anything as 'weightless' and 'ethereal' as a thought give rise to an action?"(25). He follows with, "The answer is that thoughts are not weightless and ethereal. When you have a thought, brain activity is actually going on." (Searle, 25).

Clearly, the evidence demonstrates that the Cartesian mind-body problem generates many difficulties for our formulation of a definitive theory of personal identity. Its focus on defining the self in terms of immaterial substances creates problems on both the causal level, as well as the level of experiential verification. The philosophical position of Locke shifts away from the concept of substance, focusing instead on the role of memory in defining one's personal identity. While the memory criterion eludes the problems associated with immaterial substances, it creates other issues, which are similarly potent to the formulation of a definitive theory of personal identity.

Locke and Consciousness

In *An Essay Concerning Human Understanding*, John Locke (1632-1704) develops what has become one of the most influential contributions to the discussion of personal identity – the idea that consciousness (i.e., memory) is vital to the constitution of our identity through time. Although there are some similarities between Locke's treatment of persons and the ideas discussed regarding Cartesian selves, Locke's argument is fundamentally different than Descartes' mind-body problem. This fundamental difference, as we will see, is a result of the way these two philosophers view the role of *substance*, with regard to its importance in determining issues associated with personal identity.

In short, what Locke argues is that memory accounts for personal identity and that the sorts of substances described in the mind-body problem do not determine our identity over time. Most of the criticisms against Locke's position can be generalized into four primary claims: 1) Locke conflates the concepts of consciousness and memory; 2) his theory, taken literally, requires criteria too stringent to produce a viable definition of

personal identity; 3) memory cannot define personal identity, since it presupposes it; 4) memory cannot define personal identity, because memory claims are essentially unverifiable. These criticisms are introduced here, but will be examined more fully in the following chapter, when we examine the contemporary discussions of the memory criterion and the psychological criterion.

To begin, what can we find from comparing Locke's argument with the ideas expressed by Descartes? One of the first similarities we find between Locke and Descartes is the idea that altering the mass of a living body does not alter its identity. As we saw, Descartes argued that losing a limb or other body part does not affect the *sameness* of his mind. Similarly, Locke states that, "In the state of living Creatures, their Identity depends not on a Mass of the same Particles; but on something else. For in them the variation of great parcels of Matter alters not the Identity..."⁶ and, furthermore, "The reason whereof is, that in these two cases of a Mass of Matter, and a living Body, *Identity* is not applied to the same thing." (330).

We see, then, that Locke believes the identity of living things is different from that of non-living things. For a living being, he argues, identity is communicated through the common life, or continuity of life, of that being (Locke, 331). This is why we can claim that a tree is the same tree, from year to year, even though branches and leaves may fall from it. Though its mass changes, it holds the same continuity of life. So long as the continuity of life remains intact, accordingly, we find the preservation of identity in living things. This notion produces some assumptions or implications about the differences of animate and inanimate objects – namely, it implies that living bodies (animate objects) are other than mere matter (inanimate objects). The implication we

find here is the same kind of belief in *something extra* we observed in Descartes' philosophy.

Locke creates a further distinction between being the same *man* and being the same *person*. The same man, he writes, consists in "a participation of the same continued Life, by constantly fleeting Particles of Matter, in succession vitally united to the same organized Body." (331). Locke's *man* is identical with a biological body. By way of comparison, he describes a *person* as "a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places; which it does only by that consciousness, which is inseparable from thinking, and as it seems to me essential to it. . . ." (Locke, 335). He follows by claiming that the identity of persons extends backward and forward, into our past and future, only so far as this consciousness extends – thereby uniting our identity with our actions and thoughts (Locke, 336).

Why does Locke make this distinction between *man* and *person*? I believe he does so for the same reasons we find the mind-body distinction in Descartes' argument – namely, because Locke needs a way to account for the immaterial aspects of humans. Since the distinction Locke makes is that a man is a material body, while a person is a thinking, intelligent being, it is obvious that this implied something extra is a quality belonging to persons, rather than bodies. Persons, then, refers to the vitality or animate aspect of humans, whereas a man is the inanimate body of matter. We should not assume, therefore, that Locke is equating person with immaterial substance. Clearly, Locke does not equate persons with immaterial substances, in the Cartesian sense. That is not what I am implying when I refer to the immaterial aspect of animate objects.

Rather, what I mean is that we find the same sort of tacit belief implied in Locke's writing as we do in Descartes – that mere matter cannot produce the phenomena of consciousness, intentionality, thought, and the like. However, in addressing this notion, Locke denies that immaterial substances, as well as *body*, play no role in determining one's personal identity. Hence, he states:

That if the same consciousness (which, as has been shewn, is quite a different thing from the same numerical Figure or Motion in Body) can be transferr'd from one thinking Substance to another, it will be possible, that two thinking Substances may make but one Person. For the same consciousness being preserv'd whether in the same or different Substances, the personal identity is preserv'd. (Locke, 338).

Locke's concept of person differs from a Cartesian self, in that a Cartesian self *is* an immaterial substance. In contrast, Locke's *person* is something that may reside in or be expressed through an immaterial substance, yet is independent of it. The sort of substance involved, – material or immaterial – says Locke, is irrelevant for both the determination and preservation of personal identity (336). So, we see that Locke's *person* is immaterial in the sense that it is not bound by substance, but exists so long as the *same* consciousness exists. This is precisely the point made by David Wiggins, regarding substance, when he states, "A person is material in the sense of being essentially enmattered; but in the strict and different sense...*person* is not necessarily a material concept."⁷ A consequence of this disregard for substance, with regard to the preservation of personal identity, as stated in Locke's quote, above, is that persons are things that can occupy more than one body. For we note that Locke argues that two thinking substances may make a single person, so long as the same consciousness is preserved. When we examine some of the issues about persons and duplication, in Chapter 2, we will see why Locke's assertion here strains the concept of personal identity

– primarily, because of the problem of identity with regard to the diverging futures of objects (i.e., fission).

It is Locke’s idea of the primacy of consciousness, with regard to personal identity, that generates most of the criticisms about his theory. As noted earlier, one of these criticisms is that Locke appears to confound the notions of consciousness and memory. This is a problem cited both in modern discussions, as well as in the writings of Locke’s contemporaries.⁸ Generally, it is assumed that Locke is, in fact, discussing memory when he writes about consciousness, since there is clearly a difference between the two concepts. That is to say, I can be conscious of a great many things, all of which do not require the slightest use of my memory. For instance, I am immediately aware of objects that I perceive through my senses. When I am immediately aware of an object, I perceive the object in that particular moment – it is not a matter of recalling a past idea of it. Memory, on the other hand, is referential to the past, which entails recollection. Although I can be conscious of my memories, I need not be remembering in order to be conscious of something.

The sense that we get from Locke’s use of *consciousness* is very much akin to our use of *memory*. As we have just observed, Locke remarks about consciousness extending into our past, uniting us with past actions and thoughts. This is precisely the notion we have when we speak about memory. If I make the claim, “I remember eating chocolate cake at my fifth-year birthday party,” then what my statement asserts is that I have a memory of that particular event. Although I could say, “I am conscious of eating chocolate cake at my fifth-year birthday party,” this seems a bit peculiarly stated, regarding a recollection. This is because consciousness, unlike memory, need not imply

recollection. We can easily demonstrate this notion by assuming that I made the statement at my fifth-year birthday party. By adding this fact, it is clear that my statement then becomes one regarding events of which I am immediately aware, since I am making the statement while eating the cake, as opposed to remembering the cake at some future time.

This criticism against Locke is fair, I think, although it has no real impact on the overall validity of his argument. There is no validity lost in the logic of his argument by conflating these terms, since the *concept* associated with the terms is what matters. Simply put, if Locke's use of *consciousness* carries the same logical tone (i.e. meaning) as our use of *memory*, then the two are really expressing the same idea. I think this is the general consensus view of Locke, since we find that all of the modern literature makes reference to his work in terms of memory, even though Locke specifically refers to consciousness. As such, I believe this is the least damaging of the criticisms against Locke's position, since the other three criticisms we cited earlier do take measures to weaken the foundations of logic in his argument.

Another common criticism raised against Locke's use of memory is that, taken literally, it demands too much. For instance, he states, "And as far back as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that Person..." (Locke, 335). A literal reading of this statement requires that our minds remember everything that has ever happened to us, in order to remain the same person. John Perry recognizes this problem and states that, if read literally, Locke's theory "...requires us to be able to remember everything that ever happened to us.... On the face of it, Locke has given us too stringent a necessary condition for an earlier

experience to belong to a person.”⁹ Obviously, no one can remember everything that has ever happened during his/her existence, and therein is the problem with a literal interpretation of Locke’s theory.

Predecessors of Perry’s position include Joseph Butler (1692 – 1752) and Thomas Reid (1710 – 1796). Both agree that it is not necessary for one to remember a thought or an act, in order to qualify as the one who had the thought or made the action. For instance, Reid states, “That relation to me which is expressed by saying that I did it, would be the same though I had not the least remembrance of it.”¹⁰ What is important, by Reid’s measure, is not the ability to recall, but that there is a genuine relation between the person and the thought or act in question. This relation, accordingly, exists independently of one’s memory and is of greater consideration when questioning the identity of persons. Butler agrees by confirming the idea that present consciousness of past actions is not necessary to our being the person involved in the events.¹¹

A further consequence of reading Locke literally, as Reid observed, is that Locke’s description of memory and personal identity breaks down the transitivity of one’s identity. That is, taken literally, Locke is dedicated to the position that if a person cannot remember a past act or thought, then that person is not the same person who executed the actions or thoughts in question. Reid demonstrates that Locke’s description makes it possible that, “a man may be, and at the same time not be, the person that did a particular action.”¹² In his example, Reid shows that a middle-aged man could remember an event from his childhood that he could not remember as an elderly man; yet the elderly man could remember events from when he was middle-aged, so that his consciousness remains uninterrupted throughout his whole life. The middle-aged man is the person who

is connected to both the child and the elderly man, even though the child is not connected to the elderly man. The result, says Reid, is that the middle-aged man is the same person as both the child and the elderly man; yet the elderly man is not the same person as the child.

Both Reid (214) and Butler (388) object to the idea that, as Locke's argument states, personal identity is defined by memory. They argue that Locke's claim is impossible, since memory presupposes personal identity. What this means is that for there to be a memory, there must first be a person that, in a sense, *owns* that memory. This kind of objection claims Locke is guilty of reversing the order of causation. The result, Reid states, is that memory is granted "a strange magical power of producing its object, though that object must have existed before the memory or consciousness which produced it." (214). This objection appears sound on the surface. Yet, if we restate the objection as a question, we see that the idea may not be quite so straightforward. For example, let us say we ask, "Is every memory produced by a person that, in some sense, owns that memory?" If our answer affirms this question, then some problem cases arise. For now we may ask, "What about such cases as animals or, perhaps, computers – do these objects qualify as persons, since they are capable of recollection?"

The notion expressed here is that if memory presupposes *personal* identity, then wherever we find memory we should, by necessity, find persons. Clearly, Reid and Butler are correct in that it is not possible for an effect to precede its cause (assuming a linearly unfolding timeline). What is in question, rather, is whether or not every object that produces memory is a person, since the claim is that memory presupposes *personal* identity. I only want to draw our attention to this problem here. These questions will be

examined at length, in Chapter 4, when we discuss how description affects our theories of personal identity.

The last charge against Locke's position we will consider is the argument that the memory criterion cannot define personal identity, because our memories are ultimately beyond verification. This argument is based on two assumptions: 1) memory is fallible to the point that we can have false memories, and 2) verification of memory claims requires some sort of qualification beyond our introspection. Again, Chapter 2 will deal with these ideas at length, but we will introduce ourselves to these criticisms, and some of their counterparts, here.

The first notion, of false memories, is certainly a problem for determining personal identity using the memory criterion. If false memories occur, which we perceive as events genuinely belonging to our past, then we are incorporating fictions into our concept of who we are. The fact that we make a distinction between genuine and false memories implies that we have a method for determining the differences between the two. To resolve the issue of false memories there must be some form of alternate verification we can use to determine the truth of a memory claim. Yet, the very fact that an alternative form of verification to introspection is needed demonstrates (or at the very least, implies) that memory *itself* is not enough to determine personal identity; rather, memory *and* some other phenomena may work.

If all memory claims were necessarily true, then we eliminate the need for alternative verification, since introspection alone would suffice. Yet, it is easy enough to demonstrate situations when claims we affirm are simply false memories, and it is in these situations that the need for verification arises. Seeming to remember an event is

slightly different, in that there are various ways in which I may be incorrect about the claim I assert. If I claim to remember helping lay the last stone atop the Great Pyramid at Giza, during its original construction, then clearly I am either lying or experiencing a false memory. However, on the other hand, if I claim that I seem to remember eating chocolate cake at my fifth-year birthday party, then I may genuinely have a vague recollection of the event, I could only be recalling the recounting of the event by others, or I could again be having a false memory, if in fact I never ate chocolate cake at this event.

What all of these examples demonstrate is that my introspection will not necessarily guide me to the truth of the matter. External verification, by the testimony of others, written documents, or some other source, is necessary to help support the validity of the memory claims I make. Yet, arguably, the external verification only brings us to the original objection – that memory claims are ultimately beyond verification. That is, we could argue that, even though we have external verification supporting a memory claim, the evidence we provide is itself in need of further verification. Thus, the objection is that no amount of evidence is going to produce unyielding certainty. This objection, however, has less to do with a deficiency of the memory criterion than it does the criteria we place on verifiability, as we will see in the details discussed in Chapter 2.

One of the results of the failure of Descartes and Locke to secure a definitive theory of personal identity is Hume's denial of the self. The lack of evidence for the existence of immaterial substances, coupled with the problems of the memory criterion, was enough to convince Hume that personal identity was, for the most part, created in the imagination by appearing to unite our sense-perceptions into a continuant self.

Hume's Denial and the Bundle Theory

From its very beginning, David Hume's (1711-1776) treatment reads more like a denial, rather than an affirmation, of personal identity. One of his first criticisms targets the fact that many philosophers of his day assume that we are intimately aware of our *self*. As noted earlier, both Descartes and Reid made such assumptions. Descartes drew on the belief that the nature of the self is revealed through introspection, while Reid argued that our identity is so simple a concept that any further proof only weakens the evidence of it.

Hume provides an explicit description of what he believes it takes for us to provide evidence for the existence of the self. In doing so, he also defines what others have traditionally believed it means to understand the self. For example, he states:

It must be some one impression, that gives rise to every real idea. But self or person is not any one impression, but that to which our several impressions and ideas are suppos'd to have a reference. If any impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is deriv'd; and consequently there is no such idea.¹³

Hume's idea, then, is that the self is not an immaterial substance in the Cartesian sense – one that can exist apart from the body, and which is seemingly present throughout our lives. Rather, he argues that our idea of the self is really an illusion created by the continuous overlapping of our sense-perceptions. If we remove these perceptions, we will not find some sort of immaterial phantom to which they are somehow attached, he argues. Instead, Hume states that the total removal or absence of perceptions is equivalent to annihilation, or a *non-entity* (252).

To what do we owe this apparent confusion about immaterial substances? It is Hume's belief that the problem arises from the way we judge the resemblance of objects. That is, where we typically attribute identity to an object, we should in fact attribute resemblance or similitude. This notion is not unique to Hume, for Reid also argued for this position (206). However, Reid was not drawn to Hume's conclusion that there is, therefore, no permanent self.

Arguably, there is no evidence (i.e., no conclusive evidence) to support the concept of immaterial substances. A significant part of the problem is that we cannot conceive of ways to go about verifying or testing for the existence of such substances. How does one measure (i.e. account for) the properties of immaterial substances? We will pursue this problem in detail in the following chapters. To the extent that we have little reason to believe in the existence of immaterial substances, we find some degree of support for Hume's denial of the self as such. This sentiment is implied by Paul Churchland when he writes, "If this is the correct account of our origins, then there seems neither need, nor room, to fit any nonphysical substances or properties into our theoretical account of ourselves. We are creatures of matter. And we should learn to live with that." (21).

Does this kind of materialism necessitate Hume's conclusion that we have no idea of the self? Unless we are willing to admit that our concept of self is really an empty concept, then I think the self must at least exist as some kind of idea. That is to say, we must at least have some idea of the self, no matter how vague or definitive, if our assertions about the self are to be applicable to something. Dennett reinforces this when he states, "If selves are anything, then they exist."¹⁴ In Chapter 3, we will see that

Anthony Quinton agrees, because he believes that we can talk about the “nonbodily aspect of a person,”¹⁵ without reference to an immaterial substance.

What are we to make of Hume’s claim that *real* ideas derive from single sense-impressions? Depending on how strictly we wish to interpret Hume’s words, some potential problems can arise. A strict interpretation of Hume yields some peculiar results. For instance, strictly speaking, it seems we could argue that nothing we observe results from a single impression. Hume’s objection was that there was no single impression or idea that gives rise to the self, because it is his assumption that, if the self exists, then it is simple and must arise from a single impression. Let us suppose we are observing an object, such as an oak tree. Do we describe the oak tree as a real idea that issues from a single sense-impression? Clearly, there are many sense-impressions that make up our idea of the oak tree. There is the impression we receive by looking at it, as well as those we receive if we touch it or smell it, or, if peculiarly hungry, taste it. So, it is not from a single sense-impression that we get our idea of the tree, but from several such impressions.

Are we to conclude, then, that all observable objects are products of our imagination? The question is, then – what qualifies as a single sense-impression? Is everything we experience about the oak tree collectively counted as a single sense-impression, since it is a single object, or do we break down the oak tree into the information we receive through our individual senses? If the latter, then this seems to imply not only that there is no single impression of the self, but also that there is no single impression of anything. The implied result is that *everything* is a product of our imagination.

Hume's account regards the self as a kind of abstraction, created by our imagination's ability to feign the unity of our sense-perceptions. For example, according to Hume's descriptions of the self, we can think of it as analogous to our concept of a crowd. That is, when we observe a *crowd*, all we really observe is a collection of individuals, which is unified into an organized whole by our imagination. The crowd is essentially an abstraction our brain creates in order to categorize the collection of individuals. Accordingly, a crowd exists as a description (i.e. a compound of ideas, as opposed to a simple idea), but not as something real. As such, individual members of the crowd can change, along with the actual size of the group, without necessarily affecting our description. Although we are not likely to say that the crowd is the *same* crowd if many of its original members leave, nevertheless we use the same description to apply to any significant number of individuals. We have no single impression of the crowd. What we have, rather, is a collection of the impressions of individuals. Similarly, regarding the self, says Hume, we have no single impression of it, but instead have a collection of individual sense-impressions, occurring uninterruptedly. So, the two are analogous in that we have a concept (crowd) that unifies the individuals we perceive as a group, just as we have a concept (self) that unifies our individual sense-impressions.

It is clear that Hume expected to find a single impression of the self, if such a phenomenon were to exist. Since he believed his introspection failed to reveal this impression, he was brought to the conclusion that there is no such thing as the self. Accordingly, we create the illusion of a continued self from the overlapping of our perceptions. The perceptions themselves, Hume argues, are all we truly perceive, not a

continued self. For he states, “I never can catch *myself* at any time without a perception, and never can observe any thing but the perception.” (Hume, 252).

Hume’s concept of the feigned production of a self, through the imagination, has become known as the *bundle* theory of the self, because it maintains that our *self* is really nothing more than a bundle of disjointed perceptions, which are seemingly unified by the imagination. We must return to the idea that both he and Reid agreed upon, in order to understand more fully why Hume adopts this position.

What we find in Reid and in Hume are the beginnings of what is commonly referred to, nowadays, as *type* and *token* identity. Identity is divided into two basic descriptive categories, using this distinction. Perhaps the best example of describing this idea is done as Derek Parfit does, speaking in terms of qualitative and numerical identity.¹⁶ A *type* identity, then, is identity of similar kind or category. A *token* identity, in contrast, is a specific member or instance of a type identity. For example, Reid argues that identity cannot be applied to our sensations or any operations of the mind, because “The pain felt this day is not the same individual pain which I felt yesterday, though they may be similar in kind and degree, and have the same cause.” (202). Similarly, Hume states, “Thus a man, who hears a noise, that is frequently interrupted and renew’d, says, it is still the same noise; tho’ ‘tis evident the sounds have only a specific identity or resemblance, and there is nothing numerically the same, but the case, which produc’d them.” (258).

Hume’s argument, like that of Reid, is that we mistakenly attribute identity in cases that are merely instances of similitude. This position rests on a very strict conception of identity, where only a one-one relation qualifies as genuine identity. For

instance, a one-one relation states that a thing is only equal to itself (e.g., $A=A$). Since various perceptions such as emotions and sounds happen in individual instances, we cannot therefore, strictly speaking, attribute identity to them over time. Other instances of what we would call the *same* emotions or sounds are, accordingly, instances of exact similitude, rather than genuine identity.

Each instance Hume and Reid described above is a separate perception united only by resemblance. Hume believes that our mistaken application of identity happens because of our concepts regarding change. That is, he claims that 1) the amount of change a body undergoes, in proportion to the whole, is what determines how the mind ascribes identity, and that 2) we are less likely to say identity is destroyed if change is gradual, rather than sudden and all at once (Hume, 256). His conclusion is that what generally counts as identity, is merely a quality attributed to these perceptions by uniting their ideas in the imagination when they are reflected upon (Hume, 260).

Memory is also a phenomenon that Hume believes contributes to personal identity. However, unlike Locke's treatment, he claims that we can extend our personal identity beyond memory to include those things we have forgotten, but could *possibly* remember, since there are things that we did and thought, which we cannot recall, yet they are part of our pasts. Without memory, says Hume, we would have no notion of causation or the chain of causes and effects, which constitute our self or person (261-62). This point is put to the test when we consider problems such as amnesia – a total memory loss. Yet, even in instances like amnesia, we find that what we may say varies based upon how we define the concepts associated with the problem. For example, in the next chapter we will see that there are various ways we can conceptualize amnesia. The

assertions we make about the effects of amnesia on personal identity will be determined by what we consider amnesia to be. This brings us to the conclusion Hume makes in his writing – that the questions we have about personal identity are not really philosophical in nature, but are, rather, regarded as grammatical matters (262). I think we will see more light brought to bear on this idea as we continue along our examination in the chapters ahead.

Hume's conclusion, above, is no small charge against the logical foundations of Descartes' and Locke's arguments, because what Hume implies is that all of the work regarding the analysis of personal identity thus far has been done on dubious grounds. The result is that we must use an entirely different approach to the problem. A further consequence, as we will see, is that there is a degree of open-endedness, or an arbitrary nature to the answers of our questions about personal identity.

In fact, what I will show is that even though contemporary philosophers may or may not agree with Hume's bundle theory and his ideas about the self, what we find is that this one kernel of thought – that personal identity is ultimately a matter of grammatical convention, not a philosophical difficulty – characterizes, at some level, the majority of contemporary discussions. In order to demonstrate this, however, we must further examine the concepts introduced in this chapter. Our focus now turns to the contemporary counterparts of what we found in Descartes, Locke, and Hume.

Notes

¹ Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Basil Blackwell, 1985) 69.

² Rene Descartes, *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. II (Cambridge: Cambridge UP, 1995) 16.

³ John Searle, *Minds, Brains and Science* (Cambridge: Harvard UP, 1986) 14.

⁴ Paul Churchland, *Matter and Consciousness* (Cambridge: MIT P, 1994) 9.

⁵ See, for example, Descartes' account of the interaction between the brain and the soul, as described in *The Passions of the Soul*, as well as his description of "animal spirits" in *Treatise on Man*.

⁶ John Locke, *An Essay concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon, 1975) 330.

⁷ David Wiggins, "Locke, Butler, and the Stream of Consciousness," *The Identities of Persons*, ed. Amelie Oksenberg Rorty (Berkeley: U of California P, 1976) 152-53.

⁸ See, for example, Jonathan Bennett's comments on Locke's philosophy of mind. Also see what Thomas Reid and Joseph Butler argue forthcoming in the text.

⁹ John Perry, "The Problem of Personal Identity," *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 15.

¹⁰ Thomas Reid, "Of Identity," *Essays on the Intellectual Powers of Man* (London: Macmillan, 1941) 204.

¹¹ Joseph Butler, "Of Personal Identity," *The Anthology of Religion* (London: Bell and Daldy, 1857) 389.

¹² Thomas Reid, "Of Mr. Locke's Account of Our Personal Identity," *Essays on the Intellectual Powers of Man* (London: MacMillan, 1941) 213.

¹³ David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, second ed. revised by P.H. Nidditch (New York: Oxford UP, 1978) 251.

¹⁴ Daniel Dennett, *Consciousness Explained* (New York: Little, Brown & Co., 1991) 413.

¹⁵ Anthony Quinton, "The Soul," *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 57.

¹⁶ Derek Parfit, *Reasons and Persons* (New York: Oxford UP, 1984) 202.

CHAPTER 2

THE MENTAL PHENOMENA

Mental Phenomena and Personal Identity

We will now examine the various aspects of personal identity typically classified as mental phenomena. Although it is arguable that mental states are ultimately reducible to physical states, it is not my goal to pursue such a question here. Rather, I am merely using the description *mental* to distinguish the various phenomena that we often associate with the brain. This examination is broken down into two primary themes: the memory criterion and the psychological criterion. Both play an important role in contemporary discussions about personal identity – the memory criterion, for its origins in Locke’s treatment and his influence on modern thought about personal identity; the psychological criterion, since it has developed out of our discussions about memory.

My purpose here is not to support or reject either of these theories. Instead, I am merely presenting the claims of each, then examining the objections and criticisms often raised against them. In the chapters ahead, I will offer more of my own remarks about memory and psychological continuity, as they pertain to our personal identity, and discuss why I think that both of these criteria fail to provide us with a conclusive definitive theory.

The Memory Criterion

One of the first things we notice, when examining most contemporary discussions of memory, is that it is often categorized into several types, rather than being viewed as a generalized phenomenon. For example, both Perry¹ and Shoemaker (Shoemaker and Swinburne, 86-87) make a distinction between event memory and factual memory. Event memory is the type that occurs when we remember an actual experience from our past, e.g. remembering our high school graduation ceremony. Factual memory, by contrast, is remembering *that* something is true. A genuine causally connected experience with the fact in question is not a necessity in such a case, since factual memory can (and does) expand beyond the limits of one's personal experiences. That is, we can all remember that Columbus sailed from Spain in 1492, yet there is no one alive today who actually witnessed this event. So, then, the difference between these two types of memory is that, in one case, the person remembering was actually present at the time of the event, while in the other case a fact is merely made evident.

Being present for an event type memory is what Perry refers to as the Witnessing Condition (*Personal Identity* 144). The development of this kind of argument, we shall see, is what caused the memory criterion of personal identity to evolve into the broader concept of the psychological criterion. The importance of the Witnessing Condition, and similar ideas, lies in the need for memory verification.

Verification becomes important when we begin to examine the truth-value of memories. That is, philosophers make a distinction between a genuine memory claim and one that is only *seeming* to remember. These seemingly genuine memories are

referred to by various names, such as *apparent* memories or *false* memories. Richard Swinburne (9) refers to notions of strong memory (an actually remembered event) and weak memory (false memory). No matter the terminology, we clearly understand there is a difference between something that is an actual, genuine memory of past experience, and something that is mistakenly believed to be part of one's past experience. How then are we to determine whether a memory is genuine or false?

When we make a memory claim, we are making an assertion about something from the past. Typically, when we speak about personal identity, we are making statements about something, which we believe to be applicable to *our own* past. Yet, not all assertions about our past may be the same. Bernard Williams states that we make three kinds of distinctions about our own past: 1) recalling, 2) reminding, and 3) learning again.² In the first case, there is no occurrence of new input, writes Williams. The second and third instances involve partial new input and total new input, respectively.

An example of recalling is the immediate recital (i.e. *logically* immediate, or without being prompted) of some fact with which we are familiar, such as our name or age. In such a case, we are simply reciting a fact without hesitation or forethought. We can think of recollection as *effortless*, because there is no need for external assistance and the information is something with which we are immediately aware. Reminding is a situation where we are familiar with a fact, but it is not evident to us without some sort of external prompting. For instance, I may have forgotten about playing a game of Pin the Tail on the Donkey at my fifth birthday party, until reminded by my parents. An important point here, however, is that the prompting

must actually cause me to remember the event, and not merely cause me to seem to remember. Clearly, there are often times when we hear stories recounted to us repeatedly, for which there comes a point when we question whether or not we actually recall the event or merely its recounting. It is quite possible that hearing such stories makes us seem to recall the event, simply because we have heard the story so many times, when in fact we have no real recollection of what occurred.

In the last case, learning again, there is a fact with which I was once familiar, but cannot now recall, even with another's prompting. Perhaps, when I was very young, I learned a bit of a foreign language, which I had not spoken for 15 years afterwards. If no prompting allowed me to recall any of the things I previously knew about this language, then I would have to relearn as someone with no previous exposure to the language. Williams addresses these distinctions while discussing the notion of memory and its causal connections. His idea here is that a memory must be causally *linked* to a past experience in some way. Furthermore, Williams states that our memories operate in connection with our emotions, both presently and for our future. Our reactions and emotions, he continues, can be expected to change, if our memories about these events are changed (*Problems of the Self* 188).

Addressing this concept of causal connections, Perry concludes, similarly, that memory is not so much explained by the accuracy of the claim as it is in terms of a relation (*Personal Identity* 149). This relation adopts the same kind of historical tone we read in Williams, in that experiences that *bind* identity, as it were, must be connected or linked in a special way. The causal chain that links us to our past, then, is of greatest importance for validating memory claims. Accordingly, we go about

validating a memory claim by tracing backward along the history or lineage of such a causal chain; in much the same way we would retrace a family history to discover our family's lineage. To illustrate this notion, we can think of the individual members of our family tree as analogous to the individual experiences of a person – we must be connected to past experiences, by memory, in the same kind of determinate way we are connected to members of the previous generations of our family. But is it possible to have this kind of determination with regard to memories?

Although we now have an idea of what would constitute a genuine memory and a false memory, there are other problems to consider about the memory criterion. For instance, it becomes obvious that if we try to use our own memory as the sole means of validating claims about our past experience, then our attempt is self-referential. This is like reading a story in a newspaper and then, in order to verify the accuracy of the article, purchasing another copy of the same newspaper.³ There is no external verification in such a situation.

When we speak about verifying a memory claim, we must not adopt the extreme skeptical view sometimes present during discussions of verification. Instead, the kind of verification we are seeking is more akin to what Alfred Jules Ayer has in mind, when he distinguishes between practical verifiability and verifiability in principle. He perfectly summarizes this when stating, “For it must surely be admitted that, however strong the evidence in favour of historical statements may be, their truth can never become more than highly probable.”⁴ It is clear why this is so important for discussions about the memory criterion. As we have seen, memory claims assert something about the past, thus making them a brand of historical

statement – a description relevant to both event and factual memory statements.

Since, then, we can never be certain about historical statements, we only need to concern ourselves with the practical verification of memory claims. In other words, all we are looking for to verify a memory claim is something more substantial than the words coming out of our mouths. Personal testimony about one's own past is, by itself, insufficient.

An external source of verification is a necessity recognized by both Williams and Shoemaker.⁵ They agree that there must be some sort of documentation or testimony of others, which allows us the ability to sort fact from imagination. Yet, as Williams observes, all this will tell us is whether or not a particular body was present at a given time (*Problems of the Self* 6). He introduces this notion because one could argue, for various reasons, that bodily identity is not a criterion of personal identity.⁵

Perry, on the other hand, introduces a different, potentially larger problem regarding memory. He states that we have no common understanding of the causal mechanisms of memory, and that “no such process can be observed by the ordinary man, introspectively or otherwise.”(*Personal Identity* 146-47). This is no trivial claim; it is a glaring reminder of the limits of what we know about brain processes at this time. Yet, Perry's assertion need not herald the end for theories that rest upon the memory criterion. Only if we insist that we must understand the causal mechanisms of memory, before being able to theorize about the phenomenon, does Perry's statement create difficulty.

While Perry's observation about memory may be true, it certainly has not stopped us from theorizing about the memory criterion. Perhaps, we may argue,

understanding the causal mechanisms of memory is *not* necessary for constructing a theory of memory criterion. In the same way we do not have to understand how an engine operates, in order to drive a car or teach someone how to drive, perhaps there is no conflict in our theorizing about memory, without first knowing the fundamental workings behind the phenomenon. After all, we may not know what *causes* memory, but we certainly know what memory *does*. As Shoemaker observes, "...if the word 'remember' is to have any meaning it must be possible to establish whether someone is using it correctly."(*Personal Identity* 128). He believes that this involves observing how a person uses the word in various circumstances over a period of time. And maybe that is all we need to know to apply it to our concepts of personal identity.

Furthermore, Shoemaker argues that the fact that memory claims are generally true is a conceptual truth (*Personal Identity* 129). This means that it is part of our concept of persons, he argues, that they are able to make true statements about their past. Shoemaker asserts that the truth-value of a memory claim is a conceptual truth for three reasons: 1) he believes that the memory criterion is not the sole criterion of personal identity; 2) he believes bodily criterion is not the sole criterion of personal identity; 3) he believes that we must have some way that is not self-referential to verify our memory claims.

Shoemaker's conclusion, then, is that we really use no criterion when making claims about personal identity. Instead, he argues, the verification of our memory claims comes from our conceptual truth about persons, which states that persons can generally make true memory claims about their own past. While it is generally true that persons make accurate claims about their past, as Shoemaker asserts, I am not

sure about his notion of it being a conceptual truth about persons. Examining a problem case may help demonstrate the difference between Shoemaker's claim of conceptual truth and what I believe is a matter of contingency.

When Memory Fails

Before examining amnesia to test Shoemaker's assertion that the truth-value of memory claims is a conceptual truth of persons, let us briefly explore what can be said about the phenomenon in general, as it applies to the memory criterion. Amnesia is an obvious stumbling block for any theory of personal identity that attempts to rest solely on the memory criterion. Yet, this obstacle can be overcome, if we are willing to make certain qualifications to the memory thesis. If a total memory loss occurs, then how do we account for personal identity? Let us weigh the options. In a case of amnesia, we can say that 1) personal identity is destroyed, 2) personal identity is changed to some degree, but not completely, or 3) personal identity is unaffected. I suggest that how we answer this depends on the importance we place on the memory criterion's role in defining personal identity, which, in turn, depends on how we define persons.

In the first case, we would maintain that personal identity consists solely of memory and, therefore, a total memory loss will destroy one's identity. The problem with this position is twofold: we must be able to demonstrate that personal identity consists solely of the memory criterion, and we must be able to demonstrate that a total memory loss is possible. Swinburne states, "Quite clearly, we do allow not merely the logical possibility, but the frequent actuality of amnesia – a person forgetting all or certain stretches of his past life." (24).

Swinburne's statement does not allow us to draw any definitive conclusions, however. Even though we do recognize amnesia as an actually occurring phenomenon, it is not always defined as the *total* memory loss he describes. Shoemaker believes amnesia does not pose a problem for the memory theory, because it never renders a total memory loss in the strict sense (Shoemaker and Swinburne, 86). Similarly, Perry thinks a person's identity is preserved during amnesia, if we are willing to employ the notion of *possible* memories. That is, if we include memories that one "would have if he hadn't been conked on the head,"⁶ then memory is still person preserving. Otherwise, the memory criterion is too stringent to allow personal identity to persist in such cases.

The fact that we hear about instances where people recover memories, gradually or suddenly, seems to support Shoemaker's position. But this does not mean, therefore, that his assertion holds in all cases. There may be cases where, as Swinburne suggests, memory is permanently and totally lost. A total memory loss is certainly not a logical impossibility, nor an absurdity. Even if such a case never arises, we cannot dismiss Swinburne's claim outright, as it is at the very least philosophically important because it is a conceivable case, and should therefore be entertained during our examination of amnesia and its *possible* effect on personal identity.

In the second case, we claimed that amnesia causes personal identity to change by some degree, but not completely. There are several reasons why this would be considered a suitable description. To begin, even though persons suffer from amnesia they are still the same physical body. And, as we have seen, it is

doubtful that amnesia causes a total memory loss in the strict sense. Personal identity is therefore changed, we might reason, by the degree of memory loss. This argument deviates from the idea that the memory criterion is the sole criterion of personal identity. Instead, this theory rests on the claim that bodily identity, at least to some degree, is also a criterion for determining personal identity, or that other mental criteria must be considered in order to determine personal identity. The inclusion of other mental phenomena is developed in the concept of psychological continuity. Of the various phenomena that sustain psychological continuity, memory is merely one aspect (Shoemaker and Swinburne, 90).

However, this second case also allows us the possibility of claiming that, even if a person cannot remember anything about his/her past, this is not enough to destroy personal identity. In this instance, our theory rests on an assertion like Perry's notion that possible memories prohibit a total memory loss. Part of this claim rests on the assumption that, although inaccessible, our memories are still in there somewhere. I find this assertion as unsatisfying as one that begins from a belief in immaterial substances, in that both are beyond the realm of verifiability.

Clearly, it is arguable that personal identity is *not* a matter of degree and, therefore, this second version of the theory cannot account for it. Personal identity, it may be said, is all-or-nothing and is therefore determinate or determinable at any given time. The all-or-nothing position defines personal identity as a one-one relationship, in that a thing at time T_1 is the *same* thing that exists at time T_2 . I will discuss this idea at length later in this chapter.

The third case allowed us to claim that personal identity is entirely unaffected by amnesia. This means that either 1) the memory criterion plays no role in determining our personal identity or 2) no amount of memory loss would change what we say about our identity, since total memory loss, in the strict sense, is impossible and, furthermore, memory is not the sole criterion that determines personal identity. How would we defend this position? If our own memories do not determine who or what we are, in any way, then what criterion will we use? The typical alternative to the theory of memory criterion is the theory of bodily criterion. We will explore this and other options in the next chapter.

Now, with regard to Shoemaker's claim – that a person's ability to make true statements about his/her memories is a conceptual truth – what sort of assertions can be made to support or reject its validity? If the fact that persons can generally make true statements about their memories is a conceptual truth, then how is this affected by the possibility of amnesia, if at all? If we grant Shoemaker's claim, then it appears we can say two things: 1) People who suffer from amnesia are no longer persons, or 2) personal identity is not affected by the person's state of amnesia. Here, as before, what we will say depends on what we accept as a viable definition of amnesia. Since Shoemaker rejects a total memory loss in the strict sense, he would likely assert the second of these two statements. Although our first statement may seem ridiculous, it is nevertheless a condition set up by Shoemaker's position, unless we absolutely reject the idea that amnesia can cause a total memory loss, in the strict sense. Let us set up an example that demonstrates the consequences of not rejecting this claim.

Suppose that Michael Ellis suddenly suffers from amnesia. If amnesia did, in fact, render a total memory loss, then not only would Michael Ellis no longer be the same person he once was, he would no longer be a person at all, according to Shoemaker's position. As such, Michael Ellis fails to satisfy the concept of a person, since part of Shoemaker's concept requires that he be able to make true statements about his past. Our objection here relies on the fact that we describe amnesia as a total memory loss, which Shoemaker explicitly rejects. Yet, as I stated earlier, a total memory loss is neither a logical impossibility nor an absurdity. The fact that Shoemaker rejects a total memory loss does not preclude the possibility of such an event. Given this, I think we cannot accept Shoemaker's claim as a conceptual truth, but rather as a contingent one.

The Psychological Criterion

As we have already seen, one of the criticisms of the memory criterion, particularly of Locke's treatment, is that it is too limited. That is, if we read Locke's position quite literally we might state that a complete account of personal identity must include properties such as character, values, intentions, and the like. However, some, such as Perry⁷, state charitably that Locke would agree with this idea, based on the "level headedness" he demonstrates in the *Essay*. While this is certainly possible, I prefer not to speculate about what Locke might say were he still among the quick. What cannot be denied, however, is that these properties certainly do provide much vitality for our discussions about personal identity. The addition of these features has helped develop the memory criterion into the psychological criterion. And, as we

will see, having a fundamental tie to the memory criterion brings many similar problems into discussions about the psychological criterion.

The psychological criterion is broken down into two fundamental ideas: psychological continuity and psychological connectedness. When we speak about psychological continuity, we are referring to the holding of an overlapping chain of psychological relations.⁸ Similarly, psychological connectedness refers to the holding, over time, of particular *direct* relations. Continuity, then, is a concept more concerned with our immediate relations. That is, we look for moment-to-moment continuations of these relations and, therefore, continuity can be defined without degree. Yet, connectedness has degrees, since these direct relations (memory, character, intention) hold variously during different parts of our life (*Identities* 98). David Lewis describes continuity as “the existence of step-by-step paths from here to there, with extremely strong local connectedness from each step to the next,” (*Identities* 18) and connectedness as “direct relations of similarity and causal dependence between my present mental state and each of its successors.” (*Identities* 18).

We can already begin to see the similarities between this position and the memory criterion. Both are concerned with a kind of causal link or connection of mental phenomena. The primary difference is that the psychological criterion goes on to include the relations of character, values, desires, intentions, etc. We now have a way in which one’s identity may hold, even if a total memory loss were possible. Since memory is no longer the sole criterion of the concept, our identity may persist based on the strength of the relations of these other psychological relations, even if

we lose our memories. Shoemaker believes this is a revision that answers Locke's critics, since what is needed is "that one have 'memory continuity' with that past self-memory continuity consisting in the occurrence of a chain of memory-connected person-stages...." (Shoemaker and Swinburne, 81).

The theory of personal identity based on the psychological criterion is often referred to as the Reductive View, because our identity, as such, is described in terms of various inter-connected relations, rather than in terms of a single subject of experience. This reductive classification is a byproduct of Cartesianism. Materialism is traditionally described as a reductionist view of the dualist theory of mind, so too is the psychological criterion described with regard to personal identity. With the rejection of the traditional view, we will see some rather peculiar explanations about persons and their identities, especially when we examine many of the proposed problem cases.

The psychological criterion, like the memory criterion, is understood as a kind of causal relation. That is, it was posited that memories must be linked to our past in a special way, i.e. they must represent our involvement in an actual event. In this way, a causal chain links us to that event. The concepts of continuity and connectedness are the means that explain causal relations for the psychological criterion. Therefore, we are addressing instances of sameness of character, sameness of values, and of intentions. But, as we saw, these are relationships that hold by matters of degree. How then does this affect our view of personal identity?

An aspect of psychological connectedness, advanced by Parfit, is the Psychological Spectrum. This concept explains how we understand connectedness as

a matter of degree. In the near end of this spectrum, he states, there is no change in a person's psychology. At the far end, however, all memories are lost and replaced by apparent memories (*Reasons and Persons* 232). Although Parfit mentions only memories here, we can just as easily imagine that the spectrum involves properties such as character, values, intentions and other mental phenomena, since this has been included in our concept of the psychological criterion. But Parfit does not limit his explanation to mental phenomena. He also describes a Physical Spectrum and a Combined Spectrum as part of his theory. Both of these notions operate like the Psychological Spectrum, i.e. there is someone completely continuous at the near end, but totally dissociated at the far end. So, then, a person who is the same from one time to another will be at the near end of the Combined Spectrum (i.e., the near end of both the Psychological and Physical Spectrum). As we move further toward the far end, the person is less similar and, therefore, less of a continuous person with the original.

When we ask, "Will this still be me or someone else?" we typically assume that this question has a determinate answer. Our common sense seems to tell us that there should be a sharp borderline that allows us to determine, at any given time, whether a person is the same or not, from one time to another. Yet, a sharp borderline such as this is something Parfit denies, unless, he says, we are willing to say that we are separately existing entities, in the Cartesian sense, which of course he rejects (*Reasons and Persons* 239). Instead, he argues that there is no fact involved which is all-or-nothing; a person's physical and psychological connectedness could hold to any degree. This alternative to the all-or-nothing view implies that we arbitrarily choose a

point at which we call a person the same, since there is no objective means to determine where such a borderline would lie. And it is this idea that Parfit accepts, for he states, “We must pick some point on this Spectrum, up to which we will call the resulting person me, and beyond which we will call him someone else. Our choice of this point will have to be arbitrary.” (*Reasons and Persons* 241) He concludes, similarly, “By drawing our line, we have chosen to *give* an answer to this question.” (*Reasons and Persons* 241). This idea is not unique to Parfit, however. Lewis affirms this idea, but, unlike Parfit, he offers what he thinks is a solution to making an arbitrary decision about where to create such a cutoff point. For instance, like Parfit, he says that when dealing with matters of degree we can introduce a cutoff point, and that this choice is arbitrary. The solution, Lewis argues, is that when we are faced with an arbitrary choice, the thing to do is not make the choice.⁹

How do we get around making an arbitrary decision, as described above by Parfit and Lewis? Lewis believes we avoid arbitrary choice by adopting the ideas of person-stages and tensed identity, i.e. identity at a particular time (36-37). By doing so, he believes that we can claim both survival and identity are what matter for personal identity (Lewis, 18-19). However, he does admit there is a discrepancy in the way we view these concepts:

He who says that what matters in survival is a relation of mental continuity and connectedness is speaking of a relation among more or less momentary person-stages, or time-slices of continuant persons, or persons-at-times. He who says that what matters in survival is identity, on the other hand, must be speaking of identity among temporally extended continuant persons with stages at various times. (Lewis, 20).

We will turn to a detailed examination of tensed identity later in this chapter. For the present time, we turn our attention to some of the problem cases of the psychological criterion. In doing so, we will be able to see how well our traditional concepts of persons hold up under hypothetical circumstances.

Division, Replication and other Problems

Two important concepts that are often explored in problem cases are division and replication of persons. Numerous hypothetical situations have been created to illustrate these ideas. Parfit describes a case of identical triplets involved in an accident, in which one brain is halved between the two surviving brothers. That is, the accident renders two of the three brain dead, while the third has a functioning brain and mutilated body. Therefore, the working brain is split between the two vital bodies. Under these circumstances, Parfit believes we can say that one of the following options must happen: 1) the original person associated with the brain does not survive; 2) the person survives as one of the two people; 3) the person survives as the other of the two; 4) the person survives as both (*Reasons and Persons* 254-256). What objections might we make about these claims?

Parfit states that we can object to the first claim, above, because survival can occur if half of the brain is successfully transplanted. Therefore, how can a double success be a failure? He also claims that the second and third are implausible, because there is nothing that should qualify either of the two as the genuine survivor, while excluding the other. Hence, he says, "If I survived as one and not the other, then the other would falsely think he was me. But I have no way of knowing if I am the one with the false belief. This is inconclusive, therefore." (*Reasons and Persons*

258) The only explanation is that the fourth claim is correct. Yet, it is arguable that we might also assert that there is now one person living in two bodies. Parfit admits that this claim cannot be dismissed outright (*Reasons and Persons* 256).

The two brothers receiving the divided brain are very closely connected to the original person. If we suppose that each half of the brain possessed the same contents as the other (e.g. both had all of the same memories, etc.), then the two are psychologically identical with the original person. However, even if we do not grant this supposition, the two brothers are still very closely connected to the original person. Connectedness becomes less distinctive further away from the source of origin. A greater amount of division involved in an instance results in less connection (*Reasons and Persons* 300).

Shoemaker states that our concept of personal identity must be compatible with the logical principles that govern identity in general (Shoemaker and Swinburne, 71). This creates yet another problem, since identity is considered a transitive, one-one relationship. Thus, there is a direct conflict between our concept of identity and the psychological criterion, since this criterion admits to be a matter of degree and, furthermore, can be a one-many relationship.¹⁰ For this reason, some philosophers think that we must abandon the idea of personal identity; or they argue that what matters is not identity, but survival. Others, however, simply see this as an admission of the failure of the psychological criterion to explain personal identity. Williams agrees that, in cases of fission, identity does not hold, since spatio-temporal continuity is broken (*Problems of the Self* 24).

Replication presents problems similar to those we witness in cases of division. For the purposes of discussing the psychological criterion, mental replication usually involves the concept of brain replication or merely the contents of one's brain being replicated. Although division is a form of replication, there is another type of replication I now have in mind, namely duplication. Williams, like Parfit, believes that replication contradicts the concept of identity. His examination helps clarify how we are to understand this by making a distinction between identity and exact similitude. To illustrate this, we can think of so-called *identical* twins. While we use *identical* to describe such people, they are not identical in the strict sense, since they are numerically different. What we really mean when we call them identical is that their appearance is exactly similar. For, even though they look similar, nevertheless, they have separate personal lives and are in different spatio-temporal points.

There are several reasons why spatio-temporal continuity is a consideration worthy of our attention. To begin, we can think of its application as a type of historical enquiry, not unlike the causal chain associated with the memory criterion. Ideally, this historical procession of events would reveal any instances of reduplication, so long as we can trace the histories back far enough (Williams, 24). And, as we shall see in the following section, spatio-temporal continuity involves our conceptions about time. When we theorize about personal identity, we generally make our assertions based on a linear concept of time. Yet, I will demonstrate that there are some rather peculiar possibilities for our theories of personal identity, if we adopt the notion of a non-linear time flow. Additionally, spatio-temporal continuity draws our attention to the importance of the bodily criterion, one might argue. That

is, Williams states, "...it is unclear what it would mean to say that there were *two* men who had exactly similar or the same memories, since to call them memories is to imply their correctness." (*Problems of the Self* 9). This observation is made to rule out the idea that memory is the sole criterion of personal identity. In cases where memory is replicated, continuity of body must be used in order to verify memory claims. The causal connection traced from one body back to a specific spatio-temporal point is our only means of distinguishing between two persons claiming to have exactly similar memories. This is affirmed by Williams when he says, "The only case in which identity and exact similarity could be distinguished, as we have seen, is that of the body.... Thus I should claim that the omission of the body takes away all content from the idea of personal *identity*." (*Problems of the Self* 10).

Williams thinks that if we were to replicate a person's memory, then we introduce the concept of *person-types*. He comes to this conclusion by way of the notion of an information swap, in which a person's mental contents are removed from his/her brain and mechanically transferred back into the same brain, after a reconstructive brain surgery. If we can do this with the same brain, then the same would be possible with an entirely new brain placed in the same body, he argues. And, furthermore, the fact that this would work with a new brain implies the possibility of duplication and, therefore, his notion of person-types (*Problems of the Self* 79-80).

Person-types are important for the concept of replacing ideas of personal identity with ideas of survival, because they allow us to understand how psychological continuity *can be* a one-many relationship. I emphasize this as a

possibility, rather than a necessity, because in cases where duplication does not occur (i.e. under normal conditions) psychological continuity is a one-one relationship. The hypothetical cases merely demonstrate that this need not be so. Parfit supports this idea when he states, “even if psychological continuity is neither logically, nor always in fact, one-one, it can provide a criterion of identity. For this it can appeal to the relation of *non-branching* psychological continuity, which is logically one-one.” (*The Philosophy of Mind* 150).

Parfit’s claim, then, is that under normal (non-branching) conditions, psychological continuity can provide a criterion of personal identity. If we deny this, based on the all-or-nothing view of identity, then Parfit thinks we must abandon the language of identity. We would then speak in a new way, regarding our new descriptions as having the same significance as identity (*The Philosophy of Mind* 151). By replacing ‘personal identity’ with ‘survival’ we can do this, he thinks. Perry, however, does not think Parfit’s new way of thinking is possible. He maintains that we will merely be abandoning one way of talking about objects in favor of another, but that there will still be identity among the new objects: “As long as one has predication, one will have identity....” (*Identities* 87). To this extent, I think I must agree with Perry. I see no difference in the logical tone of questions such as, “Will I really survive from my present state to a future state?” and “In twenty years will I be identical with the person I am now?” Both questions seek to answer the same fundamental end – At any given future time, will there be someone who really is me?

Returning to the fourth conclusion, from Parfit's example above, we can see why personal identity could be insufficient to describe this case. If we refer to both of the surviving brothers as Michael Ellis (to whom the original brain belonged), then Mr. Ellis will, from this point forward, have two bodies. Furthermore, each body would have a separate history, henceforth – leading to the possibility that one body might kill the other. Would we then say that Michael Ellis killed himself, was then convicted of his own murder, and died again in prison 40 years later (*The Philosophy of Mind* 146)? The standard way we talk about persons does not allow for such cases. It seems, then, that we must either change our concept of persons to include these cases, or develop a new way of describing these cases that does not imply identity. This is exactly what Parfit champions in his idea of survival. In the chapters to come, we will further investigate the role of language in the development of our concepts of personal identity.

Persons Through Time

Of all the criteria regarded when constructing a theory of personal identity, time is, arguably, the most neglected concept. In this section, we will discuss some of the basic assumptions that all theories of personal identity make with regard to time. We will also examine Lewis' idea that speaking in terms of tensed identity resolves the difficulties of personal identity, such as the difficulties of describing what happens during cases of fusion and fission, because there are factual descriptions we can make about persons at given times during their lives. That is, tensed identity, we shall see, involves the notion that the relations of aggregates of *person-stages* determine our descriptions of a person's identity.

The most basic assumption made about time and its relation to personal identity is that time is a linear phenomenon. This can be evidenced by the fact that none of the standard problem cases of personal identity ever include ideas such as time travel or time-loops. We believe with a great conviction that time flows from past to present to future, and that our everyday experiences about the world lend to this sensibility. Yet, if we were to determine otherwise, how would that affect our claims about personal identity? What would we say if time travel to both the past and the future became possible? Perhaps the most abstract example of personal identity and time travel comes from science-fiction author Robert A. Heinlein's story "*All You Zombies* —". In this story, we are introduced to a person who, through sex change operations and time travel, is able to become his own father and mother, thus giving birth to himself after an intricate series of time loops are established.¹¹

Although Heinlein's example strains comprehensibility with regard to our common-sense view of things, if we grant the details of the story (i.e. the circumstances and the possibility of time travel), it does not seem entirely implausible that such a case might be logically possible. Without the reader having the full details of the story readily available (and for which there is not time nor room enough in this thesis to provide), examining Heinlein's example in full will become quite cumbersome. I point it out here to demonstrate some of the gross peculiarities that can occur when we disregard our presumptions about a linear conception of time. Instead, I propose to examine a similar, yet slightly subdued example of my own creation. In my example, we will simply consider what we might expect from a single leap backwards in time. I choose time travel specifically because it is an

example where we may experience duplication or replication, without the kind of bodily destruction found in some of the other problem cases. That is, in problem cases that examine examples such as teleportation, we see that these examples involve a complete destruction of the body. If we create a situation that involves a time-loop, then, conceivably, we could experience a peculiar paradox, since there would be two identical bodies, without any break in psychological and bodily continuity.

As we have seen, psychological continuity and bodily continuity are both considerations for our criteria of personal identity. We also saw that it was arguable that duplication destroys personal identity, because identity is, strictly speaking, a one-one relation. Suppose that Michael Ellis builds himself a time machine. Being a rather conservative fellow, he decides that his initial test of the device should not be overly extravagant. Let us suppose that he sets his first leap through time for ten minutes into the past. Checking both his watch and the wall clock in his lab, he finds that both read 2:30 pm. He steps into the machine, configures the appropriate settings, and braces himself in his seat.

After a brief jolting of the machine, Michael Ellis opens the door and steps out to find himself still inside his lab. Immediately, he checks the time on his watch against the wall clock. The wall clock reads 2:20pm. The leap was successful. As Mr. Ellis peers around the lab, he spies a man performing advanced calculations on a chalkboard hanging on the back wall. He approaches the man at the chalkboard, taps him on the shoulder, and speaks.

“I thought I might find you here. As you can see, our time machine works,” stated Ellis. “It’s ok, don’t be alarmed. We’re the same person. I just jumped backwards in time by ten minutes,” he assured his counterpart.

“How can we be the same person – we’re in separate points of space?” asked Ellis-2.

“That’s a good question. Considering that my body wasn’t destroyed, as might happen through teleportation, I should have 100 percent bodily continuity with you, yet we both occupy different points in space,” continued Ellis, “and the same should be true for my psychological continuity. At no point has my psychological continuity been broken with yours.”

“But there is a difference – my bodily and psychological continuity will always be ten minutes behind yours. The clock now reads 2:21. Your watch should read 2:31. So, in approximately ten minutes my state of continuity will be identical with yours as it is now,” claimed Ellis-2.

“So, we’re different persons?” asked Ellis.

“Well, how can I be identical with you if I haven’t had all of your current experiences?” asked Ellis-2.

Ellis replied, “You seem to have a good point there. I appear to have all of your experiences, but you lack those that I’ve had for the past eleven minutes.”

“Well, not exactly,” said Ellis-2.

“What do you mean?” asked Ellis.

“You don’t have all of my experiences. By being the first to make the leap backwards, you’re missing the experiences I’m having right now by interacting with you,” Ellis-2 explained.

Ellis was baffled. “But I should have the experiences you’re now having in another ten minutes, after you’ve gone into the time machine and are then standing here speaking these same words I’m speaking to you now. Yet, I think there’s another difficulty,” Ellis stated.

“What’s that?” asked Ellis-2.

“Would you say that we both possess free will?” asked Ellis.

“Of course we do,” confirmed Ellis-2.

“Then couldn’t you opt not to enter the time machine at all? And couldn’t I opt to leave the room, so that I didn’t experience your interactions with me as you do now, if you did decide to leap?” questioned Ellis. “If not,” continued Ellis, “then how can we claim to have free will? Are our actions determined for us?”

“I’m not sure,” said Ellis-2, “If I don’t make the leap, then it seems to create a paradox, in that you shouldn’t be here now talking to me. But if I can’t make the decision not to leap, then, apparently, my actions are already determined for me.”

Both were baffled by the situation. “If you do, in fact, have free will – as we generally suppose of all persons – and you choose not to enter the time machine, would that pose a threat to me? Would I simply vanish, since I should not have jumped, had you not made the decision to do so?”

“It’s hard to know what to say about that,” admitted Ellis-2.

“I agree,” said Ellis. “The problem appears to include that my actions affect your timeline, if we are genuinely the same person. That is, if I refuse to make the leap backwards, then, being identical, you should not have been able to do so,” claimed Ellis-2.

The clock on the wall now read 2:25. “There seems to be two essential questions we’re asking here,” stated Ellis, “the first being that we are either identical persons, or we are instead exactly similar beings. If your actions affect my present timeline, then we must somehow be identical or, at the very least, causally linked in a way that implies identity. To say that we are merely exactly similar implies that, while similar in every respect, we are not causally linked and therefore your actions should have no bearing on my timeline. The second question asks whether we have free will, allowing us to choose the course of our own actions, or whether our actions are determined for us.”

“And we could very well complicate things all the more by both entering the time machine together,” said Ellis-2, “and in doing so wind up with three individual bodies here, after the leap: you and I – the two time travelers – and the unsuspecting gent who will then assume my role in the lab, during this scene. What do you make of this?”

They both looked at the wall clock – 2:27. “I’m not sure,” stated Ellis, “but it looks as if there are a number of ways we could continue to complicate our situation, if we have free will. Continually adding our numbers to our time-traveling group, as you’ve just indicated, is a perfect example. It seems we could infinitely duplicate

ourselves this way, all the while being psychologically and bodily connected to each other, without ever destroying our bodies and with no visible point of fission.”

The clock read 2:29. “What do you think will happen?” asked Ellis-2.

“I don’t know, but I have a feeling we’re about to find out any minute now,” replied Ellis.

Although it is difficult to say whether this example accurately describes what happens when one travels backwards in time, it is nevertheless plausible to examine such a problem case, since the principles of gravity described in Einstein’s theory of general relativity account for the possibility of time travel into the past.¹² My intention is not to provide answers to the questions raised by this example. Rather, my purpose is both to draw attention to the difficulties that arise once we disregard linear timeframes, and also draw attention to the fact that time is the most often neglected consideration in developing our concepts of personal identity. As such, our general assumption that time unfolds in a linear fashion does not allow our standard concepts of personal identity to answer these questions adequately.

Our last topic of consideration, the concept of tensed identity, combines aspects of our use of language and our concept of time. Tensed identity is an idea introduced by Lewis when he attempts to resolve difficulties that arise in the problem cases relating to the fission and fusion of persons. To begin, Lewis introduces two relations dealing with descriptions about persons over time: the R-relation, which regards the mental continuity and connectedness among various person-stages; and the I-relation, concerned with whether or not there will be a person that exists both now and at another time (20-22). The fundamental difference between these relations

is that the R-relation concerns itself with person-stages, whereas the I-relation concerns itself with wholly continuant persons, comprised of individual person-stages. That is to say, Lewis states that, “A continuant person is an aggregate of person-stages, each one I-related to all the rest (and to itself).”(22).

The first peculiarity in Lewis’ argument is that he presents us with two relations (the R-relation and the I-relation), gives us a description of the difference between the two, and then claims that the I-relation *is* the R-relation. If they are indeed the same relation, then we should find no variation among their collective descriptions. Perhaps what Lewis means is that they are essentially two ways of describing the same phenomenon. Using the familiar example of a glass being either half empty or half full, we can understand how it is possible for two descriptions to describe the same phenomenon. In fact, we can even introduce a third description, wherein the glass is described as both half empty *and* half full – the top half being empty and the bottom half being full. Whatever the case may be, the fact remains that Lewis describes two relations and claims that the two are actually the same.

Lewis next claims that the I-relation is not identity, in the strict sense, but ordinarily inherits the formal character of identity (22-23). The I-relation is important for Lewis’ concept of tensed identity, and he draws a distinction between this relation and genuine identity for a specific reason. Tensed identity, says Lewis, is not a kind of identity, because speaking in such ways does not necessitate transitivity. He understands that identity, strictly speaking, is a one-one relation and, as such, must be transitive. The I-relation, by contrast, carries the formal character of identity, but is not, itself, necessarily transitive. Indeed, it is a weaker relation that Lewis describes

as being transitive under normal circumstances, but is intransitive in problem cases where we find overlap of person-stages, such as fission and fusion (25-27).

As we have seen, Lewis believes that continuant persons are aggregates of person-stages. Speaking in terms of tensed identity, then, we should be able to determine whether or not two continuant persons share person-stages. For example, we can say that continuants C_1 and C_2 are identical at time T_1 if and only if they both exist at T_1 and their stages at that time are identical (Lewis, 26). Accordingly, the primary focus of tensed identity is to describe persons-at-times. When we ask a question such as “Is person A the same as person B?” we are asking about the states of person A and person B at a specific instance. By Lewis’ reckoning, if we can determine whether the stages of person A and person B are identical at the time in question, then we can determine whether they are the same person at that time.

Why bother with the concept of tensed identity? To what end does it aim? As Lewis observed, under normal conditions the I-relation is transitive. His notion of tensed identity, then, specifically addresses the problem cases we commonly face when discussing philosophical theories of personal identity. That is, tensed identity gives us a way to describe instances where there may be overlap of person-stages from different continuant persons, as well as overlap of person-stages within the timeline of single continuant persons. Not only do fission and fusion create these problems, but so does longevity (assuming that personalities and memories diminish over extended periods of time, similar to what we observe in ordinary human life spans).

If for no other reason, I think that Lewis' argument is useful in that it again draws our attention to the concept of time, with regard to personal identity. In doing so, it introduces an implication we have not yet seen considered among the theories discussed, namely, that the truth-value of our assertions about personal identity may be time-dependent to some degree. Regarding a definitive theory of personal identity, specifically, I do not think Lewis' position is any more successful than any of the others considered throughout our examination. Even if we speak in terms of tensed identity, we cannot escape the arbitrary nature upon which our theory rests. Lewis' foundational description rests on the ideas of mental continuity and connectedness, as well as the concept of aggregates of person-stages – none of which are concepts we are forced to accept out of logical necessity. In fact, part of the success of Lewis' position relies on pretending, for argument's sake, that some of the open-endedness of the psychological criterion has been settled (20; 30). It seems to me that we can make statements about persons-at-times, but these are nonetheless made within the relative framework we construct about persons, based on contingent criteria of personhood.

Notes

¹ John Perry, "Personal Identity, Memory, and the Problem of Circularity," *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 144.

² Bernard Williams, "The Self and the Future," *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 180-81.

³ See how Wittgenstein explains this example in his *Philosophical Investigations*, § 265.

⁴ Alfred Jules Ayer, *Language, Truth and Logic* (New York: Dover, 1952) 37.

⁵ Bernard Williams, "Personal identity and individuation," *Problems of the Self* (Cambridge: Cambridge UP, 1999) 13; and Sydney Shoemaker, "Personal Identity and Memory," *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 133.

⁶ More will be discussed regarding these issues in the following section. I merely wish to introduce them here as tensions in the thesis of memory criterion.

⁷ John Perry, "The Importance of Being Identical," *The Identities of Persons*, ed. Amelie Oksenberg Rorty (Berkeley: U of California P, 1976) 73.

⁸ Derek Parfit, "Lewis, Perry, and What Matters," *The Identities of Persons*, ed. Amelie Oksenberg Rorty (Berkeley: U of California P, 1976) 98. I use "relations" in this section for consistency with Parfit's usage. Parfit's terminology here appears to refer to the association or interconnection of various kinds of psychological states over time, relative to each other and to themselves, rather than to particular instances of psychological states. That is, each of the individual relations are causally related.

⁹ David Lewis, "Survival and Identity," *The Identities of Persons*, ed. Amelie Oksenberg Rorty (Berkeley: U of California P, 1976) 35.

¹⁰ Derek Parfit, "Personal Identity," *The Philosophy of Mind*, ed. Jonathan Glover (Oxford: Oxford UP, 1986) 148-50.

¹¹ See Heinlein's story for a full account of his example of what might be possible if we could manipulate time in a non-linear fashion.

¹² J. Richard Gott, *Time Travel In Einstein's Universe* (New York: Houghton Mifflin, 2001) 82-3.

CHAPTER 3

THE BODILY CRITERION AND REDUCTIONISM

Persons and Bodies

The issues that create problems for the mind-body distinction and the memory criterion have led to the development of the theory of the bodily criterion of personal identity. This theory seeks to resolve the problems of the above by affirming a reductionist position, reducing any and all references about a person's identity to matters of a body or a particular part of the body, such as the brain. Two of the primary claims supporting the bodily criterion are: 1) the lack of conclusive evidence for the existence of immaterial substances, such as the soul, and 2) the belief that persons appear to go where the brain goes. Let us examine the implications of equating persons and bodies at length by first attempting to discredit the validity of immaterial substances in the role of determining personal identity.

Soul Searching

If we intend to argue for a theory of personal identity that is based upon the concept of an immaterial substance, we must first be able to establish answers for a few specific questions. For example, we need to answer questions such as: 1) What is an immaterial substance or soul? 2) Are persons and souls identical? and 3) Is it

possible to establish personal identity based on the criterion of an immaterial substance? These questions are the immediate focus of our attention.

1) What is an immaterial substance or soul? Returning to Descartes' mind-body distinction, we recall that he claimed he was essentially a soul (i.e., a thinking thing) that could exist apart from his body. This sort of description has become the rule for explaining what we typically refer to as the *soul*. That is, when we discuss ideas about the soul, we are referring to what is described as a disembodied person. What is the usefulness or application of such a concept? According to Anthony Quinton, the soul has been used, traditionally, to explain both the vitality that distinguishes living and non-living things, and the seat of consciousness (53).

Before continuing to examine our three primary questions, I think it is important to point out that describing the soul in terms of a disembodied person in no way answers our original question – what is an immaterial substance or soul? Instead, all that is achieved with such explanation is that we push back our question one step, since this answer merely begs the question: What, then, is a disembodied person? It is arguable that a disembodied person is a self-contradictory concept, especially if we argue that persons simply *are* bodies. This amounts to stating that a person is a disembodied body, which is counterintuitive. However, Quinton believes a disembodied person is a viable concept. If not, he says, then we could not understand concepts such as the classic idea of a ghost (Quinton, 62).

I think Quinton's ghost example is inadequate for several reasons. To begin, it is quite obvious that there is no absurdity in claiming that we have the ability to describe things that do not necessarily exist. That is, just because language allows us

to give descriptions about certain things, this does not necessitate that there is some actually existing thing that fits our description. For example, my descriptions about Lord Wellington – the albino unicorn who lives on the front lawn of the White House, in Iowa – need not describe any actually existing thing. Although this example is slightly exaggerated, the same is true for more mundane descriptions. My descriptions about the oriental area rug in my living room need not describe anything actual, because there may be no such object in my living room. The same argument could be applied to concepts such as immaterial substance, the soul, ghosts, and the like – we can describe them, but this does not mean they exist.¹

Clearly, the argument above can be said to apply to such things as quantum particles, dark matter, super strings, and the like, since we do not directly observe these things themselves. The difference, however, is that these *objects* of scientific investigation are not, strictly speaking, posited as actually existing things. Rather, they are representations or models that allow us to describe the workings of the phenomena we *do* observe. Immaterial substances, by contrast, are not described as being representations or models, but are posited as being actually existing things.

We can also argue that Quinton's ghost example is not a genuine example of what it means to be disembodied. The classic example of a ghost, in many instances, is described as a spiritual or otherworldly human being. In other words, a ghost is described as in every way resembling the form of a human being, only made of different stuff. Given this, we can argue that this *ghost stuff*, although different from the bodies of living beings, is still a form of embodiment. This argument holds even if the ghost does not physically resemble a human being. For all that concerns the

validity of our claim is that this ghost stuff is capable of being classified as another type of body or embodiment. The ghost stuff is no less a body than is that of a human being, only made of different stuff. The kind of disembodiment I have in mind, which we might assert as *genuine* disembodiment, quickly leads us into the realm of abstract ideas, since it involves the concept of a property in-and-of itself, which describes the condition of a property existing *apart from any body*.

Genuine disembodiment, we might say – the sort that implies self-contradiction – is described in terms that allow us to ask questions such as: what would it mean to describe a person’s attributes or character entirely apart from any sort of body? To understand this concept of disembodiment, let us formulate a few questions that exemplify the point. What would it mean to have wit, charm, and arrogance apart from any particular body? Likewise, what would it mean to have memories, thoughts and intentions apart from any particular body? The problem here is similar to the kind we observe regarding the need for properties, such as color and shape, to be expressed through bodies or particular objects. Apart from any particular body, we might argue, these qualities are regarded as abstract ideas, rather than real properties. That is to say, the very concept of disembodiment requires that there be something that exists apart from the body. What are the properties, then, of a disembodied person? According to Williams, there are two possible answers for this question: 1) that there is no answer, or 2) there is an answer, but the value of the answer is always equal to zero. If the first case is true, says Williams, then:

...we shall say that the possibility of disembodiment would show, not just that a person was a sort of thing that *did not necessarily* exemplify physical determinates, but that it was the sort of thing that *necessarily did not* exemplify such

determinables. Then even embodied persons would not have physical attributes, but would be nonphysical things associated with a body, i.e., the Cartesian account would apply.²

Yet, we have seen that Williams rejects the Cartesian account, because of the difficulty involved in verifying claims regarding immaterial substances. The second answer to our question regarding properties of disembodied persons raises a different difficulty, says Williams. That is, if we claim that these properties have a quantifiable value, then we must express them in terms of how they relate to a person's *body*. For example, inquiring as to the weight of Michael Ellis would include asking about the weight of his physical body, as well as the weight of his nonphysical or immaterial substance. Upon weighing him, we might ask, "Of his 195 pounds, what percentage of his weight accounts for Michael Ellis' immaterial substance?" If this immaterial substance suddenly became disembodied as he stood on the scale, would we see a drop in the measured weight? Denying any such change requires that, even while embodied, the immaterial substance is a weightless item associated with a body.

Part of the difficulty of claiming that immaterial substances have no determinable attributes, asserts Williams, is that being able to understand what a given thing is involves having an understanding of its determinables. In other words, if we cannot know something's attributes, then we are groundlessly making assertions when we posit any claims about such things. Without knowledge of the determinables, all we are left with is conjecture. Yet, adding further difficulty to the issue, Williams argues we are forced to accept that a person is a sort of thing that

necessarily displays determinates at some time (when embodied), but does not necessarily do so at all times (when disembodied).

If the weight of an immaterial substance is indiscernible from that of its body, then we really know nothing except what we observe of the body. It may be argued, however that this is not the case. For example, it might be argued that the above claim is unreasonable, as it would be similar to asserting that the ideas of a book disappear when the book disappears. Similarly, the objection might state that the problem of indiscernibility is really a pseudo-problem, since we can explain the problem much like the velocity of a bullet *disappearing* when the bullet stops moving. Both of these objections are unsound, and fail for the same reason, namely that they are not plausible analogies of the original assertion. That is, the book analogy fails because the ideas of the book are clearly discernible from the physical incarnation of the book. We never make the confusion that the ideas in the book are a physical property of the book, nor that the weight of the book is something that exists once the book ceases to exist. Likewise, we do not confuse the velocity of the bullet with one of the bullet's other physical properties, such as the weight and length of the bullet. If its velocity were a property indiscernible from the bullet itself, then we should be able to measure the velocity of the bullet while it is moving and at rest and come up with identical figures – just as we can measure the length of the bullet moving and at rest and come up with identical lengths.

Williams believes the indiscernibility of the properties of immaterial substances from that of their bodies, as described above, demonstrates that those committed to the belief in disembodiment necessarily adopt the Cartesian position –

that bodies and immaterial substances are genuinely separate from each other. As such, the weight of a person has *always* been zero. If this were not so, then we should expect to see an increase in a body's weight, should an immaterial substance suddenly occupy it.

Quinton argues that the dispositions and character of a person can exist apart from any one particular body, so long as they are manifested within a body (59-60). I am not exactly sure what this entails, since this is similar to making the claim that height and weight can exist apart from any particular body, so long as they are manifested within a body.

On the one hand, Quinton's claim sounds very similar to Descartes' idea that the self is both separate from and, at the same time, intermingled with the body. Unless we are going to argue that dispositions and character (along with all the other properties we attribute to persons) are somehow special or different than the other properties of bodies, then I think Quinton's argument bears little fruit. I think that to say properties are not identical with a body and to say that properties can exist apart from a body are not logically identical concepts, nor does one follow logically from the other. Williams implies this as well, when he states that persons should be classified as material bodies that think, rather than a mind that has a body (*Problems of the Self* 70), and, furthermore, that persons and bodies are not identical, but that this does not entail they are two different things (*Problems of the Self* 73-74). So, then, if we intend to argue that the properties of persons are different than the properties of other bodies, in a way that allows them to become genuinely

disembodied from any sort of body, then it is necessary to explain how these properties differ in such a way. Quinton does not offer any explanation to this extent.

Ayer argues against the idea of immaterial substances, although for different reasons than Williams does above. That is, Ayer claims that, “The problems with which philosophers have vexed themselves in the past, concerning the possibility of bridging the gulf between mind and matter in knowledge or in action, are all fictitious problems arising out of the senseless metaphysical conception of mind and matter, or minds and material things, as substances.” (124). Likewise, he states, “But, when one comes to enquire into the nature of this substance, one finds that it is an entirely unobservable entity.”(Ayer, 126) For Ayer, then, we understand that immaterial substances are not a criterion for determining personal identity. What is important to note is that his objection does not rest simply on a denial of the existence of immaterial substances. Rather, what we find is that Ayer protests because he thinks propositions about immaterial substances are inherently unknowable and are, therefore, purely conjecture. This is evident when he claims that, “we shall maintain that no statement which refers to a ‘reality’ transcending the limits of all possible sense-experience can possibly have any literal significance; from which it must follow that the labours of those who have striven to describe such a reality have all been devoted to the production of nonsense.”(Ayer, 34). As we shall see, Perry echoes this idea when he claims that if we cannot observe the soul, then we have no way of knowing whether or not we are the same soul from day to day.

All of these considerations emphasize the difficulty of clearly defining and describing immaterial substances and/or the soul. Even if we concede Quinton’s

ghost example, it is clear that we have done so without conclusive observational evidence. That is, we draw our beliefs about the soul without any direct observation of immaterial substances. Accordingly, the problem becomes one of attempting to describe things that are beyond the scope of observation and experience. What we appear to be left with are abstract ideas or, perhaps, what we can call *content*. By content, I simply mean what is sometimes referred to as the special relations of memory, character, habits, beliefs, intentions, and preferences that constitute a personality. The notion is that so long as this *content* exists, then so does the person. For example, if we take all of the content that makes Michael Ellis the person he is, and successfully transplant it into another body, then, allegedly, Michael Ellis is preserved in this new body. His content, we might say, is not dependent upon any particular body. Yet, as we saw in the previous chapter, Williams points out that transplanted identities of this sort give rise to the problem of duplication, which contradicts the concept of identity, because identity in the strict sense is a one-one relation.

What we must consider is whether or not the introduction of immaterial substances, into the above concept, makes the theory any more plausible. For instance, if we grant the plausibility of transplanting persons from body to body, are we claiming that person contents are being transported by means of immaterial substances? Or, rather, are we claiming that transplanting is simply a physical process, more closely analogous to copying information from one computer to another? If we mirror a computer's hard drive onto another one, then we duplicate the contents, so that the information on each is exactly similar. The computer

analogy is useful because it demonstrates a simple transfer of information from one physical body to another, which no one (apart from, perhaps, a few misguided Star Trek fans) believes was done through any means of immaterial substances. The information on the new hard drive merely resembles (exactly) the information found on the original computer.

What may be objectionable here, however, is that computers and human beings are too different to make a useful analogy. Computers are neither conscious, nor self-conscious things, unlike human beings. This objection harkens back to Quinton's observation that the soul has been used to explain the difference between living and non-living things. If, then, we are arguing that it is the soul that transfers person content from one body to another, then we must be able to answer specific questions to support our claim. For example, what means will we use to demonstrate that the soul transfers person content from body to body? If we determine that this process is beyond demonstration, then how are we to verify our claim? Will introspection reveal the nature of the soul, as some, including Descartes, have argued?

Both Ayer and Dennett argue against the reliability of introspection. Ayer states that a substantive ego, of the Cartesian sort, is not revealed through self-consciousness, and if it is not revealed there, then it is not revealed anywhere (126). He concludes that the existence of such entities is completely unverifiable. Similarly, Dennett argues that we are fooling ourselves with our belief that introspection is either infallible or incorrigible; furthermore, we are wrong to believe that introspection is merely "looking and seeing," when what we are really doing is

theorizing (*Consciousness Explained* 67). Questions of the kind we are addressing here are what bring us to examine the next of our primary questions regarding personal identity based upon the criterion of immaterial substances.

2) Are persons and souls identical? When we pose this question, what we are speculating is whether or not a person is simply identical to a soul, or if persons are things that *have* souls. If a person simply *is* a soul, then, by the transitivity of identity, anything we find true about one is also true of the other. If, however, persons are things that *have* souls, then there must be some way to distinguish one from the other. For the purpose of our examination, let us adopt the Cartesian concept of the soul. Assuming persons and souls are identical we will see how this impacts our theory of personal identity. To begin, the most serious criticism is that if we cannot find a way to determine the identity of a soul, then, necessarily, we cannot determine personal identity. When Michael Ellis claims that he is the same person he was ten years ago, according to this position, he is making no reference to his physical body. His statement is true no matter what body his soul presently occupies, since the soul and the person are the same thing. Yet, we might argue that not all references to persons exclude a reference to bodies.

In some instances, we use language in ways that refer to persons specifically by means of their physical bodies. For example, if I say, "He only has twenty dollars on his person," I am not claiming that he has twenty dollars on his soul. There are several conclusions we may draw from this example. We could claim that, in some cases at the very least, when we refer to persons we are referring to physical bodies; or we could claim that this is simply an example of misused language, since persons

are not equal to bodies. Does including the soul into our theory of personal identity enrich our understanding of the concept of personal identity?

Arguably, the apparent benefit we gain from equating persons and souls is that we are given a definitive way to establish personal identity, when there are questionable instances of bodily identity (such as a body swap or reincarnation). If two men both claim to be reincarnations of Thomas Jefferson, then relying on the soul, we might say, will allow us to determine the truth of each man's personal identity. It may turn out that neither man is a reincarnation of Thomas Jefferson; yet, if one is in fact a genuine reincarnation, then he will turn out to be the same soul that inhabited the body of Thomas Jefferson in a previous state. But how can we really make such determinations when the very thing this position rests upon (i.e., the soul) is beyond all means of perception? Furthermore, the problem we face here, regarding the unobservable nature of the soul, is applicable whether or not we equate persons and souls – so long as the soul remains an aspect of our explanation of personal identity. That is, the fact that the soul is unobservable causes problems if we claim that persons and souls are identical or, likewise, if we claim that persons *have* souls. We cannot determine if persons and souls are identical, nor can we establish a means to distinguish the two from each other.

We can, however, speculate what it would mean were we able to determine persons are things that *have* souls, based on some of the fundamental ideas of identity. If persons are things that have souls, then the two are not identical.³ If our theory also maintains that persons are not identical to bodies, then we conclude that persons are neither souls, nor bodies. In this case, persons appear to be something

other than material and immaterial substances. What other options are left open to us? We are left with arguing that personal identity consists in a special relationship (such as person content or memory), that there is no such thing as personal identity, or that a person's identity consists in some, as of yet, undetermined criterion. Let us now turn our focus away from attempting to define immaterial substance, in order to examine the viability of a theory based on such substances.

3) Is it possible to establish personal identity based on the criterion of an immaterial substance? In asking this question, we attempt to understand the circumstances, if any, under which we can tell whether or not a person is the same from moment to moment, based entirely upon the reliance of immaterial substance. The first difficulty to consider, which we have already briefly examined, regards the problem of determining genuine identity from cases of exact similitude.

Placing immaterial substances at the base of our argument poses this problem: How are we to determine personal identity if we cannot, in any practical or conceivable way, observe immaterial substances? In fact, as we saw in the previous chapter, Williams argues that the bodily criterion is the only way to distinguish identity from exact similitude. This may seem a strong claim, but if we recall what we learned of Williams' position about cases of duplication, we remember that he claimed that such instances are counterintuitive to the concept of identity. Identity, he states, is *only* a one-one relation, never a one-many or many-many relation (*Problems of the Self* 15). Since duplication creates a one-many relation, we no longer have a case of genuine identity.

By employing the bodily criterion, we can resolve certain issues regarding personal identity that we could not have by relying on immaterial substances. For instance, if we examine bodily criterion in cases of duplication, we have a means of determining which person existed prior to the duplication and which is the duplicate person. Person A's body at time T_1 has a spatio-temporal continuity different from Person B's body at the same time. Yet, after the duplication of A to B, at time T_2 both will claim to be Person A – each having all the appropriate memories and beliefs of Person A. Indeed, to this point, Williams states, “The only case in which identity and exact similarity could be distinguished, as we have just seen, is that of the body....”(Problems of the Self 10)

Both Perry and Quinton further this idea to some extent, each in their own way claiming that immaterial substances cannot be a criterion of determining one's personal identity. Perry claims that, “If identity consisted in knowledge of the soul, then all of our beliefs about personal identity would be groundless and mysterious.”⁴ He draws this conclusion based on the premise that, if we cannot observe the soul, then we cannot observe it to be the same (*Dialogue* 17). Quinton's approach states that if the soul is a permanent and unfaltering part of our consciousness, then, being unobservable, it must be useless for purposes of identification (54). This is, in fact, why he believes Hume failed in his efforts to observe the single impression of the self. That is to say, Quinton argues that we can only observe that which changes; the fact that the soul is a constant, unchanging aspect of our existence is what makes it unobservable. For without variation we have no frame of reference. Accordingly,

Hume's mistake was to deny the existence of the self, when the very fact that an ever-present impression lends us no way to observe it (Quinton, 55).

So, again, we are brought back to the difficulty of verifying claims about an unobservable immaterial substance. This leads us to conclude that either personal identity is unverifiable (if we maintain that it rests upon immaterial substance) or that bodily identity is at least *a* criterion of determining personal identity. If we can demonstrate that the latter is false, then perhaps we must admit that there is no such thing as personal identity.

Bodies, Consciousness and Reduction

Considering the difficulties of establishing personal identity based on immaterial substances, it seems only fitting to bring the same level of scrutiny into our examination of the bodily criterion. In doing so, we draw ourselves closer to an understanding of what is necessary for establishing a definitive theory of personal identity, or to the realization that there is really no such thing as personal identity, in the strict, philosophical sense of the term. Can we establish personal identity based on bodily criterion? It would seem that Williams sufficiently answered this question when he demonstrated that the use of bodily identity allows us to distinguish between identity and exact similitude. All we need to know to determine personal identity, we might say, are the facts regarding one's body at a given time. What this implies is that persons and bodies are identical. Yet, as Williams explains, it is objectionable that persons simply are bodies, although he is doubtful that this premise is demonstrable (*Problems of the Self* 74). The objection to persons being identical to bodies asserts that notions such as, "Michael Ellis" and "Michael Ellis' body" are not

logically the same. Williams thinks, as do I, that examples such as this do not entail the necessity of immaterial substances; rather, they are examples that reflect concerns about the ways we talk about persons and bodies. Hence, he states that, “it is certainly not exactly the same thing to love a person and to love his or her body. But this does not show...that persons and bodies are two different things.”(*Problems of the Self* 74).

The thesis of the bodily criterion is that bodies (and not immaterial substances) are the subjects of psychological attributes. This view is described as reductive, since it explains psychological states by reducing them to material states of the brain. Williams argues that the concerns regarding language do not threaten this assertion. For example, in stating, “Gus Shultz loves his dog, Bingo,” rather than, “Gus Shultz’s body loves his dog, Bingo,” we have not conclusively demonstrated that persons are anything other than bodies. Instead, what we have shown is only that our conventions of speaking are such that we find it awkward to phrase expressions in the latter sense. To this extent, I agree with Williams, although, as Quinton attests, even if we grant that immaterial substances do not exist, this does not exclude the possibility of a “unitary nonbodily aspect of a person.”(57). However, if we remove the possibility of immaterial substances, what might it mean to say that persons are a unitary nonbodily aspect of bodies?

A typical response to questions regarding a unitary nonbodily aspect of persons concerns matters of experience. That is, the point might be made that the single subject of experience is what constitutes the unifying nonbodily aspect of persons. The person at time T_1 is the same person at time T_2 , because of the shared

subject of experience. We might also refer to this as the same unity of consciousness. As such, we say that the same consciousness that experienced the events of my fifth birthday party is the same consciousness that experiences these words, as I sit here writing them. I am that thing that shares both experiences – the same subject of experience.

What concerns us here is determining whether or not we are justified in believing in a single subject of experience. Are we justified in believing that our consciousness unifies our experiences? Let us refer to this unified consciousness as our *self*. Is there a single, permanent self that is present throughout each of our experiences, or are we merely something akin to Hume's bundle of impressions? From a phenomenal perspective, the answer seems blatantly obvious – we are unified, cohesive selves that experience and perceive collectively, not as a disjointed bundle of impressions that present the illusion of unity. Our main reason for accepting this notion is our own first-person perspective, through which we understand our separation from the experiences of others. Thomas Nagel emphatically rejects the idea of the unity of consciousness, as does Dennett. For instance, Dennett states, "There is no single point in the brain where all information funnels in, and this fact has some far from obvious – indeed, quite counterintuitive – consequences." (*Consciousness Explained* 102-03). Nagel's beliefs are equally critical in examining the unity of consciousness, although he is not as optimistic as Dennett about a resolution to the problem. In fact, Nagel claims that it may be "impossible for us to abandon certain ways of conceiving and representing ourselves, no matter how little

support they get from scientific research. This, I suspect, is true of the idea of the unity of a person....”⁵

The evidence used to support Nagel’s argument comes from cases of split-brain patients. In these cases, he states that subjects display behavior that indicates the existence of at least two streams of consciousness. However, it is arguable that in these split-brain cases both streams are not genuinely conscious, since the right hemisphere’s responses often appear like that of an automaton, rather than conscious mental processes. Nagel rejects this description of the split-brain phenomena; he believes the actions are “too elaborate, too intentionally directed and too psychologically intelligible to be regarded merely as a collection of unconscious automatic responses.”(*Personal Identity* 235).

Some of our investigations into the unity of consciousness begin from rather suppositional premises. For example, if we begin by asking “Can the self divide?” or “Can consciousness split?” then we start from a biased attitude. Both of these questions presuppose unity from the outset. What we should ask, rather, is whether or not consciousness is unified and, furthermore, how do we go about determining such questions.

Someone championing the unity of consciousness might respond to Nagel by claiming that, even if there are separate streams of consciousness, these streams are part of the same subject of experience, and are, therefore, ultimately unified. But in order to concede this point it seems we must accept a Cartesian view of the subject of experience. That is to say, our conception of a unified subject of experience becomes one of a separately existing thing, if we insist that our consciousness is disjointed on

one level, yet ultimately unified on another. The only way I can conceive of avoiding a move towards a separately existing subject is if we were claiming that, while aspects of our consciousness are disjointed, being a part of the same life nonetheless unifies it. I would hardly say, however, that this is sufficient to demonstrate a single subject of experience, since I can easily imagine the possibility of a disjointed consciousness that is part of a shared life, which involves no single subject of experience. This is, in fact, precisely the kind of idea Hume's bundle theory employs – a series of overlapping sensations, belonging to the same continued life, without a single subject of experience or self. In other words, "same continued life" by no means entails the need for a single subject of experience.

I agree with Nagel in that I believe our reasons for rejecting the idea of disjointed persons is a result of the habitual ways in which we conceive of ourselves, which is a direct result of our phenomenological experience. I see no absurdity, however, in the concept of disjointed persons, nor any necessity in the truth of a single subject of experience. One of the reasons to reject a disjointed view of persons is the belief that consciousness is all-or-nothing. Dennett argues that consciousness is a phenomenon of degrees, not something that separates the universe into two categories – conscious and unconscious (*Consciousness Explained* 447). He believes that our false assumption about this all-or-nothing view of consciousness is a result of essentialism – the belief that there must be sharp, definite dividing lines for the explanation or natures of things. We believe that every question has a definite answer (*Consciousness Explained* 420-21). I think we can categorize this kind of belief as a type of logical determinism, in that it uses the same type of either/or logic we find in

cases involving logical determinism. For example, an adoption of the all-or-nothing view of persons allows us to argue that either persons are conscious or they are not, or, similarly, that something is either a person or it is not.

At first glance, statements of this kind seem to carry a bit of force. However, if we take a closer look, we find they are not as telling as they appear to be. As Roy Weatherford states, “Logical determinism is, we conclude, like so many philosophical doctrines, either true but trivial or significant but false, depending on how it is construed.”⁶ What this entails is that, if true, we essentially gain nothing from statements of this sort. Conceding these kinds of deterministic statements has trivial results, since they are epistemologically vacuous. For example, let us suppose we are posed with the claim, “A person either has a soul or he does not.” While this may very well be true, it is a trivial claim in that it brings us no closer to the truth of knowing whether or not persons have souls. We have gained no knowledge about the real world from the assertion. Not only is this the case, but we are also presented with false alternatives, since there is no absurdity in claiming that personhood is a matter of degrees or that questions about our identity do not necessarily carry definitive answers. Parfit states that, “Only if we are separately existing entities can it be true that our identity must be determinate.” (*Reasons and Persons* 216).

We see that the all-or-nothing view brings us back again to a kind of Cartesian approach to personal identity. In order to work around this kind of thinking, Dennett argues that we must neutralize the illusion of what he calls the Central Meander – the idea of a single subject of experience, or Boss, which is in charge of running the system (*Consciousness Explained* 228). This concept is not unlike the reductionism

we examined previously in Searle's arguments. Searle discussed the reduction of the mind to physical, micro-level functions of the brain. Likewise, Dennett argues, "As usual, the way to discharge an intelligence that is too big for our theory is to replace it with an ultimately mechanical fabric of semi-independent semi-intelligences acting in concert." (*Consciousness Explained* 257). While someone supporting the Cartesian dualist position would reject such a notion, we can certainly see similar types of phenomena when we examine the way computers function. That is, binary informational systems provide a fantastic example of how simple micro-level states can produce various and often unexpectedly remarkable results. In binary, we begin with a system that works on the combinations of on/off states – and this is the whole of the system's logic. By merely combining various states of on/off logic, we are able to produce colors, sounds, images and video, communicate with others at nearly any given point on the planet, and much more. Arguably, none of this demonstrates that computers act intelligently. However, what is important to consider is that computers can yield all of these results, as well as execute powerful computations faster and more accurately than even the brightest human beings, yet no one ever argues that there must be some immaterial substance or unified consciousness inside controlling the system.

Another point that one might argue against this analogy is that computers do, in fact, have a sort of mechanical Central Meaner, since all of the information is processed by the Central Processing Unit (CPU). Yet, this is a bit of a mistake. While most of a computer's workload is processed by the CPU, many types of computer hardware contain instructions that allow it to bypass the CPU altogether to

access memory – a process called Direct Memory Access (DMA). Furthermore, computer networks can be setup to process information in tandem, so that many processors work on smaller parts of the information, and processing is decentralized.

So, what does all of this have to do with personal identity? It is meant to demonstrate the kinds of phenomena that are possible purely as a result of simple micro-level processes, like those described by Searle and Dennett. We are not implying that the analogy refutes the possibility of the existence of immaterial substances or single subjects of experience. Rather, what we intend to demonstrate from the analogy is that, if we can generate such remarkable phenomena and results by mechanical micro-level information systems, then this at least lends the *possibility* that the same is true for biological micro-level processes. Apart from having a personal bias or reluctance to such a notion, we can just as easily conceive of ourselves having personal identity without relying on immaterial substances or single subjects of experience. This reluctance does not negate the plausibility of the alternative explanation regarding capacities of micro-level biological systems producing remarkable macro-level results. As Dennett notes, there is a difference between finding something hard to believe and being reluctant to believe something is possible (*Consciousness Explained* 432-33). I believe, as Nagel implied, that our reluctance to accept such a view of ourselves says more about the nature of our conceptual limitations than it does about the logical shortcomings of the theory.

While Dennett and Searle agree that it is possible to describe mental phenomena in terms of reductionism (i.e. without regard to immaterial substances or single subjects of experience), they diverge about their views of consciousness. For

example, Dennett's argument champions the idea that consciousness is not some sort of special phenomenon limited to living things, but is defined, rather, in more behaviorist terms. As such, he claims that, "a suitably 'programmed' robot, with a silicon-based computer brain, would be conscious, would have a self." (*Consciousness Explained* 430). Ayer concurs with Dennett, and concludes his own argument by asserting, "For when I assert an object is conscious I am asserting no more than that it would, in response to any conceivable test, exhibit the empirical manifestations of consciousness." (130).

Searle emphatically denies any such notion that machines are able (or ever will be able) to produce genuine consciousness or *minds*. The difference between, say, robots and human beings is that even robots performing seemingly intelligent acts cannot understand the meaning behind their actions, he argues (Searle, 35). He believes the problem is a difference between duplication and simulation. That is, computers and machines can simulate the kinds of mental phenomena of human beings, but they can never duplicate it, since "no simulation by itself constitutes duplication." (Searle, 37).

Searle's argument poses one of the greatest difficulties regarding reductionism and our descriptions of mental phenomena. Suppose a life-long friend claimed that robots are capable of genuine consciousness, thought, intentions, and all of the other phenomena typically associated with human beings and living things. Upon hearing this claim, suppose that we argued, like Searle, that the idea was absurd since machines can only simulate such phenomena. After our rejection of the claim, our friend opens his head to reveal that he has a highly sophisticated computer brain. He

is, and always has been, a robot. There seem to be two positions we can adopt about this revelation: 1) We can claim that we were obviously wrong, since our friend is a robot with consciousness and all other properties associated with the mental life of human beings or 2) we can claim that we are correct in believing that robots cannot duplicate consciousness, and we have merely been tricked into believing that our friend was conscious all these years.

How are we to decide which description is correct? Remember that Searle's argument states that computers can never duplicate genuine consciousness. Therefore, if we adopt his position, then our response will resemble the second remark above. Yet, Dennett and Ayer would claim that the first description is true, since all we need to do is examine the behavior of the friend to determine whether or not he is genuinely conscious.

The kind of reductive dilemma we are faced with here is one that Williams addresses as well. Considering such cases, one question we must answer, he states, is: What are material properties? He argues that, "If they are just whatever properties material bodies have, then it painlessly follows from the thesis that psychological properties are included among material ones. If it is just defined to exclude psychological predicates, it patently begs the question." (*Problems of the Self* 74). What Williams' insight shows us is that a definitional rejection of the idea that machines are capable of genuine mental activity excludes machines from the class of conscious things by default. That is, Searle believes it is a definitional truth that computers cannot be conscious, since computer programs operate only syntactically, and minds are more than syntactical (31). Based on Searle's description, however,

we can also make the claim that much of the animal kingdom is not genuinely conscious. Conceivably, animals can react to and interact with their environment instinctually – without any real understanding of the meaning behind their actions.

The only exception that Searle takes with his position regards machines that would actually function like human beings. Thus, he states, “If you could build a machine that had the same structure as a human being, then presumably that machine would be able to think. Indeed, it would be a surrogate human being.”(Searle, 35-36). Furthermore, he argues that modern computers and similar calculating devices do not go through the same process as a human brain, even if all of the steps are formally the same, simply because such machines have no mental phenomena (Searle, 48). So, we see that Searle believes that even with identical formal processes, it is conceivably possible that there is still something inherently different about the ways in which machines and brains operate, based on the absence or presence of mental phenomena.

Searle’s attitude toward machine consciousness is not uncommon. It is a popular belief that machines simply cannot have the kinds of experiences that living organisms experience. This attitude is sometimes described as having a bias towards *wetware*, i.e. organic matter, as opposed to hardware, when dealing with the authenticity of consciousness and mental states. Disregarding examples such as robots, thinking computers, and the like, are there any existing technologies that may serve to revise our hardware prejudices?

One example of a possible shift between man and machine, which never fails to fascinate me, is the development of sensitive prosthetics. Artificial limbs that can register sensations of temperature and pressure lend evidence to the claim that

wetware may not carry the special privileged position we generally assume. If plastics and wires can achieve the same result as the organics of muscles, flesh, and nerve endings, then perhaps we are wrong to so strongly assert our biases about wetware.

While prosthetics of this sort are remarkable, clearly they are not sophisticated enough to produce consciousness. It is quite obvious that we can argue that the success of such prosthetics should not surprise us, because the brain moderates them, which is organic. What would we say if we were able to replace the entire body, save for an organic brain? Is an artificial human body, controlled by an organic brain, conscious? What about the inverse – an organic body controlled by a synthetic brain? According to Searle's argument, the first description clearly constitutes conscious, because an organic brain controls the artificial body; the second description would produce consciousness so long as the synthetic brain generates genuine mental phenomena.

Further complicating the reduction of consciousness to physical states or processes, Nagel argues that any such reductive theory must also account for the phenomenological features of experience.⁷ In other words, a genuine reductive account of consciousness must be able to explain our subjective experiences, not merely the mental processes required to produce consciousness. This would require providing an objective account of experiences, which Nagel claims is impossible. We cannot give an objective account of experience, he states, since "any shift to greater objectivity...does not take us nearer to the real nature of the phenomenon: it takes us farther away from it."(*Mortal Questions* 174).

Dennett disagrees with Nagel's assessment of the problem of providing a theory of consciousness. Dennett claims that what is necessary is that we concern ourselves with what we can know about a creature's consciousness, not whether or not we can turn our minds into the minds of these creatures on a phenomenal level (*Consciousness Explained* 442). The structure of an animal's consciousness, he states, is just as accessible as the structure of its digestive system (*Consciousness Explained* 447). The analogy of the digestive system to the stomach is the same concept we saw Searle assert in his attempt to dissolve the mind-body problem.

By Nagel's reckoning, for something to be conscious there is something it is like *to be* that thing (*Mortal Questions* 166). I believe this is similar to what Searle means by his rejection of thinking computers, namely that there are no mental phenomena and therefore it follows that there is no meaning to talk about what it is like to be a computer. And it is this phenomenal sense of being in other creatures that Nagel argues is beyond our intellectual grasp. A consequence of Nagel's argument, however, is that we are therefore locked out of an understanding of the minds of other human beings as well. Nagel recognizes this problem. His response is that the more similar we are to another creature, the more likely we are able to adopt its point of view or the quality of its experiences (*Mortal Questions* 172). So, for example, it is easier for us to understand what it would be like to be Michael Ellis than it is for us to understand what it is like to be an earthworm, since we are more similar in kind to Michael Ellis. Nevertheless, it is the inaccessibility of the subjective experience of others that Nagel claims disallows any reductive theory of consciousness.

Clearly, this problem regarding the reduction of consciousness (and, for that matter, all mental states) to physical states of the brain is a difficult one. On the one hand, Dennett's and Ayer's approach seems like a bit of an over-simplification for something as complex as consciousness – especially when we consider Searle's remarks about thinking machines and computers. On the other hand, Nagel's argument appears so stringent that we can never truly be certain about anything's consciousness, or subjective experiences, other than our own.

What is important for us to ask at this point is whether or not we can establish personal identity using the bodily criterion, even if we cannot conclusively show that consciousness and mental phenomena reduce to brain states. Or, rather, is any such theory shattered without this verification? At the beginning of this chapter, we noted that one of the primary considerations in favor of the bodily criterion is that personal identity appears to go where the brain goes. Can we then build a theory around this fact, regardless of what is said about the reducibility of consciousness to physical brain states? Let us now examine the validity of the claim that personal identity follows the brain.

Since the brain is so closely related to all of the standard mental phenomena we associate with personal identity (e.g., character, intention, memory, etc.), perhaps it is best that we phrase our question as follows: Under what circumstances would we say that personal identity does not follow the brain? If we disallow the possibility of immaterial substances, it is difficult to imagine a case where these mental phenomena would not follow the brain in a brain swapping experiment. For example, if we swapped the brains of Person A and Person B, we would be quite surprised, after the

fact, to find Person A's body still owning the habits, character, intentions, and memories associated with the brain now inside Person B's head. However, this does not necessitate that persons are merely bodies, but instead shows only that personal identity appears to follow the brain. We can just as easily imagine that persons are merely a relationship of contents, as described earlier, and that this relationship follows the brain, but would hold when attached to any brain, not simply the brain of a specific individual. As such, this entails that persons would essentially be a set of information.

The problem with describing persons solely as content or information, as Williams noted, is that it gives us no way to determine genuine identity from exact similitude. Additionally, it allows for the possibility of duplication, which is contrary to the idea of identity. If every instance of a brain swap produced results where the person's mental phenomena (i.e., his psychology) followed the brain, then we have at the very least some minimal reason for supposing that personal identity is in some way associated with the brain. Arguably, we could say that brains are types of bodies, and since personal identity follows the brain, it follows that personal identity relies on the bodily criterion. This amounts to nothing more than saying – same person equals same brain. And this notion appears to hold true even if consciousness and any other mental phenomena are found to be irreducible to physical states of the brain. That is, if mental states are some sort of metaphysical phenomena, we can still claim that personal identity follows the brain, so long as these irreducible mental states follow the brain. Michael Ellis would still be the same person, it seems, if his

brain were swapped, regardless of whether or not we discovered that mental phenomena reduce to physical states of the brain.

Under normal circumstances, then, it appears we have a strong argument for believing that personal identity follows the brain. This of course is true only if we maintain the implausibility of theories based upon immaterial substances. When we begin to consider the previously mentioned problem cases (i.e., duplication and replication), however, our claims about personal identity become less certain. And our theory must consider such problem cases, if we are to develop a working concept of personal identity. It is to these abstract considerations that we turn our attention to in the next chapter. In doing so, I believe we will uncover a pivotal understanding of why these theories have failed to provide the kind of conclusive or definitive account of personal identity that we seek.

Notes

¹ Typically, the adoption of a belief in the existence of immaterial substances, such as the *soul*, arises from religious traditions. However, the concept of *soul* as a *separately existing*, immaterial substance need not carry the ghost-like connotation that Quinton suggests. For example, in post biblical Hebrew the most commonly translated words for “soul” are *nefesh* (meaning “breath”) and *neshamah* (meaning “to breathe”) – neither of which imply the sort of distinction of body and soul that is found in Greek and early-Modern thought. Rather the post biblical concept refers to the vital or animating source of life.

² Bernard Williams, “Are persons bodies?” *Problems of the Self* (Cambridge: Cambridge UP, 1999) 71.

³ Here I am referring to *soul* in the Cartesian sense of immaterial substances (i.e. a separately existing, thinking thing), not in the sort of linguistic sense where we sometimes equate *soul* and *self* as merely a descriptive means of self-representation – not simply our first-person perspective but the immaterial *thing* that is the subject of this perspective.

⁴ John Perry, *A Dialogue On Personal Identity And Immortality* (Indianapolis: Hackett, 1978) 16.

⁵ Thomas Nagel, “Brain Bisection and the Unity of Consciousness,” *Personal Identity*, ed. John Perry (Berkeley: U of California P, 1975) 228.

⁶ Roy Weatherford, *The Implications of Determinism* (New York: Routledge, 1991) 175.

⁷ Thomas Nagel, “What is it like to be a bat?” *Mortal Questions* (Cambridge: Cambridge UP, 1979) 167.

CHAPTER 4

THE MEANING OF IT ALL

Drawing Conclusions

Having examined the primary theories and concepts related to the philosophical discourse about personal identity, it is now time to see what we can discover about our inquiry. Why is it that all of the theories generated difficulties, disallowing the sort of definitive resolution that would put the issue to rest? Is there, perhaps, some inherent commonality, which explains the lack of conclusive, ironclad certainty we strive to establish in our definitive theory of personal identity? This is our task in this chapter – to examine the nature of these theories themselves.

Specifically, our analysis throughout this chapter will consider the various ways language influences our ideas about personal identity. We will also examine meaning, contingency, and arbitrary decision, as they apply to the development of a definitive theory of personal identity. In doing so, we will not pose our own theory of personal identity, but will instead produce a meta-theory of personal identity. In short, we are creating a theory about theories of personal identity.

Contingency and Arbitrary Decision

The very beginning of our examination of personal identity concerned Descartes' notion of mind-body separation, and the problems associated with material

and immaterial substances as they related to personal identity. We will now revisit Descartes to discuss a single concept he mentions yet never fully develops – one which I believe is pivotal for understanding the difficulties about theories of personal identity from his day through our contemporary discussions. In Descartes, we find a sort of admission about a difficulty that exists regarding conceptualizations of a definitive theory of personal identity. That is, in the *Meditations*, Descartes recognizes that his beliefs about those other than himself are judgments of which he could very well be mistaken (21).

While Descartes was not concerned with judgment as it applies to personal identity, I draw attention to his statement because it explains a great deal, I think, about a kind of ambiguity or arbitrary nature that is admitted to some degree by many of the major contemporary contributors to discussions about personal identity. A similar notion of judgment is also present in Hume's treatment of personal identity. As we recall, Hume claimed that the problems typically associated with personal identity are not genuine philosophical problems; rather, they are grammatical difficulties (262). And for these difficulties, he states, we have no just standard for deciding upon the relations attributable to identity. As we shall see, similar sentiments are prominent in contemporary thought, and there is no shying away from admitting that our decisions regarding personal identity are made arbitrarily. Parfit describes questions that generate arbitrary responses as empty questions – those whose answers are neither true nor false (*Reasons and Persons* 213). Accordingly, he believes that some questions about identity simply have no real solution.

If we examine a situation involving the duplication of a person's content from one brain to another (in which the original person does not survive), then we are left asking: Is this duplicate (now the sole survivor) the same person as the original? Similarly, if we revisit our example of the life-long friend found to be a robot, we ask: Is he genuinely conscious or only simulating consciousness? It is these sorts of questions that Parfit believes are empty questions. However, being empty questions by no means implies that they are meaningless. Rather, what this implies, as Parfit states, is that, "There is no question of either of these decisions being a mistake." (*The Philosophy of Mind* 160). The implication, then, is that the answer to these questions is not a matter of fact, but is instead expressing matters of contingency.

Stating that assertions about personal identity are contingent means that they do not follow from certain premises by logical necessity. There is no absurdity or self-contradiction in adopting one framework of personal identity as opposed to another. The propositions that form our account of personal identity could have been otherwise, if we had adopted a different conception of personhood upon which to formulate our framework for our theory. What we begin to understand is that the failures or shortcomings of the theories of personal identity result from our initial approach to the problem. That is to say, our theories of personal identity have failed because we approach the issues as if they are factual matters. Our accounts of personal identity really express notions that are relative to frameworks of understanding that are not matters of fact. This may sound difficult to accept, but I think more explanation will reveal why this is the case about theories of personal identity. In making such a rather peculiar claim, one of the first objections will be, no

doubt, that statements about persons are factual statements, since such questions involve matters we can verify (or conceivably verify) at a particular time. Let us start to clarify my claim by looking at the following question: Are statements about persons factual statements? The most genuine answer we can provide is, I think, yes – so long as we are willing to accept that the facts are relative to a particular framework. In stating as much, what I am arguing is that the *facts*, inasmuch as they are facts, are relative to our concept of persons and the frameworks of understanding we build around this concept.

We must think of our statements about personal identity as *relative facts* – propositions that are either true or false based on our concept of a person, which itself is *not* a factual matter. This is why I claim that we can certainly have *a* theory of personal identity, but never *the* theory of personal identity. The truth-value of our propositions about personal identity hinges on our foundational assumptions about persons, sameness, and similar ideas. But these assumptions are also, in a sense, the wild cards in our theories. This notion of contingency with regard to our formulations of personal identity is implied in Dennett’s argument, when he admits that we have not discovered the “clearly formulatable necessary and sufficient conditions for ascription” and furthermore “there may be none to discover.”¹ If there are no genuine facts to discover (i.e., no necessary conditions), then the implication is that any truth-value these propositions may have can only be derived from the statements relative to our foundational descriptions. If we reject the foundational descriptions, then the basis of our theory collapses. And we have no reason to adopt

one description over another, apart from our own assumptions concerning which description we find most practical.

This sort of assumptive nature to our approach of personal identity is akin to the kind of approach seen regarding responses to moral decisions. For example, consider the question – When does life begin? Furthermore, does an embryo have the same rights as a fully developed human being? The logical tone of these questions is not so far removed from the logical tone of questions regarding personal identity. When does a baby become a person? Am I the same person now that I was twenty years ago? The assertions we make when answering such questions depend on our mind-sets or outlooks. Our assumptions about the concepts of life, in the first example, determine what kind of assertions we are willing to accept as valid propositions. The same is true about our approach to personal identity. Our assumptions about personhood (and the like) lay the foundation for the kinds of propositions we will accept or reject about problems associated with personal identity.

Regarding the beginning of life, there are those who will insist that life begins at the moment of conception, while others will claim that this merely constitutes the potential for life. These are the sorts of examples that Parfit described in which our answers are neither true nor false – we cannot be mistaken in either case. How is this possible? For those who affirm that human life begins at conception, being conceived is equal to ensoulment, so to speak. That is, we could argue that the *soul*, or *self*, or *person* (whatever that thing is that we are asserting exists) fully exists at the point of conception – it is *not*, we would argue, something that develops over time. This

assertion of what happens at conception is simply an assumption that results from our foundational description. What we must understand, however, is that our foundational descriptions are not logical necessities – they are not self-evident, so to speak. Simply put, we do not always agree about our foundational descriptions. What counts as a person to some sounds entirely absurd to others, and there is no definitive way of reconciling this point through logic or reason. However, we can examine each theory relative to its own framework. If we understand one's foundational descriptions of persons (e.g., persons are equal to bodies), then we can examine a theory of personal identity from that starting point. The *facts*, however, are relative to the framework, not overarching and applicable to *every* theory of personal identity, but only to those whose foundational descriptions of persons are the same.

One of the reasons we may be so eager to reject this kind of conclusion about arbitrariness and instead claim that personal identity is always determinable, is because our phenomenological experience seems to imply that the indeterminacy of one's identity is absurd. Our first-person perspective makes us very much inclined to argue that there could never be a time when we could not have a definitive answer to questions such as, "Am I the same person I was yesterday?" Only when we consider the identity of others are we less likely to insist that identity is always determinable. To this extent, Perry states, "This means I could conceivably be presented with facts that could only be interpreted as neither a clear-cut case of my own death, nor a clear-cut case of my own survival."² Furthermore, he argues that, "If indeterminate cases become common, linguistic decisions will have to be made." (*Identities* 73). Parfit

likens the arbitrary cut-off point we make, when deciding what to say about personal identity, to that of the identity of heaps. For example, when we have a heap of something (perhaps, sand), we can begin reducing the heap by its individual units. What Parfit argues is that there comes a point when it is not so clear-cut when we no longer have a heap, for it all depends on how we employ the terms (*Reasons and Persons* 233). Claiming that there must be a sharp borderline, yet we could never know where it is, is more implausible than the reductive view, he states (*Reasons and Persons* 243).

What becomes apparent is that our definitions of persons and personhood are nearly as varied as those asserting the definitions. During my own discussions with others, I have heard many equate persons with human beings; others claim that persons are equal to the sum total of our memories. I have also heard some claim that we simply *are* persons, while others argue that persons are things we become. It is arguable that only human beings qualify as persons. Wiggins shares a prejudice similar in kind to this, yet not so stringent. That is, he states that persons can never be an artifact (such as robots or machines), but are instead entities of a natural kind – though not necessarily species specific to human beings (*Identities* 161-62). For others³, freedom of the will, along with first-order desires and second-order volitions, is part of the concept of persons. We see the idea that the way we use words and language, and develop meaning all help determine or influence our concepts of persons, which, in turn, determine how we formulate our theories of personal identity. Therefore, if we gain an understanding of how we acquire and employ each of the

above, we will have some insight into how we go about formulating theories of personal identity and conceptualizing in general.

Another consideration to discuss is the notion of utility. Part of the contingent aspects associated with our theories of personal identity derives from varying applications of utility or practicality. For instance, it seems only fitting that when considering questions about personal identity, we also ask why it is important to have a theory of personal identity. Why is personal identity so important? Why does it matter? The reasons for formulating our theories of personal identity are largely based upon the specific applicability that each theory carries. What this means is that the value of a theory of personal identity is determined by its utility.

A theory of personal identity becomes important when we have a need to answer certain questions about persons – often these are questions about meaning and responsibility. For example, some of the problem cases we examined (division and replication) draw attention to questions of responsibility. If divisions of persons, replications, and brain swaps are possible considerations of personal identity, then we must ask how these problem cases affect matters of responsibility as it regards the actions and thoughts of persons involved in such cases. If Michael Ellis' mental content was duplicated into three host bodies, would each host now be responsible for any wrong-doings Michael Ellis may have committed? Do persons with certain psychological illnesses become responsible for their actions if they have no memory of the acts in question? What we find, then, is that sometimes questions about personal identity are questions of legal matters. Responsibility is important for

personal identity because it is the concept that binds a person to the accountability of his or her actions.

Although some questions about personal identity examine aspects of responsibility, this is not always the case. Some of our interest in developing a theory of personal identity is entirely devoid of moral issues. For instance, when we ask “What am I?” or “Am I the same person that I was five years ago?” these questions need not address any legal or moral considerations. We can just as easily ask these sorts of questions about personal identity from an approach of pure ontological curiosity. Accordingly, the focus of our interest would lie on a want to understand concepts about existence, rather than morality.

As stated, our reasons for formulating a theory of personal identity are based upon matters of practicality or utility. The utility of the theory is specifically expressed through its ability to explain differences between the individual and the group. Personhood, whether questioned ontologically or morally, is a concept that concerns matters of individuals.

Language and Meaning

The ability of language to influence our concepts of personal identity, as well as its ability to affect our conceptualizations in general, seems somewhat obvious. For, as Dennett claims, language plays a large role in constructing a human mind, and creatures lacking this ability should not be supposed to have similarly constructed minds (*Consciousness Explained* 447). That language in some way influences or determines our concepts of personal identity is all the more obvious when we consider the goal of language, i.e., communication. That is to say, our language

influences our concepts in that the concepts, themselves, are limited to the forms of expression allowed by the rules of a particular language. In short, we can only conceptualize what our language allows us to express, for its rules are those by which our theories must play, so to speak. As we shall see, we acquire our language socially, which implies that the rules will vary, slightly or drastically, from one language to the next.

Willard Van Orman Quine describes our acquisition of language as a process where, “Each of us learns his language by observing other people’s verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others.”⁴ As such, language is a socially constructed effort of communication, and therein lies part of the problem we face. Being that language is inherently social, there should be little surprise that we have no unanimous, definitive conclusion about many of our concepts, since how we use words and the ways we talk will vary to some degree. What is rather puzzling, however, is the degree of confidence we exhibit when using some terms that appear to have ambiguous or no meaning. For example, sometimes we hear people make statements such as, “That falls outside the scope of reality.” When we pose the question, “What, then, is reality?” we are usually met with brief silence, followed by an explanation just short of ridiculousness – usually similar to “Anything that is real.” I believe we meet similar issues when we consider statements about personal identity. At the very simple level of casual conversation, we seem to know exactly what personal identity is, since we experience it phenomenologically – the belief that we are somehow the same from one moment to the next or a continuant thing. Yet, when we are asked to produce a theory of the

principles upon which personal identity is based, we always appear to fall short of certainty. Do we, then, genuinely understand the meaning of personal identity beyond the phenomenological level?

Quine claims that “Understanding, behaviorally viewed, is thus a statistical effect: it resides in multiplicities.”(59). What this means is that our understanding of words comes about through a series of hit-and-miss attempts to use a word correctly. In some instances, we use a word and it elicits a desired result, while in others our use of a word results in bewilderment and correction. Through this process, we refine our understandings of a word’s meaning. A result of this process, however, is that it necessitates a degree of vagueness, giving no necessary sharp boundaries to the meanings of words, for the sake of fluency of dialogue (Quine, 59).

With regard to how we describe persons, any number of criteria may enter our formulation – or, as Shoemaker states, we may use no criteria at all (*Personal Identity* 127). However, if we do assert a set of criteria about what constitutes personal identity, we do so upon a conceptual formulation of persons that is ultimately a personal bias. As Perry notes, “We have to choose which criterion is most important. It’s a matter of choice of how to use our language....”⁵ Yet, that such choices are made on personal bias does not totally destroy our efforts to understand personal identity. Rather, what this entails is that we must realize that we can only examine personal identity by looking at the implications that derive from each relative formulation. We can think of this approach to personal identity as asking, “If persons are X, then what can be said about their identity?” For instance, if we accept the idea that persons are merely identical to bodies, then there are definitive things we can

infer from this starting point. Yet, it is these starting points themselves that are without solid foundation. There is not a set of rules or a logical maxim that tells us which of our formulations are correct and which are mistaken. Each formulation of our concept about persons begins on equal ground – no one formulation supersedes the others by necessity.

How does such a fundamental confusion about our approach to the problem of personal identity arise? To answer this question we will examine Gilbert Ryle's views regarding our use of language and how it leads to mistakes in our ideas about our identity. Ryle's arguments are among the strongest and most deliberate assault on Cartesian dualism. He refers to Cartesian dualism "with deliberate abusiveness, as 'the dogma of the Ghost in the Machine'."⁶ That is to say, Ryle denies the dualistic idea that minds and bodies are separately existing things, claiming that the dualist position fails because it advances on a family of logical mistakes (17). He refers to these sorts of mistakes as category-mistakes – logical errors that occur when concepts belonging to two different logical types are inappropriately cast within the same logical type. For example, if Michael Ellis states that he owns a left-footed boot and a right-footed boot, as well as a pair of boots, he commits a category-mistake by thinking that the pair of boots is some sort of third entity, rather than being a description of the function of both boots. The pair is merely the collective function of the left and right boot.

How does Ryle relate the idea of the category-mistake to personal identity? We have already seen that Descartes believed that his mind was distinct from his body. This form of dualism is based on the assumption that minds and bodies are

diametrically opposed entities of the same logical type (Ryle, 22). Yet, what Ryle argues is that this is a mistake. He states that existence is not a generic term that applies equally to all concepts and that, as such, mind and body are of different logical types (Ryle, 23). What this means is that when we talk about the existence of physical objects we are not asserting that they exist in the same sense as do things such as ideas, concepts, and thoughts. However, this does not necessarily mean that one kind of object reduces to the other. Although it is often accepted by the materialist or reductionist position that mind reduces to matter, this is not what Ryle argues. The one cannot reduce to the other since it is of an entirely different logical type. Once we recognize that they are of different logical types, the problem reveals itself as a category-mistake.

The type of mind that is described in Ryle's arguments is best conceived as a collective function of our brain processes. In other words, *mind*, as Ryle likens it, is simply a description of the collection of brain functions acting in tandem. Likewise, neuropsychologist Paul Broks describes a system where, "Minds emerge from process and interaction, not substance," and that, "The life of the self depends absolutely on the integrity of brain function."⁷ According to Ryle, the category-mistakes found in the dualist position arise as a result of the misapplication of *index words* – words that designate a particular reference, but are not proper things, themselves (188). In particular, he argues that the ways we employ pronouns create mistaken notions about our identity. Hence, he states, "Gratuitous mystification begins from the moment that we start to peer around for the beings named by our pronouns."(Ryle, 187)

“What am I?” It is this sort of formulation about questions of personal identity that Ryle believes is the source of the category-mistakes in Cartesian dualism. I believe that the reason we treat these words differently (words such as *I* and *you*) is because, as mentioned in previous chapters, we are often times still reluctant to believe that mere matter can produce phenomena such as consciousness, thought, emotions, and all of the things we think separate us from inanimate objects. However, assuming that matter *can* produce such phenomena what we must ask, then, is whether or not the phenomena are reducible to matter.

Do we argue a reductive theory of mental phenomena, which states that such phenomena are identical to mental states, or do we argue that these phenomena are emergent properties *caused by*, but not *equal to*, mental states? If this question has a determinable answer, then I think the way we will discover it is by looking again at the concept of genuine identity. If mental states are *identical* to consciousness, thought, emotions, and the like, then what we say about mental states will also be true of any of the given mental phenomena through the transitivity of identity. Yet, a reductive theory may be insufficient for answering this question. For example, it has long been part of the reductive position of materialism that all things are essentially reducible to matter.⁸ In recent years it was demonstrated that matter could be produced out of pure energy, which certainly discredits the strength of the materialist argument, simply because it makes no sense to assert that all things are reducible to matter, when in fact matter is reducible to energy. We would not, however, say that the matter is identical with the energy that produced it. The matter has emergent properties that are not inherent to the pure energy. Likewise, reductionism may prove

insufficient for describing mental phenomena because they are not, strictly speaking, identical to the mental states that produce them.

Ryle states that we believe there is something other than our body that we refer to when using 'I' – something in the background, which is unique to us and is more than the compilation of our personalia, e.g. name, age, gender (186). What Ryle implies is that there is no special ghostly thing inside our head that words such as 'I' and 'you' name. Rather, these words merely indicate a person to which certain utterances refer. In other words, when Michael Ellis states, "I am not feeling well today," his use of 'I' is not naming some incorporeal being inside his head, but is merely pointing his statement's reference to himself. It is an indication to anyone listening that his statement refers to the person uttering the statement and not to someone else. Ryle observes that there is a special sense of 'I' because it always refers to the person who utters or writes a statement, whereas words such as 'you,' 'she,' and 'they' can represent different references at different times (197-98).

Broks claims that, "People and subjects-of-experience exist as a feature of our language, but in no other way." (218). He states that our descriptions about our selves have developed because of our mistaken belief that we are genuinely unified beings, when we are actually divided and discontinuous (Broks, 41), and concludes that, "The self has no location, however natural it seems for us to believe otherwise." (Broks, 125). These sorts of statements from Ryle and Broks are essentially the same kinds of criticisms we have already cited from others in the previous chapters, such as Hume, Dennett, Searle, and Parfit. What Ryle's argument adds to our discussion,

however, is a detailed explanation of the source of these linguistic problems about personal identity.

Earlier in this chapter, I proposed the idea of relative facts, wherein we measure the truth-value of statements based on a frame of reference. John Austin offers a similar idea pertaining to the meanings of words. He argues that linguistic meaning derives essentially from sentences, rather than individual words. For example, he states, "...to say that a word or a phrase 'has a meaning' is to say that there are sentences in which it occurs which 'have meanings': and to know the meaning which the word or phrase has, is to know the meanings of sentences in which it occurs."⁹ The frame of reference for words, then, is the sentences within which they are used. Accordingly, we come to understand a word's meaning through our observations about how the word is used within various sentences. This explanation about meaning is very much akin to what Quine argued regarding the hit-and-miss behavioral understanding we use with language. That is, we understand words through the way they are used within sentences, and the proper use of these words is learned through our social interaction with others. We learn which sentences produce the desired responses and which ones do not. Quine develops Austin's point further, arguing that a sentence's meaning is not fixed, but is acquired, rather, only within frames of reference, such as theories, paradigms, and conceptual schemes.¹⁰

Austin states that part of our misunderstanding about the meanings of words derives from our belief that all words function similarly to names, in that they allegedly designate something the same way that proper names do. A second aspect of our misunderstanding arises from our inclination to analyze the individual words

within a sentence, rather than analyzing the sentence as a whole (Austin, 61).

Austin's idea is not unlike the point argued by Ryle when we examined the concept of index words. That we use 'I' in a way which names something other than the person speaking or writing a sentence affirms Austin's first criticism here.

We can see, then, that the difficulties of constructing a definitive theory of personal identity arise from not one, but a multitude of issues related to our use of language. From the arbitrary nature of defining personhood to our misconceptions about the analysis of meaning, we begin to realize that the failures or shortcomings of the traditional theories of personal identity are not, essentially, a result of faulty logic within the theories themselves. I agree with Parfit's claim that we have sufficient reasons to reject any of these theories. Yet, I also think that the fundamental problem with the theories does not derive from their internal logical construction. Rather, the problems arise from the fact that our initial approach to the subject matter is incorrect. We move ahead upon the false presumption that our assertions about persons are factually based statements, when we should understand that the foundations upon which our theories rest are ultimately contingent frameworks of meaning.

Notes

¹ Daniel Dennett, *Brainstorms* (Cambridge: MIT P, 1993) 267.

² John Perry, "The Importance of Being Identical," *The Identities of Persons*, ed. Amelie Oksenberg Rorty (Berkeley: U of Berkeley P, 1976) 68.

³ See Harry G. Frankfurt's treatment of this idea in "Freedom of the Will and the Concept of Person."

⁴ W.V. Quine, *The Pursuit of Truth* (Cambridge: Harvard UP, 1993) 38.

⁵ John Perry, *A Dialogue on Personal Identity and Immortality* (Indianapolis: Hackett, 1978) 40.

⁶ Gilbert Ryle, *The Concept of Mind* (Chicago: U of Chicago P, 1984) 15.

⁷ Paul Broks, *Into the Silent Land* (New York: Atlantic Monthly, 2003) 56.

⁸ Jeffrey Winters, "Let There Be Matter," *Discover* December 1997: 40.

⁹ J.L. Austin, *Philosophical Papers* ed. J.O. Urmson and G.J. Warnock (New York: Oxford UP, 1979) 56.

¹⁰ See, for example, Quine's account in *Two Dogmas of Empiricism*.

References

- Austin, J. L. *Philosophical Papers*, ed. J.O. Urmson and G.J. Warnock. New York: Oxford UP, 1979.
- Ayer, Alfred Jules, *Language Truth and Logic*. New York: Dover, 1952.
- Bennett, Jonathan. "Locke's Philosophy of Mind," *The Cambridge Companion to Locke*. ed. Vere Chappell. Cambridge: Cambridge UP, 1997.
- Broks, Paul. *Into the Silent Land*. New York: Atlantic Monthly, 2003.
- Butler, Joseph. *The Analogy of Religion*. London: Bell and Daldy, 1857.
- Churchland, Paul. *Matter and Consciousness*. Cambridge: MIT P, 1994.
- Dennett, Daniel. *Consciousness Explained*. New York: Little, Brown & Company, 1991.
———. *Brainstorms*. Cambridge: MIT P, 1993.
- Descartes, Rene. *The Philosophical Writings of Descartes*. vol. II, trans. John Cottingham, Robert Stoothoff, Dugald Murdoch. Cambridge: Cambridge UP, 1995.
- Frankfort, Harry G. "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy*, 68 Jan. 14, 1971.
- Gott, J. Richard. *Time Travel In Einstein's Universe*. New York: Houghton Mifflin, 2001.
- Heinlein, Robert A. *The Fantasies of Robert A. Heinlein*. New York: Tom Doherty, 1999.
- Hume, David. *A Treatise of Human Nature*. analytical index by L.A. Selby-Bigge, second ed. revised by P.H. Nidditch. New York: Oxford UP, 1978.
- Lewis, David. "Survival and Identity," *The Identities of Persons*. ed. Amelie Oksenberg Rorty. Berkeley: U of California P, 1976.
- Locke, John. *An Essay concerning Human Understanding*. ed. Peter H. Nidditch. Oxford: Clarendon, 1975.
- Nagel, Thomas. *Mortal Questions*. Cambridge: Cambridge UP, 1995.

- . “Brain Bisection and the Unity of Consciousness,” *Personal Identity*. ed. John Perry. Berkeley: U of California P, 1975.
- Parfit, Derek. *Reasons and Persons*. New York: Oxford UP, 1984.
- . “Personal Identity,” *The Philosophy of Mind*. ed. Jonathan Glover. Oxford: Oxford UP, 1986.
- . “Lewis, Perry and What Matters,” *The Identities of Persons*. ed. Amelie Oksenberg Rorty. Berkeley: U of California P, 1976.
- Perry, John. *Personal Identity*. ed. John Perry. Berkeley: U of California P, 1975.
- . *A Dialogue on Personal Identity and Immortality*. Indianapolis: Hackett, 1978.
- . “The Importance of Being Identical,” *The Identities of Persons*. ed. Amelie Oksenberg Rorty. Berkeley: U of California P, 1976.
- Quine, W.V. *Pursuit of Truth*. revised edition. Cambridge: Harvard UP, 1993.
- Quinton, Anthony. “The Soul,” *Personal Identity*. ed. John Perry. Berkeley: U of California P, 1975.
- Reid, Thomas. “Of Identity,” *Essays on the Intellectual Powers of Man*. London: Macmillan, 1941.
- . “Of Mr. Locke’s Account of Our Personal Identity,” *Essays on the Intellectual Powers of Man*. London: Macmillan, 1941.
- Ryle, Gilbert. *The Concept of Mind*. Chicago: U of Chicago P, 1984.
- Searle, John. *Minds, Brains and Science*. Cambridge: Harvard UP, 1997.
- Shoemaker, Sydney & Richard Swinburne. *Personal Identity*. Oxford: Basil Blackwell, 1985.
- . “Personal Identity and Memory,” *Personal Identity*. ed. John Perry. Berkeley: U of California P, 1975.
- Weatherford, Roy. *The Implications of Determinism*. New York: Routledge, 1991.
- Wiggins, David. “Locke, Butler and the Stream of Consciousness,” *The Identities of Persons*. ed. Amelie Oksenberg Rorty. Berkeley: U of California P, 1976.
- Williams, Bernard. *Problems of the Self*. Cambridge: Cambridge UP, 1999.
- Winters, Jeffrey, “Let There Be Matter,” *Discover* Dec. 1997.