Thinkers

Stephen Matthew Duncan
Buena Park, California

Bachelor of Science in Philosophy and Psychology, Westmont College, 2007
Master of Arts in Philosophy, Georgia State University, 2010

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Philosophy

University of Virginia
May, 2015

For Jim

ACKNOWLEDGEMENTS

It is with deepest gratitude that I thank my adviser, Trenton Merricks. He has been a tireless resource on this project. He read countless drafts, offered innumerable insights, raised powerful objections, and pushed me all along the way while still remaining unwaveringly supportive. I can't imagine this project being more rewarding to write or nearly as beneficial to me as a philosopher had it been written under any other adviser. Every advisee should be so lucky. That so few are only deepens my gratitude to Trenton.

Special thanks are also due to Brie Gertler, who read more drafts and gave more comments on this project than she should have permitted. In many ways she was my second advisor, and a great one at that.

I also thank the other members of my committee: Harold Langsam, Ross Cameron, and Michael Kubovy. I thank them for reading my dissertation, asking questions, raising concerns, offering advice, and giving guidance.

Thanks are also due to the University of Virginia Metaphysics Group, who read and dissected much of this dissertation in its infancy. And special thanks are due to Alexandre Billon, Doug Reed, Derek Lam, John Mahlan, Andrew Morgan, Paul Nedelisky, Nick Rimell, and Adam Tiller for particularly useful comments on various chapters, and for countless conversations about the work therein.

Finally, I thank Megan and Thomas Duncan. Doing stuff like this is so much easier and more joyous because of you two.

TABLE OF CONTENTS

INTRODUCTION

You and I are persons. We both exist right now, and we both existed yesterday. Hopefully we will both still exist tomorrow, the next day, and for a while thereafter. Of course, we may change in the process. People have a tendency to do that. What we are thinking and feeling changes from moment to moment. And the physical makeup of our bodies is also in constant flux. Yet people like us survive many of these changes. We continue existing despite continual transformation.

All of this seems obvious enough. But these apparent truisms give rise to plenty of further questions and puzzles about what we are both in and through time. This is the topic that I will address in this dissertation. First I will argue that introspection is an especially good way to gain knowledge about what we are. Then I will use the deliverances of this method to argue for my own view, which is that we are *thinkers*—things essentially capable of undergoing a certain distinctive form of conscious experience.

There are many good reasons to pursue a better understanding of what we are. For one thing, this topic bears on various philosophical and scientific issues—issues about moral and legal responsibility, human rights, the mind-body problem, the implications of brain damage and mental disorder, the possibility of surviving death, and many others. All of these issues are related in some way to personal identity. Thus, insofar as each of these areas of inquiry is worthwhile, it is a good idea to pursue a better understanding of persons. There is also intrinsic value to knowing more about what we are. And there is practical value here as well, since people sometimes have to make difficult decisions

about themselves or other people, such as those concerning the beginning or end of life. These are just a few reasons why the topic of personal identity is worth pursuing.

This topic has been approached in many different ways. Many different methods and sources of evidence have been used.[1] It is interesting to note, however, that those who started the debate—modern philosophers like Locke, Butler, Leibniz, Hume, and Reid—held that there is a *privileged* route to knowledge of what we are: introspection.[2] Locke (1690/1975), for instance, suggests that one's introspective awareness of past experiences is the primary means by which one is aware of one's identity through time. And this, together with other philosophical concerns, leads him to the view that personal identity consists in the sameness of consciousness through time.[3] Many of Locke's contemporaries disagree with his view. But the main evidence that they bring against it is, again, introspective. Joseph Butler (1736/2008), for example, says:

> But though consciousness of what is past does thus ascertain our personal identity to ourselves, yet, to say that it makes personal identity, or is necessary to our being the same persons, is to say, that a person has not existed a single moment, nor done one action, but what he can remember; indeed, none but what he reflects on. And one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes (p. 100).

---

[1] For instance, many philosophers appeal to intuitions and thought experiments when they talk about personal identity (e.g., Shoemaker, 1985; Dainton, 2008). Others would rather scrap these methods and base the debate entirely on what science tells us about human beings (e.g., Wilkes, 1994). Still others prefer to think of persons as social constructions; so they draw on anthropological, sociological, or historical data to answer questions about personal identity (e.g., Hacking, 2000).

[2] See Locke's *Essay* (1690, II, xxvii), Butler's *Analogy* (1736), Leibniz's *New Essay* (1739, p. 236), Hume's *Treatise* (1739, I.iv.6), and Reid's *Essays* (1855). One might also add Descartes' *Meditations* (1643) to this list, since Descartes appeals to introspection to defend a view that has clear implications for personal identity. However, Locke and his interlocutors are generally held to be the originators of the debate on personal identity.

[3] Locke's emphasis on self-knowledge when it comes to debates about personal identity is especially evident in *Essay* II, xxvii. Of this section, Harold Noonan (2003) writes, "Despite his agnosticism about Cartesian dualism, Locke himself implicitly accepts the Cartesian emphasis on the first-person viewpoint as providing a privileged standpoint from which proposals about the nature of the self can be judged" (p. 32).

Butler disagrees with Locke's view of personal identity. But he agrees that introspective evidence is the way to "ascertain our personal identity" (ibid.). In a similar vein, Thomas Reid (1855/2008) denies Locke's view of personal identity while also agreeing that introspection is privileged in this debate. He says:

> The conviction which every man has of his identity, as far back as his memory reaches, needs no aid of philosophy to strengthen it; and no philosophy can weaken it, without first producing some degree of insanity (p. 107).

And Reid later says:

> How do you know—what evidence have you—that there is such a permanent self which has a claim to all the thoughts, actions, and feelings which you call yours? To this I answer, that the proper evidence I have of all this is remembrance (p. 110).

Even Hume (1739/1975), who is skeptical about the very existence of selves, appeals to introspection to make his case (I.iv.6). He says that when he introspects he only finds various perceptions—he finds no substantial self. And this leads him to be skeptical about the self. The motivations for such a focus on introspection in early debates about personal identity are no doubt complex. What seems clear, though, is that the originators of this debate believed that introspection is a particularly good method for gaining knowledge about one's own identity. That they believed this does not, in itself, establish anything about the connection between introspection and personal identity. But if these philosophers were *right* about introspection being an especially reliable route to knowledge about personal identity, then it seems that introspection may occupy a position of privilege in this debate. In other words, if our best evidence about ourselves comes from introspection, then we may have reason to prefer this evidence to other sources of evidence about ourselves.

There has been a lot of recent discussion about the epistemic merits of introspection.[4] So bringing this discussion to bear on debates about personal identity is one way to add something to the debate while at the same time reuniting it with its origins. So that is the approach that I will take. I will start by arguing that introspection can provide us with particularly good evidence about ourselves. Then I will use this result to argue for my own view and against competing views about personal identity. I do not mean to suggest that my arguments will all be based on introspection. I will use various resources—including conceptual, empirical, phenomenological, and other resources—to show that, through introspection, we each have access to especially good information from which we can reason about personal identity.

Assumptions

Before moving on to a more detailed summary of this dissertation, let me first introduce, and to some extent motivate, a few assumptions that will be operative within it. First, I assume that there are people (I am one of them, and so are you), and that, in general, people persist for appreciable periods of time; that is, for days, months, years, decades, and sometimes even a century or more. In this dissertation I do not address the question of whether we exist and persist through time. I assume that we do. Rather, I address the question of *how*, or in virtue of what, we exist and persist through time.

Second, I assume that there are *experiences*, and that people like us undergo them. I make no substantive assumptions about the ontological nature of experiences. But I do assume that we undergo them. Here, and throughout this dissertation, I use 'experience'

[4] See Gertler (2011) and Cassam (1994) for helpful surveys of the topic.

to refer to *phenomenal states*—states that have a phenomenology or what-it's-likeness—such as pains, visual sensations, inner speech, mental imagery, and so on.

Third, I assume that people *have* or *undergo* these various experiences. That is, I assume we are substantial *subjects* of experience, rather than, say, collections or bundles of experiences. In Chapters 2 and 5 I will expound upon and defend this assumption. But, at the end of the day, I do not wish to fight over it too much.

Fourth, I assume that we often *know* what we are occurrently experiencing. And I assume that our judgments about our own occurrent experiences, when formed on the basis of careful introspection, are at least sometimes especially reliable or certain.[5] Some philosophers may worry about my approach to this dissertation because they are skeptical, or even doubtful, about our ability to know things about our selves or our minds through introspection. But the above assumptions about self-knowledge are modest enough to elude the main sources of philosophers' worries about our self-knowledge.

One such worry is that empirical research shows that our introspective judgments are not all that reliable. My assumptions about self-knowledge are untouched by this research. For one thing, much of this research doesn't concern our knowledge of our occurrent experiences. It concerns our knowledge of our personality and character traits, or our ability to predict how we will feel in the future, or our knowledge of the causes of our actions and attitudes, or our knowledge of our (non-conscious) dispositional beliefs

[5] A related claim that might be considered an assumption of this dissertation is that we can at least sometimes be *directly aware* of our experiences. I say that this claim "might" be considered an assumption because although I give some good reasons to accept it in Chapter 2, I do not engage in a sustained defense of it.

and desires.[6] That we are epistemically limited in these domains is irrelevant to my assumptions and to my project in general. Some research suggests that our judgments about our moods or emotional states, which may have an experiential component, are often mistaken. But this still doesn't threaten my assumptions. For I only assume that our beliefs about our occurrent experiences are *sometimes* especially reliable or certain. And, at any rate, I will not discuss moods or emotions in this dissertation. So worries about our knowledge of these states can be set aside. The mental states that I will deal with are all very distinctive, very simple experiences such as a sharp pain, the visual sensation of a peach on a blank screen, and the thought that 2+2=4.[7] Furthermore, my discussion will be limited to cases in which a person is in ideal introspective circumstances, with no distractions, and is carefully attending to her experiences. Thus, in the context of this dissertation, my assumptions about self-knowledge are safe from the above worry.

There are other, non-empirical worries about self-knowledge out there.[8] But, in general, these worries concern our knowledge of experiences with very complicated or otherwise difficult to discern contents, and they only cast doubt on claims such as that we are *always* in a position to know our experiences, or that our judgments about our experiences are *infallible*, or that our knowledge of our experiences is *foundational*. I do not assume that we are always in a position to know our experiences. Nor do I assume that we are infallible when it comes to our experiences. Nor do I assume that our knowledge of our experiences is foundational. And, again, the cases I will deal with all

---

[6] For a helpful summary of this research and its relevance to self-knowledge, see Gertler (2011, Ch. 3).
[7] As I will note in Chapter 3, it is controversial whether thoughts have distinctive phenomenal qualities, and thus, count as experiences (or phenomenal states). However, those who are skeptical about the phenomenology of thought can replace my talk of thoughts with talk of mental imagery or inner speech without affecting my argument.
[8] These include worries raised by Timothy Williamson's (2000) anti-luminosity argument, and by the so-called "speckled hen problem" (see, e.g., Chisholm,1989; Fumerton, 2005).

involve very distinctive, very simple experiences had by someone who is in ideal introspective circumstances and who is carefully attending to her experiences. Thus, my assumptions about self-knowledge remain safe.

Another worry about introspection is very general. It's that, since introspection is an inherently *private* activity, the deliverances of introspection are not scientifically respectable.[9] This worry is somewhat paradoxical, since a lot of seemingly respectable scientific data from cognitive science is derived from introspective reports. But, regardless, what's important for my purposes is that our judgments about our own experiences are *epistemically* respectable in that they often amount to knowledge, and are at least sometimes especially reliable or certain. I assume that this is the case. I'll let others decide whether that's good enough for science.

Moving on, my fifth assumption is a bit more technical. It is that we are *immune to error through misidentification*. The idea is that if, from the first-person perspective, I judge that I am experiencing X, then although I could be wrong in believing that X is what I am experiencing, I couldn't be wrong in believing that it is *I* who is the subject of the experience.[10] For example, if I think I see a tiger, then I might be wrong about what I see—it might not be a tiger—but I couldn't be wrong in thinking that it is *I* who is having the visual experience in question. I am immune to such errors of misidentification. Here the claim isn't that I am immune to all possible errors of experiential self-attribution. If

[9] For a helpful discussion of this worry, and of the relationship between science (and cognitive science, in particular), publicity, and the epistemic merits of introspection, see Goldman (1997).

[10] Another way to put it is: If I judge that I am experiencing X, then I couldn't be right that *someone* is experiencing X, but wrong that it is *I* who is experiencing X. Sydney Shoemaker (1968) was first to make this principle explicit. And it is now widely accepted (though there is disagreement as to *why* it is true (cf., Gertler, 2011, p. 217)).

my attribution is based on third-personal evidence, then the possibility of error arises.[11] And, as we will see in Chapter 1, there are ways of self-attributing experiences—ways other than attributing them to oneself as their *subject*—where self-misidentification can, and in fact *does*, occur. So my assumption about immunity to error is limited to cases in which one's judgment is made from the first-person perspective and is to the effect that one is the subject of a particular experience.

Sixth, and finally, in this dissertation I will speak in a way that best fits with *three-dimensionalism*, the view that objects are wholly present at each moment of their existence and thereby persist by enduring through time. However, most, if not all, of the central insights of this dissertation can be reformulated in a way amenable to four-dimensionalism, the view that objects have temporal parts and persist by perduring—by having temporal parts at various times. So although I will speak as if three-dimensionalism is true, four-dimensionalists needn't abandon this dissertation as a lost cause.

Those are my main assumptions. They are relatively modest. And even if one has qualms about one or more of them, I hope (and predict) that one will still be able to find something of interest in this dissertation.

Summary

In Chapter 1 I will argue that there is an experience of the self. I will start by getting clear on the implications of this thesis. Then I will appeal to empirical findings on

[11] For example, if I judge that I think oysters are tasty because I am looking at a photo in which a person who looks like me is scarfing down oysters, I might mistakenly think that it is I who thinks oysters are tasty.

schizophrenia to make the case in favor of it. Specifically, I will appeal to research on a phenomenon called "thought insertion". Those who experience inserted thoughts suffer from a breakdown in their experience of themselves as the author/agent of some of their thoughts, leading them to claim that their thoughts have external authors/agents. (Which is not to say that they make the sort of error of misidentification that I just said people are immune to—i.e., misidentification of the *subject* of one's thought. They don't make that mistake.) This suggests that normally—in non-pathological cases—people experience themselves as the authors/agents of their thoughts. Which implies that people normally experience themselves.

After going into some detail about thought insertion, and after making the case that the best explanation for thought insertion implies that people normally experience themselves, I will consider objections. First I will consider several alternative explanations of thought insertion that do not imply that we experience ourselves. Then I will consider objections according to which "self-experience" is nothing more than a misconception or something like an illusion—an experience that isn't really of the self. Finally, I will conclude by saying something about why various philosophers have (mistakenly) denied that there are self-experiences.

In Chapter 2, I will go beyond the previous chapter and argue that we can be, and often are, directly aware of—i.e., *acquainted with*—ourselves. First, I will describe in detail what it means to be acquainted with oneself. Then I will introduce, develop, and defend a commonly used test for acquaintance—called 'The Doubt Test'—which says: If it seems to S that she is aware of some x, and on the basis of that seeming awareness of x S cannot doubt that x exists, then S is acquainted with x. The Doubt Test is a commonly

used test for acquaintance. Its main rationale is this: If I am aware of some x, but only in virtue of being aware of some distinct y that indicates x's existence, then I can doubt that y is a faithful witness to the existence of x; however, if I am directly aware of x *itself*, then I cannot doubt that x exists, because there is no potentially misleading presentation of x that might allow me to doubt its existence.

After further detailing the nuances and motivations for using The Doubt Test, I will adopt it as a good test for acquaintance. Then I will apply it to us and show that we pass. I will consider various objections to my argument. But ultimately I will conclude that we can be, and often are, acquainted with ourselves.

In Chapter 3 I will show that the results of the previous chapters allow us to definitively rule out several of the most prominent theories of personal identity. These include, most notably, animalism and various psychological continuity theories.

I will begin with two scenarios: One in which you think '2+2=4', and one in which you perceive a briefly presented image on a theater screen. The results of the previous chapters imply that you can be directly aware of your thought or perception, as well as of yourself as their subject. In this chapter I will show that there are temporal constraints on both thought and perception that imply that when you think '2+2=4', or perceive the image on the screen, you are the subject of a temporally extended mental event. This implies that your direct awareness of your thought or perception is direct awareness of a temporally extended event. And it also implies that your direct awareness of yourself as the subject of either mental event is direct awareness of a temporally extended individual.

I will then turn to the aforementioned theories of personal identity and show that there are possible cases in which you think '2+2=4' or perceive the image on the screen even though you lose a part or property that a theory of personal identity under question says is essential to your survival. I will argue that this allows you to rule out any such theory with a very high level of confidence, since in the cases described you can be directly aware of yourself as the subject of your thought or your perception.

Given that the previous chapter rules out some of the most prominent theories of personal identity, one might suppose that there really are no criteria for personal identity. This view is 'anti-criterialism'. Anti-criterialists contend that there are no informative conditions that are both necessary and sufficient for a person's persistence through time. In Chapter 4, I will argue that, in fact, there *is* a criterion for personal identity through time.

I will start by arguing that anti-criterialists are committed to the claim that there are no informative *sufficient* conditions for personal identity through time. Then I will describe a case in which all of the qualitative (i.e., non-identity-assuming) conditions that normally obtain in a persisting person actually *do* obtain, and show that anti-criterialists are committed to the absurd conclusion that such a person might fail to persist through time. I will consider several objections to my arguments, and then conclude that anti-criterialism is false.

Finally, in Chapter 5 I will defend my own theory of personal identity. I will begin by laying out what I take to be the viable theories of personal identity that were not ruled out by previous chapters. Then I will introduce and defend my view, which I will call 'THINKERS'. This view has three essential components: (i) We are substantial subjects

of experience who (ii) are essentially capable of undergoing a certain distinctive form of conscious experience; but (iii) we do not know what this form of consciousness depends on ontologically, so we do not know what our underlying ontological nature is. Hence, we are *thinkers*, but we cannot say whether we are essentially physical or biological beings, for instance.

After describing and arguing for my view I will spell out several of its advantages. First, THINKERS helps us resolve conflicting intuitions about personal identity brought out by certain classic puzzles. Second, THINKERS can make sense of the fact that, as persons, we have certain unique traits such as the ability to reason, reflect, and act; to self-ascribe properties; and to bear moral responsibility. Finally, THINKERS suggests a direction for future research on personal identity. In order to be able to say more about our persistence conditions, we have to be able to better understand the ontological nature of consciousness. This is no easy task, of course. But I suggest that thinking and reasoning about the nature of consciousness is the best way to push the debate about personal identity forward.

CHAPTER 1

The Self Shows Up in Experience

I can be aware of myself, and thereby come to know things about myself, in a variety of different ways. For instance, I can look into a mirror, see myself, and in doing so find out what I look like, what I am wearing, or whether my hair is sticking up in the back. I can be aware of myself in other ways too. I can look down at my feet. I can go to therapy. I can get an MRI. I can listen to myself talking. I can observe my behavior. And so on. There are many different ways in which I can be aware of myself.

But is there some *special* way in which I—and *only* I—can be aware of myself? Is there a path to self-awareness that I alone can take? Can I be aware of myself just by *introspecting*? Do I somehow *show up* in my own conscious experiences?

Here's what David Hume (1739/1975) says:

> For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception (I.iv.6).

When Hume introspects, he only finds various mental states. He doesn't find anything *extra* that we might call "the self". A lot of philosophers agree with Hume on this point. For example, Gilbert Ryle (1949/2002) says:

> Self-consciousness, if the word is to be used at all, must not be described on the hallowed para-optical model, as a torch that illuminates itself by beams of its own light reflected from a mirror in its own insides (p. 194-195).[12]

[12] Ryle (1949/2002) goes on to say, "Even if the person is, for special speculative purposes, momentarily concentrating on the Problem of the Self, he has failed and knows that he has failed to catch more than the flying coat-tails of that which he was pursuing" (p. 198).

David Armstrong (1968) similarly says, "All that inner sense reveals is the occurrence of individual mental happenings," and so he concludes that Hume was right—the self can't be found through introspection (p. 337). More recently, José Bermúdez (1999) claims that when one is the subject of a mental state like the belief that p, "There is no consciousness of anything other than the state of affairs that p—or the possible state of affairs that p. *A fortiori*, there is no consciousness of self as subject."

Even some philosophers who think that we *can* be introspectively aware of the self agree with Hume's claim about the contents of experience. Robert Howell (2006), for example, says:

> The problem is that upon introspection, and upon performing the cogito, there does not appear to be anything salient corresponding to a self—the I of "I exists" … There is no acquaintance with the self or with any sort of conceptual/representational stand-in for the self. There are many ways to finesse the notion of self-acquaintance, but the basic phenomenological data adduced by Hume must be respected (p. 44, 46).

Brie Gertler (2011) says, "Most philosophers find Hume's claim phenomenologically plausible" (p. 210). And Sydney Shoemaker (1994) says that Hume's view on this matter has "commanded the assent of the majority of subsequent philosophers who have addressed the issue" (p. 188).[13] So there appears to be fairly widespread agreement that the self cannot be found in experience. All there is to experience is mental states such as sensations, perceptions, emotions, and thoughts. There is nothing *further* that is (or that represents) the self. In short, experience is self-free. The self does not show up in experience.

[13] To be fair, while it's true that few contemporary philosophers argue directly against Hume's position, more than a few philosophers at least register their disagreement with it, including Strawson (2000), Bayne (2008), Kriegel (2004), Graham (2002), and Swinburne (1985).

That's the received view, and I mean to challenge it. In what follows I will argue that the self *does* show up in experience. To make my case, I will appeal to empirical research on schizophrenia. This research is becoming more widely appreciated in the literature on self-awareness. But thus far no one has shown that this research strongly supports the claim that the self shows up in experience. That's what I will do.

1. Self-Experience

Here is my claim:

SELF-EXPERIENCE: The self shows up in experience.

Before defending SELF-EXPERIENCE, let me explain it. First, by 'the self' I mean *the person*. You are your self, and I am my self. I my*self* have experiences, think thoughts, persist, and have various other traits at various times. So when I say that my self shows up in my experiences, I mean *I*—MD, the person—show up in my own experiences. Philosophers and psychologists talk about 'the self' in many different ways. My claim—i.e., SELF-EXPERIENCE—doesn't necessarily apply to all those senses of 'the self'. It only applies to those senses according to which each person is identical to his or her own self.

By 'experience' I mean *phenomenal consciousness*; I mean that aspect of mentality that has a *what-it's-likeness*. There is something that it's like for me to feel a dull aching pain in my knee, or to hear the chirping of a bird, or to smell the aroma of

freshly brewed coffee. These are conscious experiences. And so, if SELF-EXPERIENCE is true, then there is an experience of the self that has a certain phenomenal character or what-it's-likeness.

What it means to *show up in* experience is a bit more complicated. Very roughly, when I say that the self shows up in experience, I mean that the self enters into one's own private conscious experiences. I mean that Hume is wrong: The self *can* be found in experience. There are a couple of different ways this could go. One option is to understand self-experience as awareness of oneself via awareness of a kind of mental *representation* of oneself. So then self-experience is one among the many kinds of mental representations of which one is a subject. Another option is to treat self-experience as involving direct, unmediated awareness of oneself. On this view, self-experience is the what-it's-likeness of being directly aware of oneself. In Chapter 2 I will defend this thesis. But I will remain neutral on this issue for now.

Regardless of whether self-experience involves direct or indirect awareness of the self, when I say that the self shows up in experience, I mean that an *inner* experience of the self, with its own *proprietary* phenomenology, normally forms a *distinctive* component of one's total experience. I say that this experience is *inner* to indicate that SELF-EXPERIENCE is a claim about a particular way that the self shows up in experience. It's different from the way in which I experience myself when I look in a mirror, or in which you experience me when you talk to me. According to SELF-EXPERIENCE, I experience myself *from the inside*, so to speak, in roughly the same way that I experience my pains, thoughts, or emotions. The self (or a representation of the self) is among those

items that I can be aware of through introspection.[14] Self-experience has its own *proprietary* phenomenology in that it makes its own *sui generis* contribution to one's total phenomenology; it *adds* something to one's experiences.[15] And self-experience is *distinctive* in that its phenomenal character is unique—it is different from that of other kinds of experiences.

I believe that the self *always* shows up in this way in *every* human experience. But I will not insist on it here. At this point I don't want to rule out the possibility of rare cases in which one's self is absent from one's experiences. So my admittedly vague claim will be that one's self *normally* shows up in one's experience. That's the claim that I will defend. And so, with that, SELF-EXPERIENCE can be restated in this expanded form:

> **SELF-EXPERIENCE**: The self shows up in experience; that is, an inner experience of one's self (the person) has its own proprietary phenomenology that normally constitutes a distinctive component of one's total phenomenal experience.

That's what SELF-EXPERIENCE says. Now notice what it doesn't say. SELF-EXPERIENCE doesn't specify the *manner in which* the self shows up in experience. I will remain silent on matters concerning the underlying mechanisms (neural or otherwise) responsible for self-experience.[16] And, again, I will remain neutral as to whether the self *itself* shows up in experience, or is merely *represented* in experience. Also, SELF-

---

[14] I do not wish to commit myself to any particular view of introspection here. By 'introspection' I just mean that method by which we come to know our own mental states (and perhaps selves) from the first-person perspective.

[15] This addition to experience is of course not merely *caused* by the self in the way that a perception of light may be caused by the sun without itself actually being a perception *of the sun*. Self-experience is an experience *of the self*; it is not just a byproduct or effect of the self's presence.

[16] For discussions of these issues, see Gallagher and Shear (1999), and Zahavi (2000).

EXPERIENCE doesn't say that the *essence* of the self is what shows up in experience. So I will remain silent about the essential nature of the self. Finally, it's natural to assume that SELF-EXPERIENCE has certain *epistemic* implications. For one is usually in a position to know what one is experiencing.[17] However, SELF-EXPERIENCE is not primarily an epistemic thesis about self-knowledge or self-awareness; rather, it's a metaphysical thesis about the contents of our experiences. So although I will return to certain epistemic issues in §4, I will set them aside for now.

There's a lot more that could be said here. But it's not my goal to provide a full-blown account of self-experience; my goal is just to show that we do, in fact, experience our selves. So what I've said should suffice for understanding SELF-EXPERIENCE. And it should also suffice for understanding what it takes to *deny* SELF-EXPERIENCE. Various philosophers deny SELF-EXPERIENCE in various different ways for various different reasons. At one point Hume (1739/1975) even denies that substantial selves *exist*. Others are more cautious. Many allow that selves exist, but deny that there is any sense in which we have inner experiences of ourselves. Some grant that we experience ourselves in the very limited sense that we experience our own mental properties; but they still deny SELF-EXPERIENCE, because they deny that our experiences contain any *extra* component, over and above our (non-self-implicating) mental states, that would count as an experience of the self.[18] So there are several ways to deny SELF-EXPERIENCE. But anyone who denies SELF-EXPERIENCE at least agrees that our experiences are self-free in the sense that there

[17] Now, one could be in a position to know something and yet not know it. I think this is the position that Hume et al. are in with respect to self-experience. In §4 I'll consider some possible causes of their error.

[18] See, e.g., Howell (2010). Howell grants that we experience ourselves in the minimal sense described above, but then he claims that, "… a subject's mental properties do not present themselves as properties of the subject. While he is aware of them in some sense, they are not in fact salient to the subject *as his properties*: they are phenomenologically exhausted in their presentation of the world" (p. 476).

are no inner experiences of the self over and above our various sensations, emotions, and other mental states. And that, in the end, is what it takes to deny SELF-EXPERIENCE.

2. In Defense of Self-Experience

Here is my argument for SELF-EXPERIENCE:

P1. If the self does not normally show up in experience, then the self does not normally show up in experience in any particular role.

P2. But the self *does* normally show up in experience in the role of author/agent of one's thoughts.

P3. Therefore, the self normally shows up in experience.

The above argument is valid. And P1 is beyond reproach. Thus, if P2 is true, then SELF-EXPERIENCE is true. So how might P2 be defended? I could start by saying that it seems obvious to me (introspectively) that I experience myself as the author of my thoughts. And since my experiences are probably not exceptional in this regard, I may therefore infer that P2 is true. But this isn't going to convince Hume and company. For Hume et al. report different results. They can't find their selves (in any role) when they introspect.

So I'll go a different direction. In what follows I will defend P2 by appealing to research on a phenomenon found among schizophrenics called 'thought insertion'. Those who experience thought insertion suffer from a breakdown in their experience of

themselves as the author/agent of some of their thoughts. This suggests that normally—in non-pathological cases—people experience themselves as the authors/agents of their thoughts.[19]

2.1 Thought Insertion

> Here are two standard descriptions of thought insertion:
>
> Thinking, like all conscious activities, is experienced as an activity which is being carried on by the subject … There is a quality of "my-ness" connected with the thought. In schizophrenia this sense of the possession of one's thoughts may be impaired and the patient may suffer from alienation of thought … [The patient] is certain that alien thoughts have been inserted in his mind (Fish, 1984, p. 48; cited in Stephens and Graham, 2000, p. 119).[20]
>
> In thought-alienation [i.e., thought insertion] the patient has the experience that … others are participating in his thinking. He feels that thoughts are being inserted into his mind and he recognizes them as foreign and coming from without (Fish, 1985, p. 49; in Stephens and Graham, 2000, p. 121).

People who suffer from thought insertion believe that they experience the thoughts of others. It's not just that they believe others are controlling or influencing their thoughts; they believe that others are actually *thinking* some of the thoughts they experience (see Wing, 1978, p. 105; Fulford, 1989, p. 221; Stephens and Graham, 2000, p. 121). They believe that external agents are literally inserting thoughts into their minds.

---

[19] There are other symptoms of schizophrenia and other disorders that may also support my point, including schizophrenics' experiences of alien voices and delusions of control (see Frith and Johnston, 2003, ch. 7; Graham and Stephens, 2000; Bayne, 2008), and Anarchic Hand Syndrome (see Bayne, 2004, 2008; Campbell, 2002). I will not discuss these cases in this chapter.

[20] I recognize that this passage may seem tendentious in the present context. First, one might think that whether experiences have a "my-ness" is precisely what's at issue in this chapter. That may be right. So I won't just assume that thoughts have a "my-ness". Second, some philosophers (e.g., Carruthers and Veillet, 2011; Prinz, 2011) deny that thinking is a conscious activity. These philosophers are likely to interpret thought insertion as a disorder having to do with *inner speech* rather than thought. I think this interpretation is wrong, but I won't insist on it. For my arguments don't hang on this issue.

These alarming details are illustrated by patients' vivid descriptions of their experiences. One patient is reported as saying, "Thoughts are put into my mind like 'Kill God.' It is just like my mind working, but it isn't. They come from this chap, Chris. They are his thoughts" (Frith, 1992, p. 66). And another patient is reported as saying:

> I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his … He treats my mind like a screen and flashes thoughts onto it like you flash a picture (Mellor, 1970, p. 17).

Clearly something is amiss here. But how are we to understand this bizarre phenomenon? Why would a person claim that someone else is thinking her thoughts? What could lead to such dramatic misattributions?

Many details concerning the underlying causes of thought insertion remain unknown. However, psychologists and philosophers who study schizophrenia agree that thought insertion essentially involves a breakdown in one's experience of *ownership* of one's thoughts.[21] It's not that inserted thoughts are experienced as *un*-owned. They're experienced as owned all right—as owned *by someone else*. So the experience of ownership is there.[22] And the thoughts are also experienced. What appears to be missing for those who suffer from thought insertion is the experience of *oneself* as the owner of one's thoughts.

---

[21] See, for example, Gibbs (2000, p. 196), Marcel (2003, p. 80), Sass (2000, p. 154), Radden (1999, p. 351), Stephens and Graham (2000), Campbell (2002), Carruthers (2007, p. 537), and Coliva (2002). Eilan (2000, p. 106-107), Parnas (2000, p. 139), and Sass (2000, p. 157) especially emphasize that thought insertion is a disorder of *experience*. In what follows, I will assume that thought insertion is a real phenomenon, and that the reports of those who suffer from thought insertion reflect genuine disturbances in their experiences.

[22] This is not to say that the feeling of ownership is the same across normal cases and cases of thought insertion. It is only to say that those suffering from inserted thoughts are clearly not inclined, on the basis of their experiences, to describe their thoughts as *un*owned. There is something about their experiences that leads them to attribute their thoughts to *another*. So thought insertion cannot be explained *just* by saying that there is a breakdown in some brute non-individual-specific feeling of ownership. I'll take up this issue again in §2.3.

It's worth stressing that this disruption in the experience of ownership is generally considered to be an experiential *deficit*. Something is *missing* from experience.[23] Part of the rationale for this claim has to do with underlying causal mechanisms that are sometimes associated with schizophrenia.[24] But the main rationale for positing a deficit here derives from the contents of patients' reports and from ties to other experiential deficits found among schizophrenics.[25] I'll return to this issue later (see §2.2). For now my point is just that the consensus among those who study schizophrenia is that thought insertion is to be understood as involving a certain *lack* of self-ownership experiences.

So what does it mean to experience oneself as the owner of one's thoughts? Some nuance is needed here. Lynn Stephens and George Graham (1994, 2000) draw an important distinction between the experience of oneself as the *subject* of one's thoughts and the experience of oneself as the *agent* (or *author*) of one's thoughts. The experience of subjectivity is, according to Stephens and Graham (2000), "[The] sense that something occurs in me, within my ego boundary or psychological history, rather than outside me" (p. 7). To experience oneself as the subject of one's thoughts is to experience oneself as the individual who is experiencing, witnessing, or undergoing one's thoughts. The experience of agency is conceptually distinct. To experience oneself as the agent of one's

---

[23] Sass (2000) describes the symptoms associated with thought insertion as "self withdrawal" (p. 169), and he says that these symptoms, "necessarily imply the … *absence* of something that is normally present—the sense of ownership of intentional control" (p. 154). Blakemore (2000) and Gallagher (2000) say thought insertion involves a "breakdown" or "lack" in experience. Frith (1992) suggests that inserted thoughts are a "disruption" and a "deprivation" that implies that in normal circumstances "we have some way of recognizing our own thoughts" (p. 80). See also Stephens and Graham (2000, ch. 7), Gibbs (2000, p. 196), and Zubin (1985, p. 462). Here I do not deny that inserted thoughts come with additions to experience. My only point is that the breakdown of the sense of ownership in thought insertion is an experiential *absence* or *deficit*.

[24] Frith (1992) and Blakemore (2000), for example, argue that some symptoms in schizophrenia are at least partially caused by a breakdown in a cognitive system responsible for self-monitoring.

[25] For more on experiments revealing some of the related experiential deficits found among schizophrenics, see Frith and Done (1989), Malenka et al. (1982), and Blakemore (2000).

thoughts is to experience oneself as the individual who is bringing about, producing, or *thinking* one's thoughts.

Stephens and Graham (2000) borrow the above distinction from Harry Frankfurt (1988). An example that both Frankfurt (1988, p. 61) and Stephens and Graham (2000, p. 150) use to illustrate this distinction is the difference between recognizing that one's body is moving (subjectivity) when one's reflexes are tested or when one gets bumped into, for example, and recognizing that one is *doing* the moving (agency) when, say, one raises one's arm. That we experience ourselves as the authors/agents of our thoughts in this way does not imply that we really are authors or agents in any very metaphysically loaded sense according to which we are *free* with respect to our thoughts or are in ultimate control of our thoughts. Rather, it just implies that we experience ourselves as bearing a certain active relation to our thoughts that gives us the sense that we are their authors/agents.

According to Stephens and Graham, schizophrenic patients suffering from thought insertion experience themselves as the subjects, but not the agents, of their thoughts. Stephens and Graham (2000) write:

> In the examples of thought insertion discussed in the clinical literature, patients are well aware of the subjectivity of their thoughts: of where they occur. They regard them as occurring within their ego boundaries. The patient quoted by Frith says "Thoughts are put into my mind." Mellor's patient doesn't speak of perceiving thoughts occurring outside her mind; rather, she accuses Eamonn Andrews of putting his thoughts into *her* mind: "He treats my mind like a screen and flashes thoughts onto it like you flash a picture" … The subject regards the thoughts as alien not because she supposes they occur outside her mind, but in spite of her awareness that they occur within her (p. 126-127).

What's missing for these patients, according to Stephens and Graham, is the experience of themselves as the *agents* (or authors) of their thoughts. Thus, Stephens and Graham (2000) interpret thought-insertion patients as follows:

> When she denies that the thought is her thought, she does not contradict the claim that it occurs in her. Rather, she may be interpreted as saying that, although the thought occurs in her, she does not regard herself as its agent or author. She admits to being the subject in whom the thought occurs, but denies that she *thinks* the thought (p. 153).

We can now put a finer point on the idea that thought insertion essentially involves a breakdown in one's experience of ownership. People who suffer from thought insertion fail to experience themselves as the *agents/authors* of their thoughts. As a result, they fail to attribute their own thoughts to themselves. They attribute them to other agents instead. Hence, they deny ownership of their thoughts.[26]

2.2 Thought Insertion and Self-Experience

Now let's return to my argument for SELF-EXPERIENCE. It turns on this premise:

P2 But the self *does* normally show up in experience in the role of author/agent of one's thoughts.

Here is my initial support for P2. In certain *abnormal* cases—i.e., in cases of thought insertion—the self fails to show up in experience in the role of author/agent of

[26] Stephens and Graham's (2000) application of the subject/agent distinction to thought insertion has gained widespread support among philosophers and psychologists who study schizophrenia (see, e.g., Coliva, 2002; Radden, 1999, p. 355; Gallagher, 2000; Bayne, 2004; Kriegel, 2004, p. 189). So I will proceed with this aspect of Stephens and Graham's account. But keep in mind that my arguments do not depend on these particular details. If it turns out that the experience of ownership should be understood as the experience of oneself in some role other than that of agent/author of one's thoughts, then P2 can be amended without affecting my argument.

one's thoughts, resulting in a significant experiential deficit. This suggests that, in *normal* cases, the self shows up in experience in the role of author/agent of one's thoughts.

Think of it this way. The best explanation for the difference between normal experiences and inserted-thought experiences is that the self shows up as the author/agent of one's thoughts in the former but not in the latter. So the best explanation for thought insertion implies that the self normally shows up in experience as the author/agent of one's thoughts. Or think of it this way. Something's missing. It's not the thought. And it's not the sense of ownership, since inserted thoughts are experienced as owned—just by someone else. What's missing is the sense of *oneself* as the agent/author of one's thoughts. And what's missing is missed only because it's normally not missing.[27]

This argument for P2 relies on the claim that something's *missing* from the experiences of those who undergo thought insertion. Again, this claim represents the consensus view among clinicians, scientists, and philosophers who work on the topic (see §2.1). I reiterate this point here in order to forestall the urge to tell a different story—one according to which normal experiences are self-free and inserted thoughts are different because they come with an *extra* experiential component. This story is not supported by what is currently known about thought insertion. And it also makes things much more difficult in terms of explaining or making sense of schizophrenics' thought attributions. If normal thought self-attribution doesn't involve self-experience, then why would an *extra* experiential component lead schizophrenics to attribute their thoughts to others? And if it's schizophrenics' experiences that lead them to say that their thoughts have external *authors/agents*, then why do they still maintain that they are the *subjects* of their

[27] Or as Frank Ocean puts it, "You can't miss what you ain't had" ("There Will Be Tears"). I recognize that Ocean's claim conflicts with Carly Rae Jepsen's confession: "Before you came into my life I missed you so bad" ("Call Me Maybe"). But I'm with Ocean on this one.

thoughts, if not because they experience themselves as the subjects of their thoughts? One could devise a theory according to which we attribute our thoughts in different ways to different individuals in different circumstances for different reasons. But surely the simpler theory, which happens to be favored by our best science, is to be preferred.

Again, the scientific and philosophical consensus is that thought insertion essentially involves a *breakdown* in one's self-ownership experiences, and this fits nicely within a comparatively simple view of thought attribution. Thus, I conclude that in cases of thought insertion the self fails to show up in experience in the role of author/agent of one's thoughts, but in normal cases the self *does* show up in experience in that role. So I conclude that P2 is true. And since P1 is also true, and my argument for SELF-EXPERIENCE is valid, I conclude that SELF-EXPERIENCE is true. The self shows up in experience.

2.3 Another Explanation?

A denier of SELF-EXPERIENCE might respond by coming up with a different explanation for thought insertion. This would be one strategy for explaining the data while still denying that the self shows up in experience. So let's look at some options.

Let's start with Hume. Hume denies that the self shows up in experience, but he doesn't deny that most people are led to believe otherwise. What Hume suggests is that people *infer* the internal presence of a self on the basis of their experience of certain *relations* obtaining between mental states.[28] So maybe Hume would explain thought insertion by claiming that the failure to experience these relations causes the experience of thought insertion, and causes some people to infer that their thoughts are not their own.

[28] Hume (1739/1975) specifically cites the relations of *resemblance* and *causation* (I.iv.6). But later, in the Appendix to *A Treatise of Human Nature*, he expresses doubts as to whether this account succeeds.

But this explanation of thought insertion doesn't cut it. To see why, suppose I experience two mental states, P and Q. And suppose that Q is a thought. On the basis of what experiential relation between P and Q might I infer that Q is *mine* in the sense that I am its author/agent? Do P and Q look alike? Are they about the same thing? Did P appear to cause Q? Maybe. But these experiential relations alone don't in any way support the inference that Q is my thought. After all, P and Q could look alike or be about the same thing and yet both be experienced as inserted. It seems that the only way I could (or indeed would) infer that Q is my thought on the basis of its experiential relations to P is if I already experienced P as mine. Then I could reason like this: P is mine; Q is related to P in the relevant way; therefore, Q is also mine. But if I experience P as mine, then my self shows up in my experience after all. That is, unless I infer that P is mine on the basis of its relation to other mental states. But then it would have to be that I already experience *those* mental states as mine.

Perhaps this problem could be overcome if people normally infer that their thoughts are their own on the basis of their thoughts' relations to their *overall* view of themselves, or to their long-term goals, intentions, and desires, rather than to individual mental states. Perhaps the appearance of inserted thoughts is the result of a perceived inconsistency between the content of a thought and the contents of one's overall self-conception.

This explanation of thought insertion still doesn't cut it. For inconsistency with one's overall self-conception is neither necessary nor sufficient for a thought's being experienced as inserted. It's not *sufficient* for thought insertion because some people—including those suffering from other mental disorders—experience thoughts that are

inconsistent with their overall view of themselves and their attitudes without experiencing those thoughts as inserted. Those suffering from obsessive compulsive thought disorder, for example, often report having thoughts that they find aversive, contrary to their will, and indeed, inconsistent with their overall self-conception (see Stephens and Graham, 2000, 167-168; Fish, 1985, p. 37). Yet these individuals do not experience their thoughts as inserted. Inconsistency with one's overall self-conception isn't *necessary* for thought insertion either. For although some inserted thoughts are odd or out of place (e.g., "Kill God"), other inserted thoughts cohere perfectly with their subjects' overall self-conception (e.g., "The garden looks nice and the grass cool"). Thus, inconsistency with one's overall self-conception is neither necessary nor sufficient for a thought's being experienced as inserted. So inconsistency of this sort doesn't explain thought insertion.

Are there other options? Armstrong (1968) supposes that the content of a thought could seem so *bizarre* to a subject that she experiences it as alien (p. 337). But this doesn't explain thought insertion. For bizarreness is neither necessary nor sufficient for thought insertion. The contents of some inserted thoughts are not bizarre, and the contents of some non-inserted thoughts are bizarre. Another option is Derek Parfit's (1984) suggestion that the only difference between one's own experiences and the experiences of others is that one's experiences happen *here*, in *this* mental life, whereas others' experiences happen *there*, in *those* mental lives (p. 252; n., 36). But nothing along these lines can explain thought insertion. Patients suffering from thought insertion experience their thoughts as *here*, in *this* mental life; and yet, they experience some of their thoughts as inserted. Could it be that inserted thoughts are unique in that they are

experienced as unintended, unwanted, or sinister? No, this won't work either. For some inserted thoughts are experienced as intended, wanted, and innocuous; and some non-inserted thoughts are experienced as unintended, unwanted, and sinister.[29]

Is there some *brute feeling of ownership* that normal thoughts have but inserted thoughts lack? Could this explain thought insertion without implying SELF-EXPERIENCE? It doesn't seem so. For, remember, inserted thoughts are experienced as owned—just by someone else. To be fair, the sense in which schizophrenics experience their thoughts as owned by others might be different from the sense in which we experience our thoughts as owned. After all, it's not as if those suffering from thought insertion actually experience others' brute ownership feelings. But my point is just that an appeal to a brute feeling of ownership could not, by itself, explain why one would attribute one's thoughts to one person rather than another. For it's not enough that a thought is experienced as *owned*, since the thought could be experienced as owned *by someone else*. So it seems that explaining thought insertion in terms of a breakdown in a brute feeling of ownership would, in the end, require understanding thought insertion as a breakdown in *my* (*this* person's) brute feeling of ownership. And then we are back to square one, appealing to self-experience in order to make sense of thought insertion.

I know of no semi-plausible Humean view of thought ownership that fares any better here. Pick any feature of thoughts F that might explain the difference between the experiences of normal thoughts and inserted thoughts, but which doesn't imply SELF-

[29] For a discussion of *unintended* thoughts and actions see Blakemore (2000, p. 188), Gallagher and Marcel (1999, p. 292), Marcel (2003), and Stephens and Graham (2000, p. 141-142); for *unwanted* thoughts see Bleuler (1950, p. 96) and Stephens and Graham (2000, p. 168); for *aversive* or *sinister* thoughts see Snyder (1974, p. 119), Modell (1960), Linn (1977), and Stephens and Graham (2000, p. 169). Stephens and Graham (2000) also discuss and reject the idea that inserted thoughts are unique in that they are experienced as *uncontrolled* (p. 159; see also Fish, 1985, p. 43; Hoffman, 1986, p. 536).

EXPERIENCE. It seems that, for any such F, either some inserted thoughts have F, or some non-inserted thoughts lack F. So there doesn't appear to be any non-self-implicating feature of all and only inserted thoughts that can explain the distinctive experience of thought insertion. So there is no satisfactory explanation of thought insertion that doesn't imply that the self shows up in experience.

3. Further Objections and Concerns

At this point, a denier of SELF-EXPERIENCE might admit that it *seems* like the self shows up in experience. But then she might deny that what shows up in experience *really is* the self. She might say that my self-experience is a misimpression, misconception, or something like an *illusion*—an experience that isn't really of the self.

But I think we know better. We've learned from the data on thought insertion that a *thinker*—a subject and agent/author of thoughts—normally shows up in experience. And *we* think our thoughts. I, MD—the person—am the thinker of my thoughts. Nothing else thinks my thoughts.[30] Those who suffer from inserted thoughts disagree, of course; they believe that others think their thoughts. But they're wrong. I, on the other hand, am right when I self-attribute my thoughts. So the thinker of my thoughts that I experience must be me.

Here we don't need to make any weighty metaphysical assumptions such as that I am a free agent or that I am in ultimate control of my thoughts. The point is just that *I* am the one in question—I am the person who bears the relation to my thoughts that underlies

[30] This point naturally brings to mind "Too Many Thinkers" arguments of the sort discussed by Olson (2007) and Merricks (2001). Whether or not these arguments are sound, I do believe that we should, if at all possible, avoid the conclusion that there is more than one thinker thinking my thoughts right now.

my (successful) thought self-attributions. We can set aside whether I know the precise nature of this relation or its metaphysical implications.

Think of it this way. My experiences give me the sense that my thoughts are *mine*. And that's right. My thoughts *are* mine. They aren't the thoughts of some illusory self-like entity. This suggests that my experiences of myself aren't an illusion, at least not in the sense of being experiences as of something other than me. For if my self-experiences were an illusion—if I experienced something else as my thoughts' owner, agent, subject, thinker, or whatever—then I wouldn't so regularly and so accurately attribute my thoughts to myself. I, like the schizophrenic, would at least sometimes attribute my thoughts to something else. But I don't. So it's got to be me who shows up in my experiences.

Now, a denier of SELF-EXPERIENCE might still deny that the self *itself* shows up in experience. She might admit that each mental state contains a *representation* of the self, but then maintain, with Hume, that mental states are all that can be found in experience. But this denial of SELF-EXPERIENCE is no denial at all. For remember, SELF-EXPERIENCE is neutral as to whether the self *itself* shows up in experience, or whether the self is merely represented in experience (see §1). And, anyway, I doubt that Hume and his followers would accept the above view. Hume unqualifiedly claims that he cannot find himself in his experiences. It would be awfully misleading for him to make this claim without qualification if he in fact found a representation of himself among his experiences.

Here's another way one might attack SELF-EXPERIENCE: Allow that the self shows up in experience when one is *active*—when one is (or at least seems to be) an *agent* of one's thoughts or other actions—but deny that the self shows up in experience otherwise.

Hume won't like this strategy, since he denies that the self *ever* shows up in experience. Nonetheless, this might seem like a promising way to attack SELF-EXPERIENCE.

But it's not. First of all, self-experiences aren't limited to experiences of oneself as an agent. Thought-insertion patients fail to experience themselves as the agents of their thoughts. But they still experience themselves as the *subjects* of their thoughts. So there are non-agential self-experiences. And even if we suppose that we only experience ourselves when we are active, this suits SELF-EXPERIENCE just fine. For, as experiencers, we are normally active. We are rarely if ever *completely* experientially passive. There may be *aspects* of our experiences that we are passive with respect to (e.g., some sensations). But these aspects rarely if ever completely dominate our experiences at any given time. Or, at least, they don't do so often enough to cast doubt on the claim that the (active) self *normally* shows up in experience. So the present strategy is not a promising way to attack SELF-EXPERIENCE.

One final potential objection is that SELF-EXPERIENCE implies a problematic view of self-awareness—one whereby our normal procedure for self-attributing mental states involves our *introspectively identifying* ourselves. Shoemaker (1994), among others, criticizes this view, saying that our basic self-awareness—the kind of self-awareness that governs mental self-attribution—does not involve self-identification. He says this in part because he believes both that identification always goes hand-in-hand with the possibility of *mis*identification, and that we are immune from misidentifying ourselves when we self-attribute mental states. So Shoemaker insists that our mental self-attributions do not rely on introspective self-identification. And so if SELF-EXPERIENCE entails the opposite, SELF-EXPERIENCE might be in trouble.

But SELF-EXPERIENCE does not entail the opposite. I have not offered, and SELF-EXPERIENCE does not entail, any particular view of self-awareness. SELF-EXPERIENCE is not a claim about self-awareness or self-attribution. So SELF-EXPERIENCE is not threatened (at least not directly) by the above concern.

And yet, there is clearly a connection between self-experience, self-awareness, and self-attribution. So it's worth saying something about Shoemaker's argument.[31] One thing one might say is that introspective self-identification is a special case. After all, there's never any other person who shows up in my experience as the subject of my experience. So there's no other person who I experience experiencing things for whom I could mistake myself. Thus, if I judge that I am experiencing something, and my self-identification is based on introspection, then there is no way I could misidentify myself.[32] That's one response. Another response is to grant that in general mental self-attributions do not involve introspective self-identification. Perhaps we normally take self-experience, and the experiential tie between our selves and our mental states, for granted. Perhaps it is simply assumed as a feature of our everyday experiences, and only questioned when significantly altered or impaired. Perhaps one does not need to introspectively attend to and identify oneself in order to have the (experiential) sense that

---

[31] One response that I will not consider here is to say that thought insertion proves the falsity of Shoemaker's claim that we are immune to self-misidentification. This line is suggested by Campbell (2002), but criticized (rightly, I think) by Coliva (2002) and Stephens and Graham (2000). Although those who suffer from thought insertion misidentify themselves as the *agents* of their thoughts, they do not misidentify themselves as the *subjects* of their thoughts, which is what Shoemaker's claim is about.

[32] Shoemaker (1994) anticipates this response with another argument. He says that in order to introspectively match my mental states to myself I would have to first know which mental states are *mine*. But if I already know which mental states are mine, then there is no reason to think that my self-attributions rely on a *further* step of matching those mental states to an individual that I have identified as myself. One response to this argument is to grant that introspective self-identification does not require matching oneself to one's mental states. One might say that it is part of the concept of 'I' that it refers to *this* person who shows up in *this* experience and who experiences things from *this* perspective. So the *single* fact that I identify when I introspect is something like *this-person-experiencing-P* (or *this-person-experienced-P*). This suggestion requires much more development, however.

one's mental states are one's own.[33] This second response grants Shoemaker's point about self-identification without denying SELF-EXPERIENCE.

Each of the above responses deserves further attention. But the point I want to make here is just that there's room to maneuver. SELF-EXPERIENCE can accommodate various views about self-awareness. Thus, this is no reason to shrink from SELF-EXPERIENCE. Indeed, there is no reason at all to shrink from SELF-EXPERIENCE.

4. Conclusion: The Character of Self-Experience

So then why do Hume, Ryle, Armstrong, and a host of other philosophers shrink from SELF-EXPERIENCE? If the self shows up in experience, then why do some honest introspective efforts yield strong convictions to the contrary?

Having never been privileged to observe the minds of Hume, Ryle, and Armstrong, it's difficult for me to say why they deny SELF-EXPERIENCE. But here are some tentative suggestions. First, I think that self-experience is *ubiquitous*. It's always there (at least for most people).[34] And because self-experience is ubiquitous, it's less noticeable. It fades into the background like the continuous hum of an air conditioner. Furthermore, self-experience is always accompanied by other experiences. It's never alone. So perhaps it's more difficult to pick it out from other experiences. Also, self-experience doesn't reveal much about the self. I normally experience myself as the subject and agent of my experiences, but that's about it. So self-experience isn't as

[33] Uriah Kriegel (2004) has a helpful distinction here between *focal* self-awareness and *peripheral* self-awareness. Focal awareness of something (including the self) requires attention, but peripheral awareness does not.

[34] Indeed, even those who suffer from inserted thoughts experience themselves (just not *as* agents of their thoughts).

comprehensive as one might expect. Finally, self-experience isn't very attention grabbing. A sharp pain in my knee is very attention grabbing. The feel of glasses on my face and the faint sound of traffic in the distance are less attention grabbing, and thus, less noticeable. Perhaps self-experience is the same way. It just isn't that attention grabbing.

So, for these reasons, it may be difficult to latch onto self-experience. However, thinking about thought insertion does help. It allows us to get a better grip on self-experience. For it allows us to imagine what it would be like if we didn't experience ourselves as the authors of our own thoughts. Imagining undergoing thought insertion allows us to attend to what it's like for us to experience ourselves as the authors of our thoughts. It allows us to notice ourselves. It allows us to see that SELF-EXPERIENCE is true.

Is there anything more we can say about the character of self-experience? Perhaps we can. What is self-experience like? Well, it's like the experience of the self. It isn't so much like the taste of a ripe banana, the feel of rabbit hair, or the smell of sulfur. It is different from experiences of thoughts, emotions, pains, or tickles. It is different from seeing or hearing other people, and indeed, it is different from other experiences of oneself.

And yet, there is a certain parallel between the kind of self-experience that I am talking about and other experiences of oneself. Suppose someone asks me: What is it like to see yourself in the mirror? I might say: Well, it depends. It depends on what I am wearing, for instance. If I am wearing a red shirt then I'll look one way, if I am wearing a blue shirt then I'll look another way; a hat, glasses, or a jacket would make a difference too. And what it's like to see myself in the mirror also depends on how I am standing (or

sitting), whether I am directly facing the mirror or looking at it from an angle, and so on. In fact, what it's like to see myself in the mirror may vary significantly from case to case.

So what is self-experience like? Well, it depends. It depends on what I am thinking, for instance. If I am thinking about philosophy my self-experience will be one way. If I am thinking about what to eat for lunch it will be another way. What I am seeing, feeling, hearing, smelling, and tasting make a difference too. Here I am not suggesting that self-experience is reducible to various tastes, thoughts, feelings, etc. Nor am I suggesting that there is no aspect of self-experience that is constant over time. My only point is that, just as I can be manifest to myself in many different ways when I look into a mirror, so too can I be manifest to myself in many different ways in experience. Maybe Hume was right when he said that he couldn't catch himself without an experience. But that's only because whenever one experiences oneself one experiences oneself experiencing other things.

CHAPTER 2

We Are Acquainted With Ourselves

In the last chapter I argued that SELF-EXPERIENCE can accommodate a plausible account of self-awareness. But I neither gave nor defended any such account. So in this chapter I will take the next step. I will describe and defend a view that can serve as the foundation for an intuitive and explanatorily powerful account of self-awareness. What I will argue is this: We can be, and often are, *acquainted with* ourselves. First, I will describe in detail what it means to be acquainted with oneself. Then I will introduce, develop, and defend a commonly used test for acquaintance. Finally, I will apply this test to us and show that we pass. Thus, I will conclude that we can be, and often are, acquainted with ourselves.

1. Self-Acquaintance

Again, here's the claim that I will defend:

SELF-ACQUAINTANCE: People can be (and often are) acquainted with themselves.

Let me explain. First, my understanding of *acquaintance* comes from Bertrand Russell (1912). He characterizes the notion of acquaintance as follows:

> We shall say that we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths (p. 73).

One is acquainted with something if and only if one is *directly aware* of it. Consider an example. Suppose I discover that it's raining outside by looking at a weather report. I am aware of the rain, but only indirectly. For I am only aware of the rain in virtue of being aware of something else—namely, the weather report. Now suppose that I go outside, see the rain, hear it pattering on the ground, and feel its droplets on my skin. Am I now directly aware of the rain? Not according to Russell (1912).[35] For, on his view, I am aware of the rain only in virtue of being aware of my *experiences* of the rain.[36] My rain-experiences are really what I am acquainted with. And this seems right. It seems right that I can be aware of my visual experience of the rain, for example, without there being anything that mediates my awareness of that experience. It seems that I can be aware of the experience *itself*.

The notion of 'directness' in play here has both *metaphysical* and *epistemic* elements (cf., Gertler, 2012). There is a metaphysical element in that the relation of awareness I bear to an object of acquaintance is not mediated by any distinct entity or causal process. This differs from *in*direct relations such as the one I bear to the rain when I look at a weather report. The rain causes someone to note that it's raining, which causes there to be a weather report, which then causes me to be aware of the rain. This causal process involving various entities mediates my awareness of the rain. But no such process mediates my awareness of my *experiences* of the rain. My experiences are of course

[35] Direct realists, such as McDowell (1994), Martin (2002), and Hinton (1973), might disagree here because they claim that we can be directly aware of external objects. But my point here is not to dismiss direct realism. So direct realists can just ignore this part of my exposition.

[36] Russell thinks that it's *sense data* that we are acquainted with. Here I use 'experience' as a neutral term that could, depending on one's view, refer to phenomenal properties, experiential events, sense data, etc. For detailed accounts of the acquaintance theory as applied to experiences, see Gertler (2001, 2011, 2012), Chalmers (2003), Feldman (2004), and Fumerton (1995).

caused to exist. But once they exist, no extra causal process is needed to establish my relation to them as something of which I am aware.

There is an *epistemic* directness here as well. If one is acquainted with some x, then one can form beliefs about x that are based solely on one's awareness of x *itself*. And so one's justification for those beliefs may, at least in some cases, derive solely from the nature of one's awareness of x. For example, I can form justified beliefs about my experiences of the rain just by attending to those experiences. There is nothing else that I must attend to in order to form justified beliefs about those experiences. So my knowledge of my rain experiences can be direct in this way.[37] In contrast, when I form a belief about the rain by attending to a weather report, the justification for my belief will derive, at least in part, from the epistemic credentials of a source of information other than the rain itself. Hence, I do not have direct epistemic access to the rain in the same way that I have direct epistemic access to my rain experiences.

One implication of this epistemic directness is that beliefs formed on the basis of acquaintance will typically enjoy a relatively high level of justification. If a belief about x is based solely on one's direct awareness of x, then this belief will usually be more strongly justified than a belief about x that is based on other things besides x itself. For example, if I believe that I am in pain because I am directly aware of my pain, then this belief will be more strongly justified than it would have been if it were based on my looking at a brain scan or observing my own behavior. I will return to some of these

[37] Russell (1912) thinks that acquaintance with something *automatically* yields knowledge of that thing (p. 73). Most contemporary acquaintance theorists disagree; they think that one must also conceptualize and think about objects of acquaintance in order to know things about them (see Chalmers, 2003; Gertler, 2011, p. 49). I won't insist on a side here. If one's being acquainted with something doesn't automatically yield knowledge of that thing, then it at least puts one *in a position* to gain knowledge about that thing.

issues in the next section. For now I just want to point out that beliefs formed on the basis of acquaintance appear to have a relatively high degree of epistemic justification.

Thus far I have focused on acquaintance with *experiences*, since they are generally considered to be the most plausible candidates for things we can be acquainted with. However, what I am really interested in here is *self*-acquaintance. So let's return to that topic. Here's what Russell (1912) says about it:

> When a case of acquaintance is one with which the person is acquainted … it is plain that the person acquainted is myself. Thus, when I am acquainted with my seeing the sun, the whole fact with which I am acquainted is 'Self-acquainted-with-sense-datum'. Further, we know the truth 'I am acquainted with this sense-datum'. It is hard to see how we could know this truth, or even understand what is meant by it, unless we were acquainted with something we call 'I'. It does not seem necessary to suppose that we are acquainted with a more or less permanent person, the same today as yesterday, but it does seem as though we must be acquainted with that thing, whatever its nature, which sees the sun and has acquaintance with sense-data. Thus, in some sense it would seem we must be acquainted with our Selves as opposed to our particular experiences (p. 79-80).

Russell (1912) likes the idea that we are acquainted with ourselves. But he doesn't say much else about self-acquaintance. And he also admits that the issue is "difficult" and that "although acquaintance with ourselves seems *probably* to occur, it is not wise to assert that it undoubtedly does occur" (p. 80). Perhaps this is good advice. For few have found the notion of self-acquaintance very appealing.[38]

Nonetheless, I assert that self-acquaintance occurs. And now I can say more about what I am asserting. When I say that people can be (and often are) acquainted with themselves, I mean that they can be *directly aware* of themselves in Russell's sense. I

[38] Howell (2006), for example, says (without argument as if it's obvious), "There is no acquaintance with the self or with any sort of conceptual/representational stand-in for the self" (p. 45-46; see also, Grice, 2008, p. 79-80; Hume, 1739/1975, IV.6; Ryle, 1949/2002; Armstrong, 1968). But perhaps even more notable than quotes like this is the absence of fully developed acquaintances theories of self-awareness in the literature.

mean that people can be aware of themselves, and not in virtue of being aware of anything else, such as thoughts, behavior, testimony, descriptions, brain scans, etc. I mean that one can be non-inferentially aware of one's self. And so one can, if one wants, refer to oneself directly, without relying on any description or causal chain, as '*this* person' or by using other rigid designators like 'me' or 'I'. And one can come to *know* certain things about oneself—e.g., that one exists, that one is in pain, that one is thinking that P, etc.—by forming beliefs about oneself on the basis of self-acquaintance.

Given what I have said here and in the previous chapter, SELF-ACQUAINTANCE is most naturally understood as an *introspective* thesis about self-awareness. The idea is that one is acquainted with oneself in roughly the same way that one is acquainted with one's experiences. However, strictly speaking, this isn't implied by SELF-ACQUAINTANCE. And some may wish to go a different direction.[39] So I'll just say that I consider self-acquaintance to be an introspective self-awareness. But what's essential for the purposes of this chapter is just that we can be and often are directly aware of ourselves.

This puts SELF-ACQUAINTANCE at odds with various accounts of self-awareness, including those that say we can be introspectively aware of ourselves only *indirectly*—in virtue of being aware of our mental states (see, e.g., Howell, 2006).[40] I of course don't deny that we *can* be aware of ourselves indirectly. After all, I can be aware of myself by

[39] For instance, some philosophers (e.g., Shoemaker) have argued that our most *basic* form of self-awareness couldn't rely on introspective self-identification. I've not given an account of basic self-awareness here. But still, if one wanted to develop SELF-ACQUAINTANCE into an introspective account of basic self-awareness one would have to deal with the aforementioned concerns. For what it's worth, I believe that one could succeed in doing this.

[40] Some philosophers might deny that this sort of self-awareness is indirect. Chisholm (1969), for example, thinks of mental states as modes of subjects, and so he would say that if one is directly aware of one's mental states then one is directly aware of oneself. This view might technically satisfy SELF-ACQUAINTANCE; however, here I am treating self-acquaintance as being something over and above acquaintance with one's mental states, since, as we know from the previous chapter, one's experience of one's self is something over and above one's other experiences.

looking in a mirror, wiggling my toes, or indeed, by attending to my mental states. But what SELF-ACQUAINTANCE denies is that self-awareness is *always* achieved via intermediaries.

So SELF-ACQUAINTANCE is at odds with various accounts of self-awareness. But it is almost completely neutral with respect to *ontological* accounts of the self. We should not assume that our direct awareness of ourselves yields some profound ontological insight into our underlying nature. I do assume that selves exist, and I do talk about selves as the *subjects* of their experiences.[41] But that's as far as I will go here.

Finally, I suggest that self-acquaintance can occur without any great act of attention. Acquaintance with the self is like acquaintance with experiences. We can be acquainted with our experiences without focusing on or consciously attending to them. And the same goes for the self. Most of the time we are simply aware of ourselves in the same effortless way that we are aware of our pains, itches, smells, and tastes.[42] But even so, SELF-ACQUAINTANCE doesn't say that we are *always* acquainted with ourselves. Sleep, injury, coma, drugs, and (maybe) mental disorder can prevent self-acquaintance. Perhaps even a powerful distraction could do it. So SELF-ACQUAINTANCE just says that we *can be* acquainted with ourselves, and that *often*—in many normal, waking circumstances—we are self-acquainted. That's my claim.

2. The Doubt Test

[41] In §4 I will *argue* that we are subjects of experience and that the bundle theory of selves is false.
[42] Uriah Kriegel (2004) draws a helpful distinction between *focal* self-awareness and *peripheral* self-awareness. Focal awareness of something (including the self) requires attention, but peripheral awareness does not.

At this point I'll assume that we can be acquainted with at least some of our experiences. Our goal here is to determine whether our acquaintance with things extends *beyond* experiences to the self. So we need a way of doing that.

Russell (1912) is again our help in time of need. He offers a test for determining whether one is acquainted with something. He introduces this test by considering a table sitting in front of him. Russell says that although he is directly aware of his experiences that "make up the appearance" of this table, his awareness of the table itself is a different story:

> My knowledge of the table as a physical object, on the contrary, is not direct knowledge. Such as it is, it is obtained through acquaintance with the sense-data that make up the appearance of the table. We have seen that it is possible, without absurdity, to doubt whether there is a table at all, whereas it is not possible to doubt the sense-data (p. 74).

Here Russell seems to be suggesting that one can determine whether one is acquainted with something by considering whether one can (without absurdity) doubt that it exists. Russell can doubt that his table exists—perhaps by entertaining some skeptical scenario. So he says that he isn't acquainted with his table. But Russell cannot doubt that his *experiences* of the table exist. So he concludes that he is acquainted with those experiences. This suggests the following test: One is acquainted with something that one seems to be aware of just in case one cannot doubt that it exists. Hence, Russell is acquainted with his experiences of his table, but not the table itself. We can run this test on my earlier example as well: I can doubt that the rain exists by entertaining some skeptical hypothesis. I can suppose that an evil demon is tricking me about the rain. But I cannot doubt that my *experiences* of the rain exist. Even an evil demon couldn't trick me

about *that*. So Russell's test—what I will call 'The Doubt Test'—delivers the results that I am acquainted with my rain-experiences, but not the rain.

Russell doesn't explain why The Doubt Test is a good test for acquaintance. But the reason is fairly straightforward. If I am aware of some x, but only in virtue of being aware of some distinct y that indicates x's existence, then I can doubt that y is a faithful witness to the existence of x. If a weather report says it's raining, then I can doubt that there is rain because I can doubt that the weather report got it right. Or if I see a tree in the distance, then I can doubt that the tree exists by noting that my visual experience is blurry or even by supposing that an evil demon is deceiving me. However, if I am *directly* aware of x *itself*, then I cannot doubt that x exists, for there is no potentially false or misleading presentation of x that might allow me to doubt its existence. If I am directly aware of a sharp pain, for instance, then although I can doubt certain things about the pain—e.g., how it was caused—I cannot doubt that the pain exists.[43]

The Doubt Test has a distinguished history going back even before Russell at least as far as Descartes.[44] Many philosophers have used the test. Even in the last century it has been used rather often. For example, H. H. Price (1932) uses The Doubt Test while looking at a tomato (p. 3). Price says that when he looks at a tomato there are many things about the tomato (including that it exists) that he can doubt. But he says that he

---

[43] There may be cases where I *misclassify* a pain as a non-pain. And then I may doubt that what I am experiencing (i.e., pain) is pain. But this is a separate issue. We might put the above point more precisely by saying that if I am directly aware of an experience (setting aside how I would classify it), I cannot doubt that *it* exists.

[44] Fumerton (2005), who is among those who use The Doubt Test to support an acquaintance theory, speaks to the history of the test as well as to its connection to Descartes, saying, "... many classical foundationalists sought to identify the objects of direct acquaintance by stripping from experience all that is clearly not before consciousness. One does this through something resembling a Cartesian method of doubt" (p. 123; see also Gertler, 2011, Ch. 4). These foundationalists include Descartes (1641/1993), Malebranche (1674/1997), and, more recently, Russell (1912), Price (1932), Lewis (1946), Chisholm (1957, Ch. 5), and Bonjour (1999).

cannot doubt that his *experiences* of the tomato exist and are the way they seem to be. These reflections then lead Price to conclude that his experiences of the tomato are "directly present" or "given" to his consciousness (ibid.).

> C. I. Lewis (1946) appeals to The Doubt Test in a more general form. He says: Subtract in what we say that we see, or hear, or otherwise learn from direct experience, all that conceivably could be mistaken; the remainder is the given content of the experience inducing this belief (p. 182).

More recently, Brie Gertler (2012) carries on the tradition of appealing to The Doubt Test. She asks her readers to pinch themselves and then to carefully attend to their experiences. She then says:

> When I try this, I find it nearly impossible to doubt that my experience has a certain phenomenal quality—the phenomenal quality it epistemically seems to me to have, when I focus my attention on the experience. Since this is so difficult to doubt, my grasp of the phenomenal property seems not to derive from background assumptions which I could suspend; e.g., that the experience is caused by an action of pinching. It seems to derive entirely from the experience itself. If that is correct, my judgment registering the relevant aspect of how things epistemically seem to me (this phenomenal property is instantiated) is directly tied to the phenomenal reality that is its truthmaker (p. 104-105).

Here we see a more modest version of The Doubt Test. Gertler (2012) somewhat cautiously says that she finds it "nearly impossible" to doubt that a certain quality is instantiated in her experience. She then uses this result to support her claim that she is acquainted with that aspect of her experience.[45]

David Chalmers (1996, Ch. 5) takes a similar line. He argues that acquaintance is required for certainty, where by 'certainty' he means one's ability to remove all doubt,

[45] Gertler (2012) explicitly acknowledges that she is appealing to Russell's "doubt test" (p. 94, 104). Other contemporary acquaintance theorists who appeal to The Doubt Test, or something very near to it, include Fumerton (2005), Feldman (2004) and Chalmers (1996, Ch. 5; 2003). And others who appeal to something like The Doubt Test, but not necessarily in defense of an acquaintance theory, include Hamilton (1860, XV, p. 188), Brentano (1973), Ayer (1956), Ewing (1980, Ch. 5), Malcolm (1975), Alston (1971), Bonjour (1999), Chisholm (1957, Ch. 5), Anscombe (1994), and O'Brien (2007, Ch. 2).

including doubts generated by skeptical scenarios. Chalmers then claims that we can at least sometimes be certain of our experiences, and so he concludes that sometimes we can be acquainted with our experiences. Here once again there's an inference from lack of doubt to acquaintance, which is the inference sanctioned by The Doubt Test.

The Doubt Test also crops up in less obvious forms. For instance, Katalin Balog (2012) argues that her acquaintance theory can explain why it seems to us that some judgments about our experiences can't be *overridden* (p. 23). One plausible way to interpret Balog is as saying that the reason we sometimes cannot doubt that we are experiencing what we are experiencing is because we are acquainted with our experiences. Thus, Balog appears to be offering the indubitability of our experiential beliefs as evidence for our acquaintance with our experiences. Again, The Doubt Test appears to be at work.[46]

Clearly The Doubt Test is, and has been, a commonly used test for acquaintance. But what isn't so clear is how exactly to understand The Doubt Test. In fact, it's even unclear whether all those who appeal to The Doubt Test have the same exact version of the test in mind. So some clarification is in order. In the remainder of this section I will try to explicate the core features of The Doubt Test. Although I cannot speak for everyone, I hope to capture the spirit of The Doubt Test in its most plausible form.

To start with, The Doubt Test is a *first-personal* test for acquaintance. I can use it to test whether *I* am acquainted with something. And you can use it to test whether *you*

[46] Arguments along these lines are everywhere in the literature on acquaintance and "the given" (see, e.g., Chisholm (1957, Ch. 5), Bonjour (1985), Langsam (2002), Gertler (2011, 2012), Fumerton (2005), Feldman (2004), and Chalmers (1996, Ch. 5; 2003)). So if I am right to suggest that this line of reasoning is connected to The Doubt Test, then use of The Doubt Test is extremely widespread.

are acquainted with something. But we cannot use it on each other. I cannot use it to determine whether you are acquainted with a pain, for instance.

The Doubt Test is also limited in that it is exclusively concerned with one's ability to doubt the existence of something that one is, or at least seems to be, *aware of*. One may be unable to doubt other kinds of things, like the truth of a fact such as 2+2=4. But that's different from being unable to doubt the existence of a certain something one seems to be aware of, which is what The Doubt Test is all about.

As for what it means to *doubt* something, those who use The Doubt Test are typically interested in certain *psychological* and/or *epistemic* dimensions of doubt. If one cannot doubt *psychologically* that some x exists, then one is maximally convinced that x exists and so will be unwilling to give up the belief that x exists. This is the dimension of doubt that Gertler (2012) seems to appeal to when she says that it's nearly impossible for her to doubt that her pain exists. No matter how hard she tries to doubt her pain, she can't do it. Her confidence is unshakable. This is one plausible way to think about the kind of doubt that bears on The Doubt Test. But we must exercise caution here. Since different people have different temperaments, we should not expect universal agreement as to what can be doubted psychologically. Some people may be able to doubt that their experiences exist. But many of the rest of us find this idea ludicrous. We could just ignore the outliers. Or we could be more modest and just say that we each have to decide for ourselves whether we can doubt psychologically that some x exists. This second option strikes me as the wisest course for now.[47] If on the basis of your seeming awareness of

[47] The psychological component of The Doubt Test often shows up in this way in the literature. Philosophers like Gertler (2012), Price (1932), and Lewis (1946) invite their readers to consider certain cases and then leave it up to them to decide whether to go along with what is then concluded (perhaps with the expectation that they will).

some x you find it impossible, or perhaps nearly impossible (if you have an especially cautious temperament), to bring yourself to doubt that x exists, then, by your lights, x passes what we might call 'The *Psychological* Doubt Test'. This may not yield perfect agreement. But it's one plausible way to understand the kind of doubt involved with The Doubt Test.

Another way to understand this doubt is as being *epistemic* in nature. Here the idea is that if one cannot doubt (epistemically) that x exists, then one's belief that x exists has a very high level—perhaps the highest level—of epistemic justification. Some might say that the belief in question is *infallible*. However, even among acquaintance theorists who endorse The Doubt Test there is much disagreement as to whether any of our beliefs are ever infallible. In fact, there is even disagreement as to what it would take for a belief to be infallible (see Reed, 2001). We should avoid these controversies if we can. So perhaps a better way to understand epistemic doubt is in terms of *incorrigibility*. Following Alston (1971) and Reed (2001), a belief is incorrigible just in case its epistemic justification cannot be overturned. This more closely approximates what I take to be the epistemic dimension of doubt. Yet, there is again a question of how best to understand incorrigibility. And spelling out what it takes to *overturn* epistemic justification may prove tricky. So, again, we should avoid these controversies if we can.

And we can. We can think of the epistemic dimension of doubt as *immunity to skepticism* (cf., Chalmers, 1996; 2003). This offers a particularly clear way to get a grip on epistemic doubt, and it fits well with how The Doubt Test has historically been used. The idea is this: If on the basis of one's seeming awareness of x one can rule out all skeptical scenarios in which x does not exist, then x passes what we might call 'The

Epistemic Doubt Test'. There is a very intuitive understanding of this idea that has been present throughout our discussion. We've seen that Russell's table, Price's tomato, and the rain in front of me are susceptible to skeptical scenarios, perhaps involving evil demons or hallucinations; but each of our *experiences* aren't susceptible to such scenarios. Consider the rain example. Right now it seems to me that I am aware of both the rain and my experiences of the rain. But, given the way things seem to me, I cannot rule out all skeptical scenarios in which the rain does not exist. I could be a brain in a vat. Then things might seem to me as they do even though it's not actually raining. On the other hand, I *can* rule out all skeptical scenarios in which my *experiences* of the rain don't exist. For no evil demon or computer could make it *seem* to me like my rain-experiences exist, as it now does, without also bringing it about that my rain-experiences *do*, in fact, exist. After all, those very experiences constitute part of the way things seem to me.[48] Thus, given the way things seem to me, I can rule out all skeptical scenarios in which my rain-experiences don't exist. Therefore, my rain-experiences pass The Epistemic Doubt Test, but the rain does not.

This can be captured more formally through appeal to a framework in which epistemic possibilities are conceived as centered epistemically possible worlds (considered as actual). Take a world w, considered as actual, centered on an individual and a time. We can use the rain example and suppose that w is centered on me at t, where t is a time at which it seems to me to be raining. Now consider the class of worlds

[48] Here's how Chalmers (1996) puts it: "In the case of perceptual knowledge, for example, one can construct a case in which the reliable connection is absent—a case where the subject is a brain in the vat, say—and everything will still seem the same to the subject. Nothing about the subject's core epistemic situation rules this scenario out. But in the case of consciousness, one cannot construct these skeptical hypotheses. Our core epistemic situation already includes our conscious experience. There is no situation in which everything seems just the same to us but in which we are not conscious, as our conscious experience is (at least partly) constitutive of the way things seem" (p. 195).

centered on me at t that are epistemically possible *given how it seems to me at t*. These are the worlds that, given my epistemic situation at t, I cannot rule out *a priori* as being w (assuming that I am an ideal rational agent). Here we should not think of my "epistemic situation" as being constituted by what I *know* at t. What I know may depend on factors independent of me that I cannot detect *a priori*. So what I know at t will differ from world to world. We should also be careful not to assume that my epistemic situation is constituted just by the content of my experiences, lest we beg any important questions. Perhaps for now the best way to understand my epistemic situation is as being constituted by the information that is available to me from my perspective, setting aside any empirical background knowledge that I may have about the world. With that said, consider two beliefs: (R) The rain exists, and (E) My experiences of the rain exist. Among the worlds that are possible given my epistemic situation, there will be worlds in which R is true and worlds in which R is false. For there will be worlds in which my epistemic situation is what it is and the rain really exists, and worlds in which my epistemic situation is what it is but the rain does *not* exist (because I am a brain in a vat, say). E, on the other hand, will be true in every epistemically possible world that centers on me at t. For every world in which my epistemic situation is what it is—that is, every world in which the information available to me from my perspective is the same—will be a world in which I am undergoing rain-experiences. Thus, we might say that, given my epistemic situation, E is epistemically necessary, but R is epistemically contingent. Or, to put it another way: E is certain, but R is not. Or, to put it in a way that best fits with our current discussion: E cannot be doubted, but R can. This is just a sketch.[49] But it should

[49] This sketch is an extension of the two-dimensional framework developed by Chalmers (2006) and Jackson (1998).

suffice to show that understanding the epistemic dimension of doubt in terms of immunity to skepticism can be developed within a useful formal framework.

Let's recap. The Doubt Test is a first-personal test for acquaintance that centers on one's ability to doubt the existence of something that one seems to be aware of. This test has two versions. The first version—The Psychological Doubt Test—concerns the degree to which one is convinced of something's existence. The second version of the test—The Epistemic Doubt Test—can be understood in terms of immunity to skepticism. Now, given that we now have two versions of The Doubt Test in play, we have some options. We could treat each version of the test as an independent test for acquaintance. Then we might say that if some x passes *either* version of the test, then one has reason to believe that one is acquainted with x. Alternatively, we could combine the two versions of the test and say that if x passes *both* versions of the test, then one is acquainted with x. Or, if one version of the test is deemed superior to the other, then we may just focus on that version. There is perhaps something to be said for each of these options. But I want to play it safe. So when I appeal to The Doubt Test in what follows, I will explicitly consider each version of the test, and I will use 'The Doubt Test' to refer to this relatively demanding conjunctive version of the test:

> *The Doubt Test*: If it seems to S that she is aware of some x, and on the basis of that seeming awareness S cannot doubt (psychologically) that x exists, and S can rule out all skeptical scenarios in which x does not exist, then S is acquainted with x.

So although I believe that *each* version of the test gives some indication of whether one is acquainted with some x, I will say that x has to pass *both* versions of the test in order to pass The Doubt Test.[50] This yields what seems to be a good, safe test for acquaintance. It doesn't admit of any obvious exceptions, and it is backed by the plausible justification given above. Again, the idea is that if I am aware of some x, but only in virtue of being aware of some distinct y that indicates x's existence, then I can doubt (both psychologically and epistemically) that y is a faithful witness to the existence of x; however, if I am directly aware of x *itself*, then I cannot doubt (either psychologically or epistemically) that x exists, because there is no potentially misleading presentation of x that might allow me to doubt its existence.[51] In addition to this justification, The Doubt Test also has a certain intuitive appeal. This is evidenced by the fact that a wide variety of philosophers over the past few centuries have freely and confidently used the test. Thus, although I will consider some objections to The Doubt Test in the next section, I assume that it has enough credibility to warrant our moving on.

3. The Doubt Test and Self-Acquaintance

[50] Again, this is not the only plausible approach to The Doubt Test. If you prefer one version of the test to the other, then feel free to understand The Doubt Test strictly along those lines. My arguments will be unaffected.

[51] Not everyone who uses The Doubt Test explains the rationale for the test. But those who do typically give the rationale just described. Gertler (2012), for example, points out that her awareness of the table before her is mediated by a causal process, and then she says, "the presence of this mediating factor enables me to doubt the existence of the table, since I can recognize that, for all I know, my visual experience has an aberrant cause" (p. 94). However, this is not the case, according to Gertler, when she is confronted with the phenomenal reality of her pain experiences—she cannot doubt the existence of these experiences because she is directly aware of them (p. 104). Also see Price (1932, p. 7-12) and Chalmers (1996, p. 195-196; 2003, p. 34).

Let's return to our main topic: *self*-acquaintance. We are now in a position to apply The Doubt Test to ourselves. We know from the previous chapter that the self is something that we can be aware of in experience. So we are fit to be tested. Now the question is: Do we pass? I will argue that we do indeed pass The Doubt Test. Thus, I will argue that SELF-ACQUAINTANCE is true—we can be acquainted with ourselves.

Here's my argument:

1. If we pass The Doubt Test, then we are acquainted with ourselves.
2. We pass The Doubt Test.
3. Therefore, SELF-ACQUAINTANCE is true.

This argument is valid. So let's consider the truth of its premises, starting with (2). I can report that *I*, at least, cannot doubt that I exist. I am maximally convinced that I exist—I couldn't be convinced otherwise—and my belief that I exist is immune to skepticism. So I pass The Doubt Test. And since I am not an exceptional case, I assume that you and everyone else pass The Doubt Test too. So (2) appears to be true.

At first this seems like an obvious result. Indeed, few philosophers (and no rational folk untainted by the wiles of professional philosophy) have professed an ability to doubt their own existence (at least in print).[52] But, to be fair, some philosophers may find themselves unable to doubt their own existence only because they recognize that doing so would be absurd for purely logical or conceptual reasons. The way Descartes (1641/1993) puts it is that one has to exist in order to doubt something. So, necessarily, if

[52] Hume (1739/1975), Unger (1979), and Sider (2013) are philosophers who have doubted their own existences. Billon (2014) argues that the self-doubt expressed by those who suffer from Cotard's Delusion or Depersonalization Disorder may also be rational. But I will not take up that debate here.

one doubts one's own existence, then one exists. Hence, it's absurd to doubt one's own existence. Another way of getting at the point is to say that 'I exist' is *self-verifying*. By rule, 'I' refers to the person who thinks (or says) it. So whenever 'I exist' is thought, it must be true, since its thinker has to exist in order to think it. Thus, there is something absurd about the thought 'I do not exist'. These points may make (2) seem less obvious. For if it's not our *awareness* of ourselves that renders us unable to doubt our own existence, then it is inappropriate to appeal to The Doubt Test here. One has to be aware of x in order for x to pass The Doubt Test. So if one's inability to doubt one's own existence has nothing to do with self-awareness, then (2) remains unaddressed.

What we need to do here is set aside the logical barriers to self-doubt just mentioned and ask ourselves whether, *on the basis of self-awareness*, we pass The Doubt Test. We know from the previous chapter that the self shows up in experience. So I, for example, can attend to myself and ask whether I can doubt that *this* person in my experience exists.[53] Perhaps it is misleading to talk only of my *inability to doubt* my existence here. For this puts the point in purely negative terms (as a *lack* of an ability), when in fact a positive cognitive achievement is at stake. We might instead put the question thus: Am I *certain* of my existence? And this is quite separate from the question of whether it would be absurd, for logical reasons, to assent to 'I do not exist'. I might know that it would be absurd for anyone to say or think 'I do not exist' without ever actually considering my own existence. So I could think it absurd on purely logical grounds to deny that I exist without actually being *certain* that I exist (or indeed, without

[53] And even if you were not convinced by my arguments in the previous chapter, it's worth noting that most philosophers accept that we have a certain form of self-awareness—a form of awareness that is present in cases where I think 'I exist'—that would make us fit to be tested by The Doubt Test. For particularly helpful discussions on this topic, see O'Brien (2007), Billon (2014), and Howell (2006).

believing anything at all about myself). This shows that what's at stake here goes beyond my recognition of the logical peculiarities of doubting my own existence. And that's where The Doubt Test comes in.

So let's look more carefully at (2), beginning with The Psychological Doubt Test. The best way that I know of to assess this version of the test is to engage in self-reflection. Here I follow the lead of Descartes (1641/1993). Descartes sits down in a comfortable place next to his fireplace, frees himself of distractions, and then asks himself what he can and cannot doubt. Eventually Descartes comes around to thinking about his own existence. First he notes that he is unable to doubt his own existence for the logical reasons mentioned above. Then he goes further, saying, "I am; I exist—this is certain … it is obvious that it is I who doubt, I who understand, and I who will, that there is nothing by which it could be explained more clearly" (*Med*., II, 27-29). When he introspects, Descartes discovers himself thinking. And on the basis of this discovery, he finds himself unable to doubt (psychologically, at least) his own existence. Thus, by his own lights, Descartes passes The Psychological Doubt Test.

Now you try. Pause for a moment, close your eyes, attempt to clear your mind of all distractions, and ask yourself: Can I doubt that I exist? Given the way things seem to me right now, can I bring myself to doubt that *this* person exists? Or, to put it another way: Can I doubt that *these* experiences—*these* thoughts, sensations, emotions, and feelings—are *my* experiences? Can I doubt that I am their subject?

I, like Descartes, find that I cannot doubt my own existence. I just can't do it, at least not in good faith. And I take it, though I don't know for sure, that most other people agree. A few people apparently disagree (see fn. 52). And I find this incredible, much like

I find doubting one's own experiences incredible. However, I will simply reiterate that we each have to decide for ourselves what we can doubt (psychologically). So I will leave it up to you to decide whether you pass The Psychological Doubt Test.

Let's move on to The Epistemic Doubt Test. Here the question is: Is my belief that I exist, or your belief that you exist, immune to skepticism? And the answer is: Yes, it is. An evil demon could be tricking me about all sorts of things, but it couldn't be tricking me about my own existence. One barrier to skepticism about my own existence is, again, purely logical. An evil demon couldn't trick *me* into falsely believing that I exist, since I have to exist in order to be tricked. But, again, this isn't relevant for our purposes. We need to bracket this purely logical barrier to skepticism. We need to ask whether, on the basis of my self-awareness, I can rule out all skeptical scenarios in which I do not exist. And the answer is still: Yes. There is no skeptical scenario in which things seem as they do now (to me), but in which I am absent. One might be tempted to think that there are skeptical scenarios in which I am mistaken—that is, scenarios in which it seems as if I exist and am experiencing various things, but in which I do not actually exist. But a little reflection eases this temptation. For a little reflection reveals the absurdity in the idea that I might not be the one who is occupying this particular mental perspective that I seem to be occupying right now. The issue isn't with the idea of another person (or no one at all) undergoing experiences that are qualitatively similar (or even identical) to my experiences. The issue is with the idea that someone else (or no one at all) might turn out to be the one occupying *this* particular perspective and undergoing *these* particular experiences that I seem to be undergoing right now.[54] That's what seems

[54] This point is captured nicely by the phrase 'putting oneself in another's shoes'. When one puts oneself in the shoes of another, one tries to imagine undergoing the same (qualitatively) experiences as someone else;

absurd. We might put the point by saying that I am *built into* my perspective (hence 'the first-person perspective'), or that my perspective *depends on* me.[55] Or we might put it in terms of SELF-EXPERIENCE, saying that since my self shows up in my experiences, I am part of the way things seem from my perspective. The crucial point is this: I am inseparable from my perspective and the way things seem from my perspective. So there is no skeptical scenario in which things seem as they do and yet in which I do not exist.

Consider the point from a slightly different angle. Suppose that, right now, I have a sharp pain in my knee. This pain is part of the way things seem to me right now. But it's not just that *pain* is part of the way things seem to me right now. Rather, it's *my* pain—it seems to me that *I* am the one hurting. And, as many philosophers have been careful to point out, if it seems to me that I am the one in pain, then I *am* the one in pain.[56] Even an evil demon couldn't trick me about *that*. Thus, given the way things seem to me right now, I can rule out any skeptical scenario in which I am not in pain. So I can rule out any skeptical scenario in which I do not exist. Thus, my belief that I exist is immune to skepticism.

All of this can be captured within the formal framework mentioned earlier. Consider my belief (M) I exist. Since 'I' refers to me, and all of my epistemically possible worlds are centered on me, M comes out trivially true in every world. But we can bracket this trivial truth (as above) by conceiving of epistemically possible worlds as centered, not on a particular *individual*, but rather, on a particular *epistemic situation*,

---

but, of course, it's crucial that one imagines *oneself* undergoing those experiences. That is, one considers the experiences of another *from one's own perspective*.

[55] This, at least, is how some philosophers put the point. See, for example, Shoemaker (1968), Baker (2000), Nagel (1986), and O'Brien (2007).

[56] See, for example, Shoemaker (1968), O'Brien (2007), Evans (2001), Howell (2006), Cassam (1994), Gertler (2011, p. 215-217), McGinn (1983), and the Introduction to this dissertation (§Assumptions).

which again is constituted by the information available from a particular perspective—in this case, the particular perspective that I happen to occupy at t—setting aside background empirical information about the world. This way we do not prejudge whether the same person exists in every world. So take a world w, considered as actual, centered on q at t, where q is the epistemic situation that I happen to be in at t. Now consider M. M is true in every epistemically possible world centered on q at t. For I am an essential part of q. Thus, any world centered on q at t is a world in which I exist at t. We can see this especially clearly by considering my pain P at t. P partly constitutes q at t. So there is no epistemically possible world centered on q at t in which P does not exist. And, again, it's not just that P is *someone's* pain. Rather, q is such that P is manifestly *my* pain. So P is my pain in every epistemically possible world centered on q at t. So I exist in every such world. Thus, given q, M is epistemically necessary. It is certain. It is beyond doubt (epistemically). In other words, my belief that I exist is immune to skepticism.

Again, there are a lot of things that I can doubt (epistemically). But my existence is not one of them. So I pass The Epistemic Doubt Test. And I am not an exceptional case. So I assume that we each pass. This, together with the (hopefully well-founded) assumption that we each also pass The Psychological Doubt Test, means that we pass The Doubt Test. So I therefore conclude that (2) is true.

Let's move on to (1). I have already said my piece about The Doubt Test *in general*. So I will not consider objections that reject The Doubt Test wholesale. Anyone who denies that The Doubt Test is *ever* a good test for acquaintance will most likely deny that we can be acquainted with anything, including our experiences. And it is beyond the

scope of this chapter to consider this position.[57] So what I want to do in the remainder of this chapter is address those who agree that The Doubt Test is a good test for acquaintance with experiences, but who are less optimistic about it being a good test for acquaintance with the self.

One reason one might think The Doubt Test works especially for experiences is because experiences are exactly as they appear to be. As Russell (1912) puts it, he can know one of his experiences "perfectly and completely", and so "no further knowledge of it itself is even theoretically possible" (p. 73-74). One might adopt this position and then point out that the self is different—its properties go beyond the way it appears. So one might argue that our partial self-awareness is insufficient to warrant an appeal to The Doubt Test.

I grant that there are important differences between experiences and the self. And, of course, there is more to the self than what we experience it to be. But none of this is relevant here. Remember why The Doubt Test is a good test. If I am aware of some x, but only in virtue of being aware of some distinct y that indicates x's existence, then I can doubt that y is a faithful witness to the existence of x. However, if I am directly aware of x *itself*, then I cannot doubt that x exists, for there is no potentially misleading presentation of x that might allow me to doubt its existence. That's why The Doubt Test is a good test. And that has nothing to do with whether or not x has some properties of which I am unaware. I can be directly aware of something without being aware of all of its properties. One might point out that we can be aware of the *essence* of an experience, which is arguably not the case when it comes to the self. But, again, this is not relevant

[57] However, even those who are skeptical about acquaintance should see that The Doubt Test reveals a key asymmetry in our awareness of things. We can doubt the existence of some things, but not others. Why? I suggest that it's because we are directly aware of the latter, but not the former.

here. I can be directly aware of something without being aware of its essence. So, while I grant that the differences between experiences and the self are very important *in general*, I deny that they are relevant in the present context.

A related concern might be that, whereas with experiences there is no gap between appearance and reality—between how an experience *seems* and how it *is*—with the self there is such a gap. Here the point isn't that the self has hidden properties; rather, it's that the properties of the self that we actually do seem to be aware of could be misleading in a way that the properties of experiences couldn't be misleading. Maybe this difference could be exploited to argue that The Doubt Test works for experiences but not for the self.

The problem is, there is no such difference. Sure, we might be misled about some aspects of the self. But then, the same goes for experiences. The pain in my knee seems worse than it was yesterday. My red visual sensation appears to be caused by a book on my desk. These appearances may be misleading. After all, I'm not acquainted with past pains or the book on my desk. So, even with experiences, the lack of appearance/reality gap is limited. And that's precisely how it is with the self, too. There are aspects of ourselves that we are not acquainted with. So a gap between appearance and reality emerges. But this isn't the case with *every* aspect of the self. As we've seen, there is no appearance/reality gap between my seeming to be the subject of a particular mental state and my actually being the subject of that state. Given that a particular pain exists, if it seems to me that it is my pain, then it *is* my pain. More generally, there is no gap between my seeming to occupy a certain perspective and my actually occupying that perspective. If it seems to me that *these* are my experiences and *this* is my point of view, then these

are my experiences and this *is* my point of view. So, upon further reflection, the apparent asymmetry between experiences and the self dissolves.

Indeed, it seems that anything that can be said about experiences that is relevant for our purposes can also be said about the self. If The Doubt Test is a good test for acquaintance with experiences, then it is a good test for acquaintance with the self. And since at this point I assume that The Doubt Test *is* a good test for acquaintance with experiences, I conclude that it's a good test for acquaintance with the self. Thus, I conclude that (1) is true. Again, (2) is also true—the self passes The Doubt Test. And my argument is valid. So I conclude that we can be acquainted with ourselves. SELF-ACQUAINTANCE is true.

4. Conclusion

SELF-ACQUAINTANCE is a substantive thesis about the nature of self-awareness. However, it is also a somewhat limited thesis. For it doesn't say what role self-acquaintance plays with regard to broader issues such as those having to do with self-reference, indexicals, I-thoughts, action, and so on. So a lot more needs to be said if SELF-ACQUAINTANCE is to become a fully formed account of basic self-awareness. I believe that it can be developed into such an account. But that will have to wait for another day.

For now, let's focus on the progress that has been made. In this chapter I have shown that a widely used, widely appreciated, and quite plausible method for determining whether one is acquainted with something—namely, The Doubt Test—can be extended beyond experiences to the self. So anyone who believes on the basis of this method that

we can be acquainted with our experiences should also believe that we can be acquainted with ourselves. Furthermore, even those not yet wise in the ways of acquaintance have a reason to get on board. For The Doubt Test is an intuitive and well-supported test for acquaintance. So it gives us all reason to think that we can be acquainted with both our experiences and ourselves.

And there's one final implication of this chapter—one that I'll conclude with now. Earlier I said that SELF-ACQUAINTANCE is *almost* completely neutral with respect to ontological accounts of the self (§1). And indeed it is. However, what we've learned in the last two chapters does allow us to say something about our ontology. Specifically, it allows us to rule out the *bundle theory* of selves, which says that we are nothing but bundles of mental states (as opposed to substantial subjects that *have*, and are distinct from, their mental states). This view is often attributed to Hume (1739/1975), who at one point says, "[Each of us is] nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement" (I.iv.6). Throughout this dissertation I have been assuming that we are substantial subjects of our mental states, not mere bundles of mental states. So I have been assuming the falsity of the bundle theory. But the bundle theory is still taken seriously today.[58] So it's worth discharging my assumption.

And the results of the last two chapters make this possible. In Chapter 1, I argued that the self shows up in experience. And, in particular, I argued that there is an experience of the self that adds something to one's total experience over and above one's

[58] However, few contemporary philosophers explicitly endorse the bundle theory. Derek Parfit (1984) is a notable exception. Some philosophers endorse a different view that goes by the name 'the bundle theory', which is the view that all objects are bundles of tropes or property instances (see, e.g., Van Cleve, 1985; Casullo, 1988). But I will not address this view here.

various *self-less* sensations, emotions, thoughts, etc. So if the bundle theory is taken as the view that people are nothing more than successions of these *self-less* mental states, then the arguments in Chapter 1 allow us to rule out the bundle theory.

However, the bundle theory might instead be taken to allow that self-experiences are among the various kinds of mental states that constitute the self. But notice, this implies that self-experiences are *representations* of the self. For one's self shows up in experience as a *subject* of thought (Ch. 1, §2). And if the self is a bundle of experiences, then it couldn't be that one experiences the self *itself* as a subject of thoughts—it would have to be that the self is merely represented as such. But this conflicts with the results of the present chapter. For I have argued that we can be, and often are, directly aware of ourselves as the subjects of our mental states. So, contrary to the present suggestion, self-experience is *not* just a representation of the self; it is an experience of direct self-awareness.[59] Thus, given the results of the last two chapters, it seems we can rule out the bundle theory.

A dyed-in-the-wool bundle theorist may have further responses. But I will leave it at that. At very least we can see that the results of the last two chapters put serious pressure on the bundle theory. And we can also begin to see that, even though claims like SELF-EXPERIENCE and SELF-ACQUAINTANCE are not ontological claims *per se*, they do have ontological implications. They can help us learn about what we are. This is an

[59] Now, interestingly enough, a bundle theorist can agree that we are acquainted with ourselves. For she can agree that we are acquainted with our experiences; and this, on her view, counts as acquaintance with the self. So she can maintain that self-experience is a representation of the self, and nonetheless grant that we are acquainted with the self in virtue of being acquainted with our experiences. But, again, the bundle theorist cannot allow that we are acquainted with subjects (or agents) of our experiences. And this is where things get dicey for the bundle theorist. The bundle theorist has to say that although we are acquainted with ourselves, we are only *represented* as being the subjects of our experiences. But this means that we should be able to doubt that we are the subjects of our experience. For one can, in general, doubt what's represented as being the case. But, as I've argued, I cannot doubt that I am the subject of my experiences (§3).

important point. For in the chapters to follow, I will continue to use what we've learned about self-experience and self-awareness to gain an even better understanding of what we are.

CHAPTER 3

I Think, Therefore I Persist

Consider two rather mundane scenarios. First, suppose that you're lying in bed. You just woke up. But you're alert. Your mind is clear and you have no distractions. As you lie there, you think to yourself, '2+2=4'. The thought just pops into your head. But wanting to be sure of your mathematical insight, and having nothing better to do, you once again think '2+2=4', this time really meditating on your thought.

Now suppose that you're sitting in an empty movie theater. The lighting is normal and the screen in front of you is blank. Then at some point an image of a peach is flashed on the screen. The image isn't up there for long. In fact, it's only on the screen for what seems like an instant—*just* long enough for you to see the peach image on the screen. That's odd, you think to yourself, but at least it's better than being hounded by advertisements. So you focus on the screen. And again you see the peach image pop up on the screen for just an instant.

Here's what the previous chapters teach us about these scenarios. In the first scenario, when you think to yourself, '2+2=4', you are (or at least can be) directly aware of your thought and of yourself as the thinker of your thought. In the second scenario, when you perceive the peach image on the screen, you aren't directly aware of any peach, or indeed, of any peach image on any screen. Still, you are directly aware of your *perception* of the peach image on the screen and of yourself as the perceiver of that image.

These two scenarios are a bit mundane. So I hope to liven things up in what follows by showing that reflection on these scenarios, and on the lessons we've learned from previous chapters, can yield significant philosophical results concerning the nature of persons and their persistence through time. First I will argue that thought and perception have certain temporal constraints such that your direct awareness of yourself as a thinker or perceiver is direct awareness of yourself persisting through time. Then I will argue that this allows us to rule out several prominent theories of personal identity.

1. Thought and Perception

Go back to the first scenario. You are lying in bed and you consciously think to yourself, '2+2=4'. You can be—and so let's suppose that you *are*—directly aware of your thought and of yourself as the thinker of that thought.

Now here's an observation: *thinking takes time*. I don't just mean that thinking a really long, drawn-out thought takes time. I mean thinking *any* thought takes time. No thought is instantaneous. Your thinking '2+2=4', for example, takes time. It may not take a lot of time to think—only a few milliseconds, perhaps—but it does take time. And it's not just that it takes time to produce or generate a thought. Thinking *itself* takes time.[60]

[60] In what follows I will take it for granted—and as obvious—that it takes time to think thoughts like, '2+2=4'. Some philosophers (e.g., Reid, 1855) claim that the structure of experience is such that temporally extended conscious events can be broken down into experiences that are instantaneous (see Dainton, 2008, for further discussion). I will return to this issue later in connection with certain epistemic issues (see §2.2). But note that my claim that '2+2=4' is temporally extended does not depend on the claim that *all* experiences are temporally extended or on the claim that temporally extended experiences cannot be broken down into instantaneous parts.

Thoughts are occurrent conscious events with cognitive (i.e., non-sensory) content.[61] They may include occurrent attitudes such as beliefs, desires, and intentions, as well as non-attitudes such as the mere entertaining of content. Your entertaining the content, '2+2=4', is a thought. So is your consciously intending to write this content down. So is your conscious desire to remember it. Thoughts are not sensations, perceptions, or emotions. They are their own kind of conscious event. And thinking takes time.

Here's another observation: In order for you to think a given thought, you have to be the subject of that thought for as long as it takes to think it. To think '2+2=4', you have to think the whole thought. You could think shorter thoughts like '2' or '2+2'. But that doesn't count as thinking '2+2=4'. Thoughts don't finish themselves. And you don't get full credit for incomplete thoughts. So just as you can't run a mile without taking the time to run it, you can't think a thought without taking the time to think it.

Thus, here's what I will call 'The Thought Claim':

> *The Thought Claim*: In order for you to think '2+2=4' on a given occasion, you must be the continuous subject of the thought, '2+2=4', for some temporally extended duration, t' to t, where t' is some time before t, and the difference between t' and t is equivalent to the length of time it takes you to think '2+2=4' on that occasion.

[61] Some (e.g., Carruthers, 2011; Prinz, 2011) deny that thoughts are conscious. These philosophers may feel more comfortable substituting my talk of thoughts with talk of inner speech or some other conscious event. Doing so will not affect my point. Also, I will talk of conscious *events* rather than, say, conscious *states*. But by doing so I don't mean to be picking sides on any substantive issues.

The Thought Claim combines each of the above two observations. First, it says that your thought is temporally extended. Second, it says that when you think your thought, you are the continuous subject of that thought for as long as it takes to think it. So if you think '2+2=4' from t' to t, you exist as the subject of that thought from t' to t.

I am tempted to say that the Thought Claim is a necessary truth. But I won't say it here. For some may say that it's at least *possible* for you to think '2+2=4' instantaneously. And I don't want to fight about it. In fact, I don't even want to fight about my earlier claim that all actual thoughts are temporally extended. So the Thought Claim is just that it actually takes time for you to think the thought, '2+2=4'. Which is true.

Indeed, the Thought Claim may seem obvious—perhaps so obvious that it's not even worth mentioning. But consider this interesting result: You can be directly aware of a temporally extended event. This is surprising, since it's natural to assume that direct awareness is limited to the present instant. However, you can be directly aware of your thought, '2+2=4'. And this thought is temporally extended. So when you are directly aware of your thought, you are directly aware of a temporally extended event. Here it helps that '2+2=4' is a short thought. You cannot be directly aware of thoughts that take a long time to think, since your awareness of longer thoughts will inevitably rely on short-term memory (and maybe long-term memory, if you think really long thoughts). And arguably you cannot be directly aware of something by remembering it. So you can only be directly aware of shorter thoughts like '2+2=4'. Still, such thoughts *are* temporally extended. So, again, when you are directly aware of your thought, '2+2=4', you are directly aware of a temporally extended event.

And there's more. In order to think a thought you have to be that thought's subject for as long as it takes to think it. So when you are directly aware of *yourself* as the thinker of your thought, you are directly aware of a temporally extended person. Again, one might have thought that our direct awareness of things is only ever instantaneous. But what we've learned so far suggests a different picture. I'll return to these epistemic issues in section 2. However, since the Thought Claim isn't a claim about self-awareness, I'll set these issues aside for now.

Let's move on to the second scenario. You are in an empty theater and you see an image of a peach on a screen. Your perception is not a mere appearance. It isn't a case where it *seems* to you as if some object is before you when it's really not. You aren't hallucinating, for example. Here you actually *perceive* an image of a peach on a screen. So the image exists, it's there, and you see the image because it's there. The image is what causes your experience.[62]

Perception, like thinking, takes time. That is to say, actual perceptual experiences are temporally extended. So all of the lessons from the first scenario apply here. But I want to talk about a *further* temporal constraint that is unique to perception. So, to simplify matters, I won't assume that your perceptual experience is temporally extended.

The temporal constraint that I want to talk about now is this. Given the actual laws of nature and the actual facts about human perception, your ability to consciously perceive a peach image on a screen requires that it be continuously presented to you for

[62] These are conditions on perception. We might say that perception is *factive*. Or we might say that, when it comes to perception, there is a certain *relationship* between a perceiver and an object of perception. Or we might follow Chisholm (1957) in saying that, in perception, an object *appears* to a perceiver, where, in the context of visual perception, 'S appears to x' means that, "as a consequence of x being a proper visual stimulus of S, S senses in a way that is functionally dependent upon the stimulus energy produced in S by x" (p. 149).

some minimum duration. This minimum duration is sometimes called a person's 'visual duration threshold' (VDT), which Dent et al. (2007) define as "the minimum exposure duration required before an individual can correctly identify a briefly presented picture" (p. 304). A person's VDT is not very long, and its length is determined by several factors, including the size and brightness of the stimulus. Dent et al. (2007) and Legge (1978) estimate that, in normal lighting conditions, a visual stimulus must be continuously presented to a person for about 10 milliseconds (ms) in order for that person to *detect* the stimulus, and for about 40 ms in order for the person to correctly *identify* the stimulus (e.g., identify it *as a peach*) (p. 304).[63]

Here it doesn't matter what the precise duration of your VDT is at any given time. What matters for my purposes is just this claim: In order for you to see a peach image on a screen at time t, it must be continuously presented to you for some minimum duration leading up to t; otherwise, you won't see the image. So let's call this 'The Perceptual Claim':

> *The Perceptual Claim*: In order for you to perceive an image of a peach on a screen at time t, it must be presented to you continuously from t' to t, where t' is some time before t, and the difference between t' and t is equivalent to your perceptual duration threshold.

[63] See also Effron (1970) and Warren and Morton (1982). For research on factors that influence a person's VDT at any given time, see Kulikowski and Tolhurst (1973), Keesey (1972), and King-Smith and Kulikowski (1975). For a description of the physical mechanisms that determine visual threshold, see Rudd (1996).

The Perceptual Claim is a claim about *perception*. This is important to keep in mind because otherwise it might be tempting to object to the Perceptual Claim by describing a counterexample like the following. At time t, you have a visual experience of a peach image; however, the image isn't presented to you on the screen from t' to t. What happens is, prior to t, a computer receives and processes information about the image, and then the computer stimulates your visual cortex so that you experience a peach image at t.

It might be tempting to think that this case is a counterexample to the Perceptual Claim. For it might be tempting to think that it is a case whereby you perceive the image even though it is not presented to you on the screen from t' to t. But remember, in order for you to *perceive* something—an image of a peach on a screen, say—you must actually see *it*; that is, the image on the screen must be what causes you to see the image.[64] If a computer causes you to have a visual experience as of a peach image on a screen in virtue of some process carried out independently of you, then that experience is not a perception. We might call it a mere *seeming*. It *seems* to you as if you see an image on a screen. But you don't actually *perceive* the image.[65]

What this brings to light is that The Perceptual Claim has both a necessary conceptual element and a contingent empirical element. It's a necessary conceptual fact

---

[64] And it must cause your experience *in the right way*. A deviant causal chain could go from the image on the screen to your experience of a peach without you actually perceiving the peach. Spelling out what it means to be caused *in the right way* is a notoriously tricky task for any account of any causal process. But I'll leave it at that.

[65] Perhaps the case can be adjusted though. Suppose that you are exposed to an image of a peach for part of the time between t' and t, but computers do the rest. So you experience the peach at t partly in virtue of your exposure to it, and partly in virtue of a computer's antecedent processing. Perhaps now you actually *perceive* the image. I don't know whether this case counts as perception. But suppose that it does. This just shows that in some non-normal cases you can perceive an image while only being exposed to it for part of the time that it normally takes you to perceive an image. In other words, the above case shows (at most) that a computer can make your VDT shorter than normal. But it's still true that you must be exposed to the peach for a temporally extended duration—for the length of your VDT—in order to perceive it.

about perception that you have to bear a particular kind of relation to an entity in order to perceive it. It's a contingent empirical fact about humans that establishing this relation takes time. But a fact is a fact. And the fact is that you have to be exposed to something for a temporally extended duration in order to perceive it.

Thus, in order for you to perceive the peach image on the screen at t, you have to be the continuous subject of that peach image for some interval (i.e., from t' to t). Now, of course, if you are a continuous subject throughout some interval, then you *exist* throughout that interval. So the Perceptual Claim implies that if you perceive the image on the screen at t, then you existed from t' to t. We can put the point more generally: In order for a person to perceive something at a particular time, that person must exist for a minimum duration prior to that time. Thus, if a person *does* perceive something at some time, then she existed continuously for some minimum duration prior to that time. In other words, a perceiving person is a temporally extended person.

This result is different in kind from what we get from the Thought Claim. The Thought Claim says that in order to think throughout some interval you must exist during that interval. The Perceptual Claim goes further and says that in order for you to perceive *anything at all*, at any time, you have to exist throughout some extended interval. With thinking, you may be in some state of thinking at t without having existed prior to t. Hence, from the fact that you are in some such state at t, you cannot infer that you existed prior to t. Perception is different. Since you can't perceive something at t unless you existed prior to t, you *can* infer that you existed prior to t from the fact that you have a perceptual experience at t.

And, again, there's an epistemic upshot. When you are directly aware of your perception of the image on the screen, you are directly aware of a mental event that implies a certain temporally extended relation between you and the image. You may not realize this. Everything about your perception may *seem* instantaneous. But even if the perceptual experience itself were instantaneous, your direct awareness of your perception would still be direct awareness of something that implies your prior existence. Thus, when you are directly aware of yourself as the perceiver of the image on the screen, you are directly aware of an individual who must have been temporally extended. But, again, I'll set these epistemic issues aside for now.

Here's what we've learned so far. Thinking is temporally extended. And you have to think a whole thought in order to think it. Thus, the Thought Claim is true. Perceiving is also temporally extended. And even if we assume that it's not, it remains true that in order to perceive something, you have to bear a certain temporally extended relation to that thing. Thus, the Perceptual Claim is also true.

2. Personal Persistence

The claims of the previous section may seem modest. However, in this section I'll argue that they imply the falsity of several prominent theories of personal identity through time (aka, theories of personal persistence). So, modest or not, the Thought Claim and the Perceptual Claim have big philosophical payoffs.

2.1 Personal Persistence

A theory of personal persistence says what it takes for a person who exists at one time to be identical to a person who exists at some later time. It does this by giving a *criterion*—i.e., metaphysically necessary and sufficient conditions—for personal identity through time.

There are many different theories of personal persistence. A lot of philosophers claim that some sort of *psychological* continuity is necessary and sufficient for personal persistence. On this view, a person P persists from time t to time t*—that is, P at t is identical to P* at t*—if and only if P at t has the relevant psychological connections with P* at t*. Other philosophers claim that people persist in virtue of some sort of *physical* or *biological* continuity. These philosophers maintain that P persists from t to t* if and only if P at t is physically or biologically continuous (in the relevant way) with P* at t*. Other philosophers endorse hybrid theories.

The Thought Claim and the Perceptual Claim aren't theories of personal persistence. But they do imply the falsity of several theories of personal persistence. Specifically, they imply the falsity of theories which suggest that there are (or could be) circumstances in which a person thinks or perceives something, and yet, does *not* persist (i.e., exist continuously) through the time it takes her to think or perceive, or through the aforementioned interval prior to her perception. I now turn to some of those theories.

2.2 Falsified Theories

I'll start with a relatively simple theory. It's not a very popular theory, but its simplicity makes it a good one to start with. So consider the combination of *physicalism* and *mereological essentialism*. Physicalism is the view that persons are wholly physical.

Mereological essentialism is the view that all of an object's parts are essential to it. If mereological essentialism is true, then objects do not, and indeed *cannot*, survive any change in parts.[66] So if *both* physicalism and mereological essentialism are true, then:

(1) Necessarily, if a person P at time t is identical to a person P* at time t*, then P has the exact same physical parts as P*.

The Thought Claim can be used to demonstrate the falsity of (1), and thus, the falsity of physicalism plus mereological essentialism. To see this, suppose that you are back in bed and you think '2+2=4'. This thought is temporally extended. Let's say it takes you from t to t* to think it. Now suppose that you lose a physical part—an atom on the tip your nose—sometime between t and t*. For the sake of simplicity, let's just suppose that this atom on the tip of your nose is completely annihilated sometime between t and t*.

If the Thought Claim is true and you do in fact think '2+2=4' in the above case, then you persist from t to t*. For the Thought Claim says that your thinking '2+2=4' from t to t* implies that you persist from t to t*. But if (1) is true, you *don't* persist from t to t*, for you don't have the exact same physical parts at t* that you had at t. So if the above case as I have described it is possible, then either (1) is false, or the Thought Claim is

[66] I am specifically concerned with *three-dimensionalist* versions of mereological essentialism. (A four-dimensionalist version of mereological essentialism would require a different treatment.) Ted Sider (2001) describes three-dimensionalist mereological essentialism as the view that, "(necessarily:) if x is ever part of y, then x is always part of y (provided y exists)" (p. 180). Roderick Chisholm (1973, 1975, 1976), James van Cleve (1986), and Dean Zimmerman (1995) defend this view. Also, Joseph Butler and Thomas Reid arguably held this view. However, all of these philosophers are *non*-physicalists. Indeed, I'm not sure that anyone holds the combination of mereological essentialism and physicalism. I consider this combination here largely because it helps to illustrate the argumentative strategy that I will go on to apply to more prominent theories of personal persistence.

false. The above case is surely possible. And the Thought Claim is true. So (1) is false. Therefore, the combination of physicalism and mereological essentialism is false.

Let's consider our options in a bit more detail. They are: (a) Deny the Thought Claim by denying that it takes time to think '2+2=4' or that you have to exist for as long as it takes to think '2+2=4' in order to think it; (b) Deny that the above case is possible by denying that you could succeed in thinking '2+2=4' in such a case; (c) Deny that the case is possible for some other reason; (d) Deny (1). I've endorsed (d). But what about the other options? Well, the Thought Claim is unimpeachable, as far as I can tell. So (a) is out. And, aside from the question of whether you think '2+2=4', the above case is clearly possible. So (c) is out as well. Thus, we are left with (b). What a defender of physicalism and mereological essentialism has to say, it seems, is that since it takes from t to t* to think '2+2=4', and since you don't exist from t to t*, it's not really *you* who thinks '2+2=4'. She has to say that although in the above case it *seems* to you as if you are thinking '2+2=4', in fact you aren't.

This is a very bad option to be left with. First of all, it smacks of a dubious distinction between your thinking '2+2=4' and it merely *seeming* to you as if you are thinking '2+2=4'. There is no such distinction. There is certainly no *phenomenal* distinction here. If it seems to you phenomenologically that you are thinking '2+2=4', then you *are* thinking '2+2=4'. This is how it is with conscious events in general. Take pain, for instance. Your seeming to be in pain *just is* your being in pain. In the same way, your seeming to think a thought *just is* your thinking a thought. And there is arguably no *epistemic* distinction in this case either. For in the very act of judging (or considering, or supposing) that you are thinking '2+2=4', you entertain the content, '2+2=4', and thereby

think, '2+2=4'. So on both the epistemic and phenomenal senses of 'seeming', there is no gap between your thinking '2+2=4' and it merely seeming to you as if you are thinking '2+2=4'. So it's simply not plausible to insist that although it seems to you as if you are thinking '2+2=4', in fact you aren't.

This point gains even more force when we consider the *specific kind* of mistake that you would have to be making here. You wouldn't necessarily be wrong to believe that '2+2=4' was thought. It was. Or, at least, each part of it was. If physicalism and mereological essentialism are true, then you thought part of '2+2=4', and the person you replaced—the one with the slightly bigger nose—thought the other part of it. So your mistake wouldn't be in believing that '2+2=4' was thought. Rather, it would be in believing that *you thought it*.[67] But this is not a mistake that you can make. You are, as they (i.e., philosophers) say, *immune* from such errors.[68] If you know that a thought is thought, and you judge on the basis of the way things seem to you that *you* are thinking it, then you are right, you *are* thinking it. So the notion that you are wrong to believe that you thought '2+2=4' simply doesn't gain any traction. There's just no denying that you thought '2+2=4' in the above case. So (b) is false. The defender of physicalism and mereological essentialism doesn't have a leg to stand on.

We can see just how compelling this argument is by comparing it to two less compelling arguments. Here's the first argument: You've been digesting your lunch for the last two hours; which implies that you've existed for the last two hours; yet you've

---

[67] One might characterize the situation as one in which you and a series of other people just like you combined to think '2+2=4' in virtue of each of you thinking part of that thought. The fact remains: *you* did not think the thought. At most, you thought part of it. And, again, you don't get full credit for partial thoughts.

[68] Again, see for example Shoemaker (1968), O'Brien (2007), Evans (2001), Howell (2006), Cassam (1994), Gertler (2011, p. 215-217), McGinn (1983), and my Introduction (§Assumptions).

lost several atoms since then; so physicalism plus mereological essentialism is false. This argument may be sound. But it's not very compelling. A defender of physicalism and mereological essentialism will simply deny that it was really *you* who digested food for the past two hours. And what can you say in response? It's not as if you have any special evidence that it was *you*, as opposed to a series of people continuous with you, who digested the food. For all you know given the way things seem to you, you might not have done any digesting. Thus, the assumption that you digested your lunch is not, by itself, a compelling reason to reject physicalism and mereological essentialism. If you have a good reason to accept physicalism and mereological essentialism, then the rational course may very well be to revise your beliefs about your past digestion.

Here is a somewhat better argument: You've lost several atoms in the last day or so; yet you *remember* existing yesterday; thus, physicalism plus mereological essentialism is false. This argument is better than the previous argument because memory is a pretty reliable source of evidence about the past. Still, the argument is hardly conclusive. For a defender of physicalism and mereological essentialism can, without too much embarrassment, just say that your memory is mistaken. She can grant that your memory is of a person who really did exist yesterday, who is connected to you in various important ways, and who for all practical purposes we can think of as you. But then she can say that, *strictly speaking*, you are not identical to this person; while it may *seem* to you as if the person in your memory is you, in fact, it isn't you. This might be surprising, but it isn't incoherent or even particularly absurd. Memory isn't perfect, after all.[69] The

[69] One might suggest that the potential mistake here isn't, or isn't *just*, a mistake of memory. Perhaps it is a mistake having to do with other cognitive abilities, such as your ability to identify an object that you seem to remember. Whatever the mistake is, the point is just that there is room for the defender of physicalism and mereological essentialism to cast doubt on the basis for your belief that you existed yesterday.

defender of physicalism and mereological essentialism can therefore resist this argument by simply denying the apparent deliverances of your memory. And so while this argument is better than the first, it's not at all conclusive.

Now return to my argument. Here denying the appearances—that is, denying that things are as they seem—simply doesn't work. For, in the case that I described, there is no gap between appearance and reality. If it *seems* to you that you are thinking '2+2=4', then you *are*. So it just isn't reasonable to say that although it seems to you as if you are thinking '2+2=4', in fact, you aren't. There is no epistemic wiggle room here. There is no shred of doubt that might yield the means for resistance.

Think of it this way. There are scenarios—including various skeptical scenarios—in which things seem to you exactly as they do now but in which you neither digested your lunch nor existed yesterday. Perhaps an evil demon is tricking you. Perhaps the universe popped into existence five minutes ago. Or perhaps you are just wrong about what it takes for you to persist through time. These are ways things could turn out to be given the way things seem to you right now. So you can doubt that you digested your lunch or that you existed yesterday. Thus, if you have a good reason to believe physicalism and mereological essentialism, then you may, without absurdity, give up your belief that you digested your lunch or that you existed yesterday. But your belief that you are thinking '2+2=4' is different. It isn't open to doubt. There is no skeptical scenario in which things seem to you as they do but in which you are not thinking '2+2=4'. Given the way things seem to you, it couldn't turn out that you are not thinking '2+2=4'. But (1) implies that, in fact, you are not thinking '2+2=4'. So it couldn't turn out that (1) is true. You can conclusively rule (1) out. That's why I say my argument

against the combination of physicalism and mereological essentialism is especially compelling.[70]

Now, one might wonder why it is that you can be so sure that you are thinking '2+2=4'. I suggest that it's because you can be *directly aware* of your thought and of yourself as its subject (see Ch. 2). This is why there is no room for doubting that you exist from t to t* as the thinker of your thought. The evidence is *right there*. It's incontrovertible. There's no rational way to deny it. But, to be fair, some philosophers say that we can only be directly aware of *instantaneous* mental events. So they would say that while you may be directly aware of instantaneous *parts* of the thought, '2+2=4', you cannot be directly aware of the whole thought at once.[71] Let me offer a few good reasons to reject this view. First, shorter thoughts aren't, or at least don't seem to be, made up of a series of instantaneous parts. They seem *unitary*. And, with these shorter thoughts, *the whole thought* seems to be immediately presented to consciousness. It's not as if only one instantaneous part of the thought is ever immediately experienced. The whole thing is. Furthermore, a good reason to believe that you can be directly aware of the whole thought, '2+2=4', is that it passes The Doubt Test. The Doubt Test says that if one seems

---

[70] There are several different ways to express what makes my argument more compelling than the other two arguments. We might say that whereas facts about digestion provide no non-question-begging evidence of my persistence, and memories provide some fallible evidence of my persistence, my awareness of thinking '2+2=4' provides infallible evidence of my persistence from t to t*. Or we might say that whereas one would be likely to accept the first premise of each of the first two arguments only if one already accepted the conclusion, one can be certain that one is thinking '2+2=4' without any preconceptions about what it takes to persist through time.

[71] Thomas Reid (1855), for example, defends this view. But there are a lot of problems with it (see Dainton (2008) for a helpful discussion). Another view called 'the retentional model' says that each experience contains two components: an instantaneous present component and a component that represents the recent past. This view also has problems. One especially relevant problem is that since your seeming to think '2+2=4' *just is* your thinking '2+2=4', there doesn't seem to be any distinction between your thinking '2+2=4' and your entertaining a representation of the thought, '2+2=4'. So, on the retentional model, it would seem that when you think '2+2=4' you actually think *two* thoughts—one spread out in time, and one instantaneous. In fact, you think *many* thoughts, since each successive instant over a certain period of time will contain a representation of your thought which itself counts as a thought. This feels like a bad result (see Dainton (2008) for other problems with the retentional model).

to be aware of some x and cannot doubt that x exists, then one is directly aware of x (Ch. 2, §2). When I think '2+2=4' I cannot doubt that this thought exists. I assume the same goes for you. So this thought passes The Doubt Test. And since The Doubt Test is a good test for whether you are directly aware of something, we have good reason to think that you can be directly aware of the thought, '2+2=4'. Now, perhaps it's only *very* short thoughts that we can be directly aware of. Some might even say that '2+2=4' is too long. If it is, then we may just pick a shorter thought to talk about. But what we need to resist is the idea that we are only directly aware of *instantaneous* mental events. It's doubtful that any such events exist. And even if they do exist, our *thoughts* that we are directly aware of are not among them. Thinking takes time. As does feeling, hearing, hurting, itching, smelling, and every other kind of conscious event. We can be, and often are, directly aware of these events. Thus, the conclusion that we are directly aware of temporally extended mental events is unavoidable.

Even so, some may still prefer a different explanation for why your belief that you are thinking '2+2=4' is immune from error.[72] The key point here is just that it *is* immune from error. Even if it's not because you can be directly aware of your thought, still, what's important is that my argument does not allow for the epistemic flexibility—the toehold in doubt—that one would need in order to resist it. Therein lies the potency of my argument.

With that said, let's set aside this talk of direct awareness of temporally extended events. Indeed, let's forget for a moment that thoughts and other mental events are

[72] Philosophers do indeed differ on this point. See Gertler (2011, Ch. 7) and O'Brien (2007) for helpful discussions of the various strategies that philosophers pursue.

temporally extended. There is a way to make my case without dealing with these issues. It requires turning to the Perceptual Claim. So let's do that now.

Like the Thought Claim, the Perceptual Claim can be used to demonstrate the falsity of (1), and thus, the falsity of physicalism plus mereological essentialism. To see this, suppose that you are back in the empty theater, and at time t* you see a peach image on the screen. Next, suppose that your VDT is equivalent to the time between t and t*. Finally, suppose that you lose an atom on the tip your nose sometime between t and t*.

If the Perceptual Claim is true, and you perceive the peach image on the screen at t*, then you persist from t to t*. For the Perceptual Claim says that your seeing the peach image on the screen at t* implies that you persist from t to t*. But if (1) is true, you *don't* persist from t to t*, for you don't have the exact same physical parts at t* that you had at t. So if the above case as I described it is possible, then either (1) is false, or the Perceptual Claim is false. Again, the above case is possible, and the Perceptual Claim is true. So (1) is false. Thus, the combination of physicalism and mereological essentialism is false.

Much like the previous scenario, a defender of physicalism and mereological essentialism has to deny that you perceive the image on the screen at t*. She could deny the Perceptual Claim or say that the above scenario is impossible for reasons that I've overlooked. But these don't really seem like live options.[73] According to physicalism plus mereological essentialism, no single person is exposed to the image of the peach from t to

[73] One might be tempted to say that what we have here is a case where your VDT is shorter than normal—that for some reason you have a special ability to see the peach despite only being there in front of the screen for part of the time between t and t*. But this is highly implausible. You wouldn't normally be able to see the peach if you were exposed to it for less than the time between t and t*. And you don't have any special powers in the present case. The displacement of one atom certainly doesn't help. So the obvious conclusion is that your VDT is normal.

t*. Thus, given the Perceptual Claim, a defender of this view has to say that no one perceives the image on the screen.

This is implausible, though. First of all, it *seems* to you as if you perceive the peach image on the screen. And, in fact, it *really is* there. It's not as if computers are stimulating your brain and giving you a false impression of what's in front of you. It's not as if you are hallucinating. The peach image on the screen is actually what causes your experience of it. Thus, not only do you have the right kind of experience—i.e., the experience *as of* a peach image on a screen—but all of the relevant causal connections to your environment are there as well. So the obvious conclusion is that you perceive the image on the screen and you therefore persist from t to t*. Which means that (1) is false.

This argument does not invoke the claim that you can be directly aware of temporally extended events. This might make it a bit less forceful than my previous argument. However, it still possesses much of the same force. For, at t*, you can be directly aware of your perceptual experience. And, in the present scenario, your having this experience implies that you persist from t to t*. So it implies the falsity of (1). There really is no way for a defender of physicalism and mereological essentialism to elude the refutation brought on by the Thought Claim and the Perceptual Claim. The view is simply false.

Maybe you knew this already. As I've said, the combination of physicalism and mereological essentialism is not a popular view. I've discussed it here mainly as a way of greasing the wheels. It's a relatively simple view. So it's a good one to introduce my argument on. However, now that the argument is on the table, we can move on. Let's turn to some more prominent theories of personal persistence.

Consider *animalism*, for instance.[74] Animalism is the view that a person is a certain kind of living organism—namely, a human animal. According to animalism, a person who exists at one time is identical to a person who exists at some later time if and only if she is the same human animal. So if animalism is true, then:

(2) Necessarily, if a person P at time t is identical to a person P* at time t*, then P and P* are the same human animal.

Now return to the scenario where you think '2+2=4' from t to t*. Imagine the following. While you are lying there thinking, aliens destroy most of your body. They do this sometime between t and t*, and the only part of you that they *don't* destroy is your cerebrum. In fact, the aliens manage to sustain the normal functioning of your cerebrum. So, while you lose most of your body sometime between t and t*, you still think '2+2=4'.

Since you think '2+2=4', and it takes from t to t* to do that, the Thought Claim implies that you persisted from t to t*. But (2) implies the opposite. The person lying in bed at t is not the same human animal as the cerebrum at t*. Cerebra aren't animals, after all.[75] Thus, if (2) is true, you don't persist from t to t*. So if the above case is possible,

[74] There are various biological theories of personal persistence that *might* be called 'animalism'. Here I consider the view that Eric Olson (2007) defends and specifically refers to as 'animalism'. Other defenders of biological views that are more or less like Olson's view are Peter van Inwagen (1990), Mark Johnston (1987), Judith Jarvis Thompson (1997), and Bernard Williams (1973).

[75] Olson (2007) says that detached cerebra aren't even *organisms*, let alone animals. He writes, "A detached cerebrum is no more an organism than a detached arm is an organism" (p. 41). However, some philosophers who *might* be called animalists say that cerebra are (or at least can be) organisms if they are separated from the body (see van Inwagen, 1990). These philosophers might say that you *do* persist from t to t*. So the present argument does not apply to those versions of animalism. However, we might amend the above case so that your biological cerebrum and visual system are replaced with an *inorganic* cerebrum and visual system sometime between t and t*. Of course, it's open to animalists to deny that this is possible (given the actual laws of nature). But if it is possible, then the present argument can be applied to *any* version of animalism.

then either (2) is false, or the Thought Claim is false. This case may seem outlandish, but it *is* possible.[76] And the Thought Claim is true. Thus, (2) is false, and so is animalism.

The rest of the story is the same. The same considerations that applied above, apply here. The animalist has to deny that *you* thought '2+2=4'. But, again, this is unbelievable, since you can be directly aware of your thought and of yourself as its thinker. And even if this is denied, your belief that it was you who thought '2+2=4' remains undeniable.

Plus, there's the case of perception. Return to the theater scenario. At time t* you see a peach image on a screen, and your VDT is equivalent to the time between t and t*. Sometime between t and t*, the aliens destroy all of your body except for your cerebrum, eyes, and the rest of your visual system. The aliens manage to sustain the normal functioning of your cerebrum and visual system throughout the procedure. So you still perceive the image on the screen at t*.

The Perceptual Claim implies that you persist from t to t*. But, again, if (2) is true, you don't persist from t to t*. So if the above case is possible, then either (2) is false, or the Perceptual Claim is false. This case is possible, and the Perceptual Claim is true. Thus, again, (2) is false, and so is animalism. And we don't need to assume that you can be directly aware of temporally extended events. So animalism cannot escape refutation.

At this point it may seem like my arguments favor *psychological* theories of personal persistence. But this is not necessarily the case. For The Thought Claim and Perceptual Claim rule out various psychological theories. Consider the *memory view*, for

[76] If this scenario seems too outlandish, consider this: Apparently there have been cases in which someone remained conscious for a short period of time after being decapitated (at least that's what the evidence suggests). I don't know if anyone ever spent her last moments thinking '2+2=4'. But these cases should lend some credibility to the claim that a scenario like the one I've described is possible.

instance. The memory view says that a person who exists at one time is identical to a person who exists at some later time if and only if the later person has *first-personal memories* of events occurring to the earlier person.[77] If the memory view is true, then:

(3) Necessarily, if a person P at time t is identical to a person P* at time t*, then P* has at least some first-personal memories of events occurring to P.

Now return to our two scenarios. This time around suppose that mad neuroscientists erase all of your first-personal memories, including both your explicit and your tacit first-personal memories, sometime between t and t*. In the first scenario, you think '2+2=4'. In the second scenario, you see a peach image on a screen at t*. And in both scenarios you don't remember anything occurring to anyone at t.

If either the Thought Claim or the Perceptual Claim is true, you persist from t to t*. But if (3) is true, you don't persist from t to t*. So if the case I described is possible, then either (3) is false, or both the Thought Claim and the Perceptual Claim are false. Surely this case is at least possible. And both the Thought Claim and the Perceptual Claim are true. So (3) is false. Thus, the memory view is false.

The memory view is one prominent psychological theory of personal persistence. There are other actual and potential psychological theories of personal persistence. One might say that people persist in virtue of continuously possessing other, non-memorial psychological states or processes, such as beliefs, dispositions, personality traits, or

[77] John Locke (1690/1975) defends this view. Sydney Shoemaker (1970), H. P. Grice (1941), Anthony Quinton (1962), and John Perry (1976, 2008) defend more sophisticated versions of it—versions that appeal to continuous chains of memory-connectedness or the ancestral of the remembering relation, for example—that are nonetheless sufficiently similar to what I am calling "the memory view" for the purposes of my arguments.

character traits. However, the Thought Claim and Perceptual Claim also rule out many of these views. For, as you may now realize, these claims imply that you can survive the loss of *any* part, state, or process that is inessential to conscious thought or perception.[78] This includes a wide variety of psychological states and processes. You can lose beliefs, dispositions, personality and character traits, etc., without thereby losing your ability to think or perceive. So the Thought Claim and Perceptual Claim rule out other psychological theories as well.

Indeed, the Thought Claim and Perceptual Claim rule out a *wide variety* of actual and potential theories of personal persistence. I have only discussed a few of those theories here. There may be (and no doubt are) other theories of personal persistence that conflict with these claims. Those theories are also false.

Here I am speaking rather pointedly about the falsity of some of the most prominent (both currently and historically) theories of personal persistence. This may seem immodest. After all, defenders of the above theories have their own arguments. They have given reasons to prefer their respective theories. So perhaps a better way to describe the state of play is just to say that I have given *some* evidence against various theories of personal identity—evidence that is to be considered equally alongside the other evidence.

But that's not how I see it. The evidence that I have given is especially potent. It is evidence that we all have access to. And it is evidence that we can be directly aware of

---

[78] What psychological processes are essential to conscious thought and perception? That's an interesting question. But I think the real question is more general: What processes are essential to *consciousness*? For although I am focusing on conscious thought and perception in this chapter, these arguments could be applied to *any conscious experience*. So there is no *particular* psychological process that is essentially implicated in my argument here. If you experience something, then you persist through some interval. Hence, the question is: What is essential to having conscious experiences? I don't have an answer to that question.

whenever we think or perceive. Thus, rather than being just one consideration among many, this evidence carries special weight. It is certain. It is undeniable. We are all eyewitnesses, so to speak, to the falsity of the above theories of personal persistence. But unlike many eyewitnesses who rely on hazy memories and questionable assumptions, we have the full force of airtight evidence at the ready whenever we need it. We should therefore not hesitate to decry each of the above theories. For we know that they are false.[79]

We want to know what makes a person the same person through time. We want to know what makes someone like you the same person now as you were yesterday, and what it will take for you to be the same person tomorrow, next year, or three decades from now. We may form an opinion as to whether you've persisted by observing the way you look or act. However, there is a more basic, more secure way for you to learn something about what it takes for you to persist through time. Just think to yourself, '2+2=4', then ask: What is required for my thinking this thought? You will have thereby discovered a sufficient condition for your persistence. It's not a necessary condition, since you don't *need* to think '2+2=4' in order to persist. But by noticing that you do in fact persist when you think '2+2=4', you will be able to learn that there's a lot else you don't need in order to persist through time. You don't need to keep all the same physical parts. You don't need to be the same animal. You don't need to remember the past. You

[79] Some philosophers express doubts about using thought experiments to settle these issues about personal identity. They say that thought experiments rely too heavily on intuitions that are unreliable, idiosyncratic, or at the very least unsuited to deal with wild thought experiments about aliens and mad scientists. However, in appealing to the above thought experiments, I am not trying to pump your intuitions. I am not asking you to reflect on your gut instincts or to consider whether certain of your concepts apply to various cases. Rather, I am trying to draw your attention to decisive, incontrovertible evidence about yourself—evidence that you actually have directly before you.

don’t need to have the same personality or character traits, beliefs or desires, plans or goals. These things are superfluous. You can persist without them.

CHAPTER 4

There Is A Criterion of Personal Persistence

A criterion of personal persistence is a set of informative necessary and sufficient conditions for personal persistence. To count as a criterion of personal persistence, the obtaining of these conditions must be metaphysically necessary and sufficient for any person's persistence through time. And these conditions must also be *informative*; that is, they mustn't presuppose the identity of the person in question. If I assert that person P at time t is identical to person P* at time t* if and only if P is *the same person* as P*, then although I have given a necessary and sufficient condition for personal persistence, I haven't given a criterion of personal persistence, since my condition isn't informative. If instead I assert that P at t is identical to P* at t* if and only if *it is raining in France from t to t**, then I have given an informative condition for personal persistence—it doesn't presuppose anyone's identity across time—but I haven't given a criterion of personal persistence, since my condition isn't necessary and sufficient for personal persistence. Hence, to count as a criterion of personal persistence, the conditions on offer must be both informative and also necessary and sufficient for personal persistence.[80]

At this point one might despair of finding a criterion of personal persistence. With several of the most widely held potential criteria now off the table, one may begin to doubt that any such criterion exists. One might simply accept that there is no criterion of personal persistence.

---

[80] It's worth emphasizing that I am not using 'criteria' in an *epistemic* sense. That is, I am not talking about conditions under which we could *know* that a person has persisted through time. Rather, I am talking about conditions under which a person persists through time, regardless of whether we can know it or not.

This reaction to the perceived inadequacy of various criteria of personal persistence is not uncommon. Joseph Butler (1736/2008), for example, has this reaction. Butler acknowledges that people can be consciously aware of having existed in the past. So he grants that people do persist through time (p. 100). However, Butler rejects any and all views according to which personal persistence consists in some further fact or conditions regarding the physical or psychological continuity of persons. He says:

> Now, when it is asked wherein personal identity consists, the answer should be the same as if it is asked, wherein consists similitude or equality; that all attempts to define, would but perplex it (p. 99).

Thomas Reid (1785/2008) takes a similar line. He agrees that people persist through time, and can be consciously aware of it; but, like Butler, Reid thinks that personal persistence is unanalyzable (p. 115-116). Butler and Reid are two of the first and most prominent defenders of this view. But many have followed their lead. For example, Geoffrey Madell (1981) describes his view this way:

> I argue that the correct view of the nature of personal identity is the one associated with the names of Reid and Butler above all: that personal identity is strict and unanalysable … The vast majority of the books and articles which have appeared clearly take the view that some sort of empiricist analysis of personal identity must be given, some account in terms of the observable connections between experience, perhaps … The fundamental error in nearly everything which has been written in this field recently has been the failure to take note of the importance of the first person perspective. This failure, I argue, shows itself in very many ways, and it is in itself part and parcel of the empiricist outlook which I am attacking (preface to *The Identity of the Self*).

Notice that Butler, Reid, and Madell's claim is that there is no *definition* or *analysis* of personal identity. This is a common way for those in Butler and Reid's

tradition to express the claim that there is no criterion of personal persistence.[81] And yet, this talk of analysis is not entirely without controversy.[82] So in what follows I will stick with the claim that there is no criterion of personal persistence.

Contemporary philosophers who endorse this claim include Trenton Merricks (1998), Roderick Chisholm (1976), Richard Swinburne (1985), E. J. Lowe (2009), and Harold Langsam (2001). These philosophers agree that although people do persist through time, there are no informative necessary and sufficient conditions for personal persistence. That is, there is no *criterion* of personal persistence. This is what I will call 'anti-criterialism'.

Anti-criterialism may seem like a good way to go given the results of the last chapter. But in this chapter I will argue that anti-criterialism is *not* the way to go. First I will describe the commitments of anti-criterialists and their opponents (criterialists). Then I will argue that there is a criterion of personal persistence.

1. Criterialism and Anti-Criterialism

Criterialists claim that there is a criterion of personal persistence. Trenton Merricks (1998) characterizes this claim as follows:

> The Criterialist Claim: Necessarily, some object O at t's being identical with O* at t* obtains if and only if there is some criterion of identity over time C such that O at t's satisfying the criterion C with O* at t* obtains (p. 116).[83]

[81] See, for example, Swinburne (1985, p. 20; 2012), Madell (1981, p. 80), Gasser and Stefan (2012), Kanzian (2012), and Langsam (2001, p. 251).
[82] Merricks (1998), for example, argues that questions about criteria of personal persistence do not have the tight connection with questions about analyses of personal persistence that many have assumed exists.
[83] This claim applies to all objects, but in this chapter I will focus only on its application to people.

The Criterialist Claim implies that there are informative necessary and sufficient conditions for personal persistence. There are various criterialist positions. Some criterialists claim that a certain form of psychological or phenomenal continuity is necessary and sufficient for personal persistence. On this view, a person P persists from t to t*—that is, P at t is identical to P* at t*—if and only if P at t has the relevant psychological or phenomenal connections with P* at t*. Other criterialists claim that physical or biological continuity is necessary and sufficient for personal persistence. They say that P at t is identical to P* to t* if and only if P is physically or biologically continuous in the relevant way with P*. Psychological and physical/biological theories are particularly prominent kinds of criterialist positions.[84] But what ultimately unites *all* criterialists is the belief that there is a criterion of personal persistence.

Anti-criterialists reject this belief. They deny the Criterialist Claim.[85] Anti-criterialists do not deny that people persist through time, but they do deny that there is a criterion of personal persistence. They claim that there are no informative necessary and sufficient conditions for personal persistence. Hence, anti-criterialists claim that all supposed persistence conditions are uninformative, unnecessary, or insufficient for personal persistence.

According to anti-criterialists, many proposed criteria of personal persistence are *uninformative*. Again, an alleged criterion of personal persistence is uninformative if it presupposes the identity of the person in question. To use an earlier example, if a

[84] Sydney Shoemaker (1985) and Noonan (2003) hold versions of the psychological view. Barry Dainton (2008) and Galen Strawson (1999) hold the phenomenological view. Peter van Inwagen (1990) and Eric Olson (2007) hold the biological view.

[85] As indicated above, Reid (1785/2008) and Butler (1736/2008) were early proponents of anti-criterialism, and Merricks (1998), Lowe (2009), Langsam (2001), Madell (1981), Swinburne (1985), and Chisholm (1976) are more recent defenders of anti-criterialism. See also Zimmerman (1998).

criterialist claims that P at t is identical to P* at t* if and only if P* is *the same person* as P, then her alleged criterion of personal persistence is uninformative. Or, perhaps less obviously, the notion that P at t is identical to P* at t* if and only if P* has *genuine* memories of P's experiences at t might also be uninformative, and thus, not a criterion of personal persistence. For if by 'genuine memories' one means to assert that the memories which P* believes are of her experiences at t really are of *her* experiences at t, then of course P* at t* is identical to P at t if P* has the genuine memories of P. But that's only because 'genuine memories' is defined in terms of the identity of the person whose memories they are (cf., Merricks, 1998). So this alleged criterion assumes the identity of the person in question. Thus, it is uninformative. And this is the sort of mistake that anti-criterialists accuse many criterialists of making.[86]

But anti-criterialists do not claim that this is the only mistake that criterialists can or do make. Anti-criterialists typically grant that some criterialists provide informative persistence conditions. But they claim that all such conditions are either unnecessary or insufficient (or both) for personal persistence. For example, one might claim that spatial continuity is necessary and sufficient for personal persistence. This alleged criterion is informative; it doesn't presuppose personal identity. But an anti-criterialist might claim that spatial continuity is *insufficient* for personal persistence, since it seems that a person could die and therefore be spatially continuous with a corpse, and yet, not persist as a corpse.

And anti-criterialists claim that a similar story can be told for *any* alleged criterion of personal persistence. That is, anti-criterialists claim that any potential criterion—

[86] See, for example, Merricks (1998), Swinburne (1985, 2012), and Madell (1981).

regardless of whether it has been proposed, defended, or even mentioned by anyone—will be uninformative, unnecessary, or insufficient for personal persistence.

2. A Challenge to Anti-Criterialism

In the previous chapter I showed that several of the most prominent theories of personal persistence are false. So anti-criterialists are at least right about the failure of the criteria suggested by those theories. Nonetheless, I believe that anti-criterialists are wrong to claim that *every* potential criterion is bound to fail. That's what I will argue in this section. I will start by bringing to light some specific commitments of anti-criterialism. Then I will show that those commitments lead to absurdity. So I will conclude that anti-criterialism is false.

First, consider a person named 'Sam' who exists at a particular time t. We might ask whether Sam persists from t to some later time t*. And we might also ask whether there are informative necessary and sufficient conditions for Sam's persistence from t to t*. I will call this a *criterion of Sam's persistence*. Criterialists claim, and anti-criterialists deny, that there is a criterion of Sam's persistence. That much is clear.

Now, if anti-criterialists are right—if there is not a criterion of Sam's persistence—then one of the following must be true:

(i) There are no informative *necessary* conditions for Sam's persistence.

(ii) There are no informative *sufficient* conditions for Sam's persistence.

(iii) There are informative necessary conditions for Sam's persistence and there are informative sufficient conditions for Sam's persistence, but there

> are no informative conditions that are *both* necessary and sufficient for Sam's persistence.

Among these available options, most anti-criterialists defend the claim that there are no informative *sufficient* conditions for personal persistence. Richard Swinburne (1985), for example, accepts that there are some informative necessary conditions for personal persistence (e.g., having certain mental capacities; see p. 26), but he argues that informative conditions such as those involving brain or memory continuity are "not *enough* to ensure personal identity" (p. 22; emphasis added).[87] Harold Langsam (2001), who also defends anti-criterialism, similarly writes, "According to the Non-Reductionist [i.e., anti-criterialist], the existence of a self is a further fact, a fact not logically implied by any of the facts adduced by the Reductionists [i.e., criterialists]" (p. 256; see also p. 248).[88] Thus, Swinburne and Langsam's view appears to be that there are no informative sufficient conditions for personal persistence. And other philosophers' comments appear to confirm this understanding of anti-criterialism.[89]

---

[87] Swinburne (1985) also claims that, "complete knowledge of what has happened to a person's body and its parts, and of the extent of the apparent memory by later persons of the deeds and experiences of the earlier person, would not automatically give knowledge of what has happened to the earlier persons …" (p. 35). Again, this suggests that Swinburne's view is that there are no informative sufficient conditions for personal persistence. Harold Noonan (2003) claims that Swinburne's view is that there are no informative *necessary* conditions for personal persistence (p. 94). But I think Noonan is mistaken. Swinburne (1985) does argue that *bodily continuity* is unnecessary for personal persistence. But, elsewhere, Swinburne claims that there are certain other informative conditions that are necessary for personal persistence (e.g., having certain mental capacities, indivisibility; see p. 21).

[88] Here I equate Non-Reductionism with anti-criterialism and Reductionism with criterialism because this is consistent with Langsam's (2001) treatment of Reductionism and Non-Reductionism. I do not wish to make any claims about other, distinct views that may go by the name 'Reductionism' or 'Non-Reductionism'.

[89] For example, David Shoemaker (2002) writes, "So the non-reductionist [i.e., anti-criterialist] believes that even when we have gathered all the facts together regarding the body, brain, and experiences of the person in question, we still do not have the key further fact necessary to determine questions of identity" (p. 146). See also Lowe (2009, p. 139), Madell (1981, p. 79-80), Parfit (1984, p. 309), Chisholm (1976, p. 111), and Eklund (2004). These philosophers say things that *suggest*, but do not entail, that anti-criterialists' view is that there are no informative sufficient conditions for personal persistence.

Thus, it looks like anti-criterialists might accept and defend (ii) above. Of course, this does not mean that anti-criterialists *must* or even *should* accept (ii). So what I will now argue is that anti-criterialists should indeed accept (ii)—the view that there are no informative sufficient conditions for Sam's persistence. Then I will argue that (ii) is false. And so I will conclude that anti-criterialism is false.

To begin with, I take it that everyone, including anti-criterialists, should reject (i). For clearly there are informative necessary conditions for Sam's persistence. Here is one: In order for Sam to persist from t to t*, the universe must not be annihilated sometime between t and t*. Here is another: In order for Sam to persist from t to t*, Sam's intrinsic features must continue to fail to be the intrinsic features of an ordinary cardboard box over that time. These are just two examples of informative necessary conditions for Sam's persistence. Other such conditions abound. So (i) is false.

So if there is no criterion of Sam's persistence, then either (ii) or (iii) must be true. One reason anti-criterialists might prefer (ii) over (iii) concerns a common objection to criterialism: the objection from *fission*. The idea is this. There are possible cases in which a person undergoes fission—is split in two—and the physical and/or mental connections that obtain between the pre-fission person and the post-fission people are enough to commit various criterialists to the absurd conclusion that, in such cases, the one pre-fission person is identical to each of the two non-identical post-fission people. This is one common objection to criterialism.[90] But notice, if an anti-criterialist were to accept (iii), and thus were to accept that there is an informative sufficient condition for personal persistence, then she would be subject to the same objection—she would be hoist by her

[90] See, for example, Swinburne (1985), Merricks (1998), and Gasser and Stefan (2012). My case against anti-criterialism can be made without addressing such objections to particular criterialist views. However, just as a special bonus, I will discuss fission in more detail in section 3.

own petard. For if fission is possible, and if the continuity that could be doubly maintained throughout fission is sufficient for personal persistence, then a single pre-fission person could be identical to two non-identical post-fission people. Same bad result. Hence, anti-criterialists motivated by fission have a reason to prefer (ii) over (iii).

The same goes for another objection to criterialism—namely, Merricks' (1998) objection from *modal coincidence*. Merricks starts by arguing that there is no *analysis*—no definition or account of the nature of—personal persistence. He then points out that, if The Criterialist Claim is true, then "the obtaining of one [contingent] state of affairs (O at t's being identical with O* at t*) is broadly logically necessary and sufficient for the obtaining of a distinct [contingent] state of affairs (O at t's satisfying the criterion C with O* at t*)" (p. 116). Finally, Merricks (1998) suggests that this necessary connection is odd—a striking modal coincidence—given that there is no analysis of personal persistence. He says:

> … we ought to assume, for any distinct and contingent states of affairs S and S*, either that S can obtain in some possible world where S* does not obtain or vice versa, unless there is *some* reason to think otherwise … This is a reasonable assumption, and I think it is presupposed by a great deal of our reasoning about what is broadly logically possible (p. 117-118).

According to Merricks (1998), if there were an analysis of personal persistence, then the necessary connection in question would not be odd. For if there were an analysis of personal persistence, then the satisfaction of some criterion is likely *what it is* for a person to persist through time. But since there is no such analysis, the necessary connection implied by The Criterialist Claim is just an odd modal coincidence. That's the objection. But notice, if an anti-criterialist were to accept (iii), and thus were to accept that there is an informative sufficient condition for personal persistence, then she would

be subject to the same objection. For if there is a sufficient condition for personal persistence, then every world in which that condition obtains between person P at t and person P* at t* is a world in which P and P* are identical. So if there is a sufficient condition for personal persistence, then there is a necessary connection between two distinct contingent states of affairs. Again, same bad result. Thus, anti-criterialists who are compelled by Merricks' (1998) argument have a reason to prefer (ii) over (iii).[91]

And even if we set aside these *ad hominem* considerations, there are independent reasons to reject (iii), which again says that there are informative necessary conditions and informative sufficient conditions for Sam's persistence, but no informative conditions that are *both* necessary and sufficient for Sam's persistence. To see why I say this, suppose that there is just one informative sufficient condition for Sam's persistence—some form of biological continuity, say.[92] This means that Sam can persist through time in virtue of the continuity of his biological makeup. This continuity is sufficient to bring it about that Sam persists; its obtaining absolutely *guarantees* Sam's persistence through time. So no matter what else happens—even if Sam loses all of his memories, personality and character traits, capacities for consciousness, etc.—Sam will

[91] The same could be said for yet another anti-criterialist objection to criterialism: the objection from graduality (see Noonan, 2003; Gasser and Stefan, 2012, p. 8). This objection begins by pointing out that the connections featured in many proposed criteria of personal persistence *admit of degrees*—that is, they may hold to a greater or lesser extent. But, the objection continues, personal persistence does not admit of degrees. So criterialists have to defend the implausible claim that, although the connections featured in their proposed criteria admit of degrees, there is a precise point at which those connections are just strong enough for a person to persist. That's the objection. But notice that this very same objection could be raised against a defender of (iii) so long as the informative sufficient condition for Sam's persistence that she posits admits of degrees.

[92] I leave the precise nature of this biological continuity open. One might think that biological continuity is some sort of *causal* process or relation connecting entities at different times. Or if one is inclined to think that the same biological entity could exist at two different times without being causally connected (if one believes that 'gappy existence' is possible, for instance), then biological continuity might be construed differently (e.g., in terms of qualitative similarity of certain biological attributes). Of course, if one prefers to use a different informative condition here (e.g., psychological continuity, phenomenal continuity, etc.), then that's fine. One could even suppose that there is more than one informative sufficient condition for Sam's persistence. None of these differences make a difference to the arguments that follow.

continue to exist so long as his biological makeup remains roughly intact. Suppose that this is true. Now we might ask: What could explain this? How could it be that Sam persists in virtue of his biological makeup? How could biological continuity guarantee Sam's survival? One natural answer to these questions is that people like Sam *just are* biological entities—that being an instance of a certain kind of biological entity is *what it is* to be a person. But then presumably Sam, who is necessarily the person that he is, couldn't fail to be the particular biological entity that he is. And then it looks like the biological continuity we are supposing to be sufficient for Sam's persistence is also *necessary* for Sam's persistence. In which case (iii) is false.

What a defender of (iii) would have to say here is that even though biological continuity is sufficient for Sam's persistence, Sam is not *just* a particular biological entity, at least not in any sense that implies that Sam *must continue to be*, or necessarily is, the particular biological entity that he is. For if any informative sufficient condition for Sam's persistence obtains in every possible case in which Sam persists—whether that condition is biological continuity, or psychological continuity, or whatever—then the obtaining of that informative condition is both necessary and sufficient for Sam's persistence. And then (iii) is false. Or, if it is supposed that there is more than one informative sufficient condition for Sam's persistence, some of which obtain in some possible cases and others of which obtain in other possible cases, then just take the disjunction of those conditions; if that disjunction is satisfied in every case in which Sam persists, then that disjunction is an informative necessary and sufficient condition for Sam's persistence (cf., Merricks, 2001, p. 196). And then (iii) is false. Thus, the defender of (iii) has to say that there are at least some possible cases in which Sam persists even

though no informative sufficient condition for his persistence obtains. In other words, the defender of (iii) has to say that there is an informative sufficient condition for Sam's persistence, and, *in addition*, there are *un*informative sufficient conditions for Sam's persistence (e.g., Sam persists) that possibly obtain even if no informative sufficient condition for Sam's persistence obtains.

This view may be coherent, but I doubt that it can be developed into a plausible and principled account of Sam's persistence. It's hard to see how one could motivate *both* the claim that there is an informative sufficient condition for Sam's persistence *and* the claim that there are further uninformative sufficient conditions for Sam's persistence. If one has good reason to believe that there is an informative sufficient condition for Sam's persistence—that is, if one has good reason to think that something like biological continuity ensures Sam's survival and can thus explain why Sam persists through time—then it's natural to conclude that uninformative conditions such as *Sam persists* are not truly *further* conditions on Sam's persistence. It's natural to conclude that such conditions add nothing to, or do no extra work for one's account of Sam's persistence. On the other hand, if one has good reason to believe that there are possible cases in which Sam persists without satisfying any informative sufficient condition for his persistence, then it's hard to see why one would think that there's an *informative* sufficient condition for Sam's persistence. If an account of Sam's persistence has to include an uninformative condition such as *Sam persists*, and if this condition can obtain even if no informative condition for Sam's persistence obtains, and if this condition obtains in every case in which Sam persists (as I assume it would), then it's not clear what work any informative

sufficient condition could be doing in that account. It's not clear why one would need to appeal to biological continuity at all.

Here's another way to look at it. One natural reason for one to think that something like biological continuity is sufficient for Sam's persistence is because one thinks that such continuity is *what it is* for Sam to persist through time. But if one thinks this, then one should also think that biological continuity is necessary for Sam's persistence. If, on the other hand, one denies that biological continuity accounts for what it is for Sam to persist, then one should deny that biological continuity is sufficient for Sam's persistence. For otherwise one is committed to an unaccounted-for necessary connection between biological continuity and Sam's persistence that, to borrow from Merricks (1998), makes for a rather strange modal coincidence. So, either way—whether or not one thinks that biological continuity is what it is for Sam to persist through time—one should think that (iii) is false.

For this reason, and for the other reasons outlined above, anti-criterialists should not accept (iii). The view is implausible, and it is not well suited for anti-criterialists' purposes. So anti-criterialists should reject (iii).

Thus, I conclude that if anti-criterialism is true, then (ii) is true—there are no informative sufficient conditions for Sam's persistence. This conclusion may seem like an obvious one, since (ii) is the sort of claim that anti-criterialists typically focus on and defend.[93] But, as obvious as it may seem, this conclusion has not been adequately appreciated up until now. For it has not been made *explicit* up until now. No anti-

[93] And no anti-criterialist that I know of explicitly defends anything like (i) or (iii). Contrary to (i), anti-criterialists like Swinburne (1985, p.21) and Merricks (1998, p. 118) grant that there *are* necessary conditions for personal persistence. And while Merricks (2001) seems to allow that something like (iii) is *consistent* with anti-criterialism, he does not explicitly endorse any such view.

criterialist that I know of has explicitly said that theirs is the view that there are no informative sufficient conditions for personal persistence.[94] So it's useful to get clear about the fact that the lynchpin of anti-criterialism is the claim that there are no informative sufficient conditions for personal persistence.

I now turn to my main challenge to anti-criterialism. It is a challenge to (ii), and it runs as follows. Recall Sam. Sam is a perfectly normal person at time t. Now suppose that all of the psychological, phenomenal, physical, biological, and any other qualitative (i.e., non-identity-assuming) connections that normally obtain in an average persisting person, connect Sam at t with an object we might call Sam* at t*. Sam and Sam* are continuous with respect to all relevant psychological connections—all of the qualitative connections having to do with memory, cognitive abilities, and even personality traits. They are also continuous with respect to all phenomenal connections; we might even suppose that a single, continuous phenomenal stream connects Sam at t and Sam* at t*.[95] Sam and Sam* are also continuous with respect to all of the biological and physical connections that normally obtain in a persisting person. And the same goes for any other such (qualitative) connections. Finally, the connections between Sam and Sam* are *non-branching*.[96] Sam and Sam* are continuous *only* with each other. Sam is not also

---

[94] Anti-criterialists rarely talk about the *specific* formal commitments of their view. They tend to focus their attention on attacking various criterialist positions (see e.g., Lowe, 2009; Swinburne, 1985), or on other issues connected with anti-criterialism (see e.g., Langsam, 2001).

[95] This stream will include *self*-experiences (see Ch. 1). So we must be careful to not make assumptions about the *identity* of the self (or selves) being experienced from t to t*. We can assume that Sam experiences himself at t and that Sam* experiences himself at t*, but we must not assume that the self that Sam experiences is identical to the self that Sam* experiences. So we must not assume that phenomenal continuity from t to t* entails continuity in the identity of the self being experienced from t to t*. If one is worried about this assumption creeping in, one may focus just on the underlying qualitative features (neural or otherwise) that are responsible for self-experience.

[96] I stipulate that these connections are *non-branching* because some (e.g., Shoemaker, 1985, p. 85; see also Noonan, 2003, p. 12-15) maintain that this is a necessary condition on personal persistence. These philosophers point to cases of *fission* (as well as *fusion*)—i.e., imaginary cases where one person is psychologically or physically continuous with two (or more) persons or entities—and they suggest that, in

connected in any relevant way to anyone or anything other than Sam*; nor is Sam* connected in any relevant way to anything other than Sam. Thus, in sum, Sam is just like a normal person in every qualitative way, both in time and through time. According to any plausible theory of personal persistence, Sam and Sam* are the same person.

This example can be made even clearer by comparing Sam and Sam* to Bob and Bob*. Let us assume that Bob, who exists at time t, is the same person as Bob*, who exists at t*. In other words, let us assume that Bob persists from t to t*. Now, Sam's case is such that we can assume that all of the psychological, phenomenal, physical, biological, etc., connections that connect Bob and Bob* are qualitatively identical (in kind) to the connections that connect Sam and Sam*. So Sam really is *exactly* like a normal persisting person in every qualitative way.

This is a problem for anti-criterialism. For assume that anti-criterialism is true. This means that (ii) is true. So, to be more specific, assume for *reductio* that there are no informative sufficient conditions for personal persistence. Then it is the case that all of the non-branching psychological, phenomenal, physical, biological, etc., connections that connect Sam and Sam* do not strictly speaking imply that Sam is the same person as Sam*. In other words, because *ex hypothesi* there are no informative sufficient conditions for personal persistence, the conditions in question that obtain from t to t* are insufficient for Sam's persistence over that interval. There is still a further fact about whether Sam persists from t to t*. Sam might persist over that interval, but he might not. Hence, it is possible that Sam does *not* persist from t to t*. That is, there is possibly a being (or series of beings) that is qualitatively just like a single persisting person in every way except that

---

such cases, the original person does not survive, even though various psychological, phenomenal, biological, physical, etc., connections obtain between that person and some other person(s). Again, I'll discuss fission below in section 3.

he does not actually persist through time. Thus, it is possible that every qualitative connection that normally obtains in a persisting person obtains over some interval without the person in question persisting over that interval.

This result is absurd. It's absurd to think that Sam could be just like a normal persisting person *in every single qualitative way* and yet not be a persisting person. It's unbelievable that all of the psychological, phenomenal, physical, biological, etc., states of Sam at a particular time could continue (unbroken and undivided) in the normal fashion without Sam persisting.[97] If such things were possible, then you or I could be in Sam's shoes! After all, anti-criterialism is a fully general view about personal persistence. If anti-criterialism is true, then no amount of qualitative continuity of *any* kind is sufficient to guarantee *any* person's persistence through *any* period of time. Thus, if anti-criterialism is true, then it's possible that you will be unable to finish reading this chapter, not because you are bored or distracted, or because your mental or physical capacities are about to break down or divide, but because you will simply and inexplicably fail to persist. And it's also possible that you *just now* came into existence, even though your mind and body have existed for years.[98] This, I think, is unacceptable.

And therein lies my challenge to anti-criterialism. Anti-criterialists are committed to the claim that there are no informative sufficient conditions for personal persistence. But surely the obtaining of the sort of continuous non-branching connections and states

---

[97] In fact, we can go even further. We might just suppose that all of the qualitative facts that have anything to do with Sam remain fixed from t to t* so that Sam and Sam* are qualitatively indistinguishable. We might even suppose that the entire universe remains qualitatively fixed from t to t*. In this scenario, it is even more absurd to think that Sam could possibly fail to persist from t to t*.

[98] It is tempting to push this point further and say that since you can be certain of your existence over the past few moments (see Ch. 3), it is not possible that you *just now* came into existence, and so anti-criterialism must be false. However, anti-criterialists can respond by saying that although it's true that your being self-aware from t to t* is sufficient for your persistence from t to t*, this does not show that the obtaining of *purely qualitative* conditions is sufficient for your persistence from t to t*, since the notion of *self*-awareness is identity-assuming.

that begin in Sam at t and continue to t* are sufficient for Sam's persistence from t to t*. Thus, there have got to be informative conditions that are sufficient for personal persistence.

Now, it is important to be clear about what I am *not* imagining, doing, or suggesting here. First, concerning Sam, I am *not* imagining a scenario in which a person exists at t and then is switched out by aliens or annihilated and replaced by God sometime between t and t*. Nor am I imagining a scenario in which a person divides into more than one person. In the scenario I am imagining, there is no interruption in the non-branching psychological, phenomenal, physical, biological, etc., connections that normally obtain in a persisting person. No such connections are severed or split; no interruptions occur.

Second, I am *not* providing persistence conditions for Sam or any other person. For instance, I am not suggesting that the satisfaction of all of the supposed criteria of personal persistence mentioned in the philosophical literature is both necessary and sufficient for personal persistence. I will argue for my own account of personal persistence in the next chapter. But, for now, I am remaining neutral on the question of what is the correct criterion of personal persistence. I am only arguing that there *is* such a criterion, and that, from t to t*, Sam satisfies it. Anyone who thinks that the obtaining of the conditions described in my example is sufficient for Sam's persistence should agree.

Finally, my case for there being a criterion of Sam's persistence does not rest merely on the fact that an average observer would think that Sam persists from t to t*. In other words, I am not just appealing to what would be *observable evidence* for Sam's persistence to support my argument. I grant that Sam could *seem* to persist and yet not persist. That is, I grant that there are conceivable scenarios—e.g., one in which Sam is

annihilated and quickly replaced—whereby Sam fails to persist even though an observer's evidence for Sam's persistence is just about as good as it could be. I grant that such scenarios are possible. But the scenario that I am imagining is one in which an observer's evidence for Sam's persistence is just about as good as it could be *and in fact* all of the non-branching psychological, phenomenal, biological, physical, etc., connections that normally obtain in a persisting person obtain from t to t*. So I am not just appealing to observable *evidence* for Sam's persistence; I am appealing to *facts* that I take to be relevant to Sam's persistence.

Again, if anti-criterialism is true, then (ii) is true—there are no informative sufficient conditions for personal persistence. But if (ii) is true then it is possible for Sam to be just like a persisting person in every qualitative way and yet not persist through time. This is not possible. So (ii) is not true. Anti-criterialism is therefore false. There *is* a criterion of personal persistence, even if we haven't heard of it yet.

3. Objections

Objection 1: Anti-criterialists have given reasons to reject every alleged criterion of personal persistence currently on offer. So criterialists have to come up with persistence conditions that are both true and informative if they want to fend off anti-criterialism.

The anti-criterialist rejection of each criterion of personal persistence currently on offer does not imply that there is no criterion of personal persistence. A criterialist needn't have a specific answer to the question of how a person persists through time. For instance, I haven't given an answer yet—I haven't endorsed any specific criterion of

personal persistence. But my lack of answer does not imply that there is no answer. I think there *is* an answer. In this chapter I have argued that there is a criterion of personal persistence, even if we haven't heard of it yet. And anyone who thinks that the obtaining of the conditions described in my example is sufficient for Sam's persistence should agree with me.

Objection 2: Perhaps Sam's case does count in favor of criterialism. But don't possible cases of fission count equally against criterialism, since cases of fission show that various criterialist positions are committed to the absurd conclusion that one person could become two? So aren't we just at a stalemate?

At most, the possibility of fission shows that certain *particular* criterialist positions are absurd. If you think that some alleged criterion implies that one person could be identical to two distinct people, then you should reject that alleged criterion. But you shouldn't thereby reject criterialism; you shouldn't thereby reject *all* criteria of personal persistence, including those that have yet to be proposed. Fission is impossible; one person can't become two. So no true view about personal persistence will imply the opposite. But criterialists can accept this. Criterialists can (and should) say that the correct criterion of personal persistence rules out the possibility of fission. Perhaps it's that the connections that figure in this criterion couldn't possibly divide into two continuous branches. Or perhaps it's that the correct criterion of personal persistence has a "non-branching" condition built into it, so that even if the connections in question *could* divide, such a division would result in the death of the pre-fission person. The point is

that criterialists have options. There's room to maneuver. The possibility of fission may threaten some particular criterialist positions. But it doesn't threaten criterialism *as such*.

And it's also worth pointing out that the results of Sam's case are clearer and more obvious than those of fission cases. Fission cases are notoriously difficult to get a grip on. It's not at all obvious which alleged cases of fission are possible. This isn't to say that philosophers shouldn't talk about or appeal to fission. It's just to say that our judgments about Sam's case are clear in a way that our judgments about fission cases are not. It's obvious that Sam persists from t to t*. It's far less obvious what's going on in cases of fission. So even if cases of fission did threaten criterialism as such, the results of Sam's case would still be more compelling.

4. Conclusion

Anti-criterialists claim that there is no criterion of personal persistence. But in this chapter I have shown that if anti-criterialism is true, then it is possible for an entity (or series of entities)—i.e., Sam—to be just like a persisting person in every qualitative way and yet not persist through time. This result is unpalatable. So we should reject anti-criterialism. We should accept that there is a criterion of personal persistence.

CHAPTER 5

Thinkers

In this chapter I will describe and defend my own theory of personal persistence. My view is that we are *thinkers*—things essentially capable of undergoing a certain distinctive form of conscious experience.

But first let me set the stage by reviewing what has been established in previous chapters. In Chapter 1 I argued that the self shows up in experience, and in Chapter 2 I argued that we can be directly aware of ourselves. I put these results to work in Chapter 3 to show that we can definitively rule out several of the most prominent theories of personal persistence. But then after arriving at these seemingly pessimistic results, I argued in Chapter 4 that there is, after all, a criterion of personal persistence.

So now we must find the theory of personal persistence that delivers that criterion. One thing that we are looking for here is a theory of personal persistence that is consistent with the above results. In particular, we are looking for a theory that, together with the details of possible cases, is consistent with The Thought Claim and The Perceptual Claim. This will be a view that implies that the continuity of a certain form of conscious experience—namely, the form of experience required for thinking '2+2=4' or perceiving an image—is sufficient for personal persistence (see Ch. 3). So it will imply that a person can survive the loss of any part or property that isn't necessary for continuity in this form of experience.

But that's not all that we are looking for. For consider this theory: Person P at time t is identical to person P* at time t* if and only if P and P* are continuous with

respect to thinking '2+2=4' or perceiving an image on a screen. This view is consistent with The Thought Claim and The Perceptual Claim. But it is false. For thinking '2+2=4' or perceiving an image is not *necessary* for personal persistence. People can (and do) persist without thinking or perceiving such things. Consider another theory: P at t is identical to P* at t* if and only if P and P* are continuous with respect to the form of conscious experience involved in thinking '2+2=4' or perceiving an image. This view doesn't say that P has to think any particular thought or have any particular perception in order to persist. It just says that P has to be the subject of *some* experience of the relevant form from t to t* in order to persist from t to t*. Still, this view is false. For undergoing experiences isn't necessary for personal persistence. People can (and do) persist through periods of unconsciousness—during sleep, for example.

So in addition to finding a theory of personal persistence that is consistent with the results of the previous chapters, we need to find a theory that can stand up against various other problem cases, empirical data, and so on. And that isn't so easy.

But it is the task of this chapter. I will begin by laying out a number of theories that are plausible and also consistent with the results of the previous chapters. Then I will describe my own theory, argue for it, and highlight its merits. Finally, I will show how my theory handles certain classic problem cases. My goal here is not to refute all competing theories, nor is it to show that my theory is completely satisfying in every way. Rather, my goal is to develop and motivate the kind of theory of personal persistence that I believe can deal with the major hurdles brought out in this dissertation and elsewhere.

1. Options

Most theories of personal persistence are construed either in terms of bodily/physical/ biological continuity, or else in terms of psychological continuity. We now know that many of these theories are false (see Ch. 3). For example, *animalism*—the view that a person P at time t is identical to a person P* at time t* if and only if P and P* are the same human animal—is false. For animalism falsely implies that people cannot survive the (almost) complete destruction of their bodies (Ch. 3, §2). Various psychological theories are also false. This includes views that construe personal persistence in terms of the continuity of memory, belief, dispositions, or personality. For people can survive complete losses or changes in each of these psychological states or properties. Indeed, people can survive the loss of *any* state, property, part, or process that is unnecessary for conscious thought or perception (Ch. 3, §2). This fact diminishes the range of acceptable theories of personal persistence.

But it does not completely diminish the range of such theories. Consider, for example, *Cartesian dualism* as a theory of personal persistence. This view construes personal persistence as consisting in the continuity of an immaterial mental substance. So P at t is identical to P* at t* if and only if P and P* have the same immaterial mental substance.[99] A dualist of this stripe can plausibly say that this immaterial mental substance is necessary for conscious thought or perception. And she can say that nothing unnecessary for conscious thought or perception is necessary for personal persistence. Thus, for all that has been said so far, Cartesian dualism is still on the table.

[99] Another version of substance dualism is one according to which people *are*—that is, are identical to—immaterial mental substances. In order to make this a criterialist position, and thus to make it a live option, one would need to give a criterion of identity over time for immaterial mental substances.

But Cartesian dualism is not the only view still on the table (which is good, since few philosophers are attracted to Cartesian dualism). Another view that is consistent with the results of previous chapters is "the brain view", which says that P at t is identical to P* at t* if and only if P and P* have the same brain. A defender of this view can plausibly say that brains are necessary for conscious thought and perception, and that nothing unnecessary for conscious thought or perception is necessary for personal persistence. So we can't rule out the brain view.

Another view agrees that some sort of physical brain-like processing is necessary for conscious thought and perception, but denies that it has to be a *biological brain* doing the thinking or perceiving. One who holds this view might claim that some artificial material, for instance, could support human consciousness. One would need to specify the conditions under which a person's capacity for consciousness could be maintained in a way sufficient for personal persistence. But perhaps that could be done.

These are just three options. The theories mentioned above do not exhaust our options. But they do show that the results of the previous chapters can be incorporated within various theories of personal persistence that are at least live options. They also show that our options are ontologically diverse—that various views about what we are made up of are consistent with what we've learned so far.

Taking a look at some of our options also helps us get a better sense of what kind of theory we are looking for, where disagreement is likely to emerge, and how our own epistemic limitations might confront us. These options suggest that we should look for a theory of personal persistence that construes our survival as being essentially tied to the continuity of a certain form of consciousness (or capacity for consciousness). This

suggestion leaves ample room for further disagreement, though, especially about what is necessary for consciousness at the ontological level. And here is where it's crucial to recognize our limits. We don't know a lot about what is necessary for consciousness. So we should be careful to not overstep our bounds when it comes to speculating about our underlying ontological nature.

These themes will be developed in more detail as I lay out my own theory of personal persistence. So let me do that now.

2. We Are Thinkers

My view is that we are *thinkers*. We are things essentially capable of undergoing a certain distinctive form of conscious experience. In this section I will describe my view and mention some of its immediate payoffs. Then in the next section I will show how my view solves certain puzzles concerning personal identity.

2.1 We Are Things

We are things. We aren't states, events, or properties. We are things that can be *in* states, *undergo* events, and *bear* properties. So I deny that we are *bundles of mental states*.[100] I agree with Thomas Reid (1785/2002) when he says, "I am not thought, I am

[100] I also reject the view, which unfortunately also goes by the name 'bundle theory', according to which all ordinary objects are bundles of tropes or property instances (see, e.g., Van Cleve, 1985; Casullo, 1988). But this is not my main target here. The main thing I want to reject here is the view, which I argued against in Chapter 2, that people are collections of mental states, events, or properties, which, in itself, does not imply whether or not there exist *things* (or substances) that form a different ontological category from those of states, events, or properties.

not action, I am not feeling; I am something that thinks, and acts, and suffers" (p. 264). And furthermore, I agree with Peter Unger (2006) when he says:

> [The] bundle theory of ourselves may be seen, I think, to be perfectly absurd … [It is] obviously absurd to think that an experiencing may be its own subject, or to think that there may be no difference at all between an experiencer, however small and impoverished, and, on the other side, the experiencing that this subject enjoys (p. 57).

I agree. We are not mental states. We are substantial subjects of mental states. We can be *in* mental states. But we are distinct from them. For we are things that are *in* states, *undergo* events, and *bear* properties. In this way, we are more like rocks, airplanes, dogs, and planets than we are like redness, World War II, or the desire for chocolate ice cream. For, unlike properties, events, and mental states, we are things.

2.2 We Are Thinking Things

We are *thinking* things. Here I am using 'thinking' in the sense that Descartes used it—to refer to *phenomenal experiences* of the sort that we humans undergo. We undergo many kinds of experiences. They include bodily sensations, perceptions, emotions, thoughts, inner speech, and mental imagery. There is something that it's like to undergo each of these experiences. Then there's *self*-experience. Not only do we experience various sensations, emotions, thoughts, etc.; we also experience *ourselves* experiencing them (Ch. 1). Self-experience is the experience of direct awareness of oneself as the subject of one's experiences (Ch. 2). Many—perhaps all—other creatures capable of undergoing experiences do not undergo self-experiences. But we do. And the kind of total phenomenal experience that we undergo as a result of having this distinctive

form of consciousness—namely, total experiences that contain self-experiences—is what I am calling 'thinking'.[101]

We are thinking things. And my view is that we are thinking things *in essence*. So, on my view, personal persistence is essentially tied to continuity in thinking. Persons P and P* are *continuous* with respect to thinking if and only if P and P* are subjects of the same uninterrupted episode of thinking.[102] Personal persistence is essentially tied to continuity in thinking in that no condition that is unnecessary for continuity in thinking figures in or is any part of the correct theory of personal persistence. Thus, either continuity in thinking *itself* is necessary and sufficient for personal persistence, or else continuity in something that is necessary for continuity in thinking is necessary and sufficient for personal persistence.

Let me demonstrate. First, take some condition for personal persistence, C. We know that continuity in thinking is sufficient for personal persistence (Ch. 3). So if the obtaining of C isn't necessary for continuity in thinking, then a person can survive without C obtaining (Ch. 3, §2). So then C isn't necessary for personal persistence. Thus, if the obtaining of C isn't necessary for continuity in thinking, then C is not the correct theory of personal persistence.

---

[101] This gets at what is sometimes called "the depth problem" for experiential theories of personal persistence. The problem concerns how rich, complex, or robust a person's experiences have to be in order for that person to persist through time. Some who defend experiential theories of personal persistence contend that in order for a person to persist through time, her experiences must be fairly cognitively sophisticated (e.g., Unger, 1990). However, given Ch. 3, it should be clear that we are capable of surviving without very much cognitive sophistication. Others, such as Dainton and Bayne (2005), contend that, "no cognitive sophistication is necessary for our survival, and that we could survive with a consciousness of the simplest of forms, e.g., a few basic bodily feelings" (p. 560-561). But I don't go that far. For I say that personal persistence requires experiences that contain (or are capable of containing) *self-experiences*. This strikes me as a plausible middle ground.

[102] There are various ways to put this. Instead of 'episode of thinking' we might say 'series of phenomenal states' or 'phenomenal stream'.

But now suppose that C is a part of a *conjunctive* theory of personal persistence whereby person P at time t is identical to person P* at time t* if and only if both C *and* condition *D* obtain. If the obtaining of either C or D is unnecessary for continuity in thinking, then a person can survive without the conjunction of C and D obtaining. So if any conjunctive theory of personal persistence contains a condition that is unnecessary for continuity in thinking, then it is not the correct theory of personal persistence.

The only other option would be a *disjunctive* theory of personal persistence whereby person P at time t is identical to person P* at time t* if and only if either C *or* D obtain. It's hard to know what to say about such a theory, since no one (that I know of) defends one. Of course, any theory of personal persistence can be *stated* in disjunctive form. A psychological continuity theory, for instance, can be stated as a disjunction of all the *determinate* mental continuities that make for psychological continuity. But the kind of theory we want to consider here is one with multiple sufficient (and only *jointly* necessary) conditions that are divergent enough from each other that the theory isn't better stated in terms of a single, non-disjunctive necessary and sufficient condition. And, again, no one (that I know of) defends such a theory. Perhaps this is because disjunctive theories of personal persistence have an air of implausibility, or because there is no good motivation for adopting such a theory, or because philosophers are averse to disjunctive theories *in general*. Or it could be because disjunctive theories face a special threat of contradiction. Here's an example to illustrate this worry: Suppose that for C to obtain between x and y is for x and y to have the same *brain*, and for D to obtain between x and y is for x and y to have the same *right arm*. Now consider Peter, whose right arm is severed and then attached to a different body—to someone we might call 'Quinton'—

shortly thereafter. So we have Peter who exists at t, Quinton who exists at t*, and then we have the disarmed person—call him 'Peter*'—who exists at t*. D obtains between Peter and Quinton, since they have the same right arm. And C obtains between Peter and Peter*, since they have the same brain. Thus, according to the disjunctive theory in question, Quinton is identical to Peter, Peter is identical to Peter*, and so by the transitivity of identity, Quinton is identical to Peter*. But Quinton is *not* identical to Peter*. So we get a contradiction. For a more likely (though more tendentious) example, we might suppose that for C to obtain between x and y is for x and y to be *psychologically continuous*, and that for D to obtain between x and y is for x and y to be *biologically continuous*. Then we might imagine a case in which some person P's psychology is transferred to a new body so that C obtains between P and one person, and D obtains between P and another person. We would get the same result: a contradiction.

These are just two examples. In fact, *any* disjunctive theory with disjuncts that can possibly diverge in this way will lead to contradiction.[103] So what the (non-existent) disjunctive theorist would need to do is come up with conditions that do not lead to contradiction when disjoined. Of course, she would also have to make sure that her disjunction is a plausible theory of personal persistence. This is a tall order. One that I doubt can be filled. Thus, since disjunctive theories of personal persistence face serious problems, and since no one actually defends such a theory, I will set disjunctive theories aside.

[103] This threat is especially salient in the present context. For we know that continuity in thinking is sufficient for personal persistence (see Ch. 3). So we can assume that this condition is one of our disjuncts. Thus, any disjunctive theory that contains a disjunct that can possibly diverge from continuity in thinking (in the relevant sense) will lead to contradiction.

Let's recap. My claim is that no condition that is unnecessary for continuity in thinking figures in the correct theory of personal persistence. So I claim that personal persistence consists either in continuity in thinking, or else in continuity in something that is necessary for continuity in thinking. Any simple or conjunctive theory of personal persistence that does not meet this standard will be false. And the only other option—a disjunctive theory, which no one defends—is not an appealing alternative. Thus, I conclude that no condition that is unnecessary for continuity in thinking figures in the correct theory of personal persistence. Hence, personal persistence consists either in continuity in thinking, or else in continuity in something that is necessary for continuity in thinking.

2.3 We Are Essentially Capable of Thinking

Personal persistence consists either in continuity in thinking, or else in continuity in something that is necessary for continuity in thinking. Personal persistence does *not* consist in continuity in thinking. For people can survive episodes of unconsciousness. If you want proof, just go take a nap (this chapter may aid this proof in more ways than one). You will survive despite not being the continuous subject of an uninterrupted episode of thinking. Thus, personal persistence doesn't consist in continuity in thinking *itself*; it consists in continuity in something that is necessary for continuity in thinking.

What is that something? My answer: the *capacity* for thinking. On my view, continuity in the capacity for thinking is necessary and sufficient for personal persistence. Here's how I understand the capacity for thinking:

> Person P has the capacity for thinking at time t if and only if, at t, P's parts and their interrelations are such that either P is thinking or P would be thinking if P were to be appropriately stimulated.

Let me explain. A capacity for thinking is a dispositional feature that people have, the exercise of which results in thinking. I assume that people have this capacity in virtue of the nature of their parts and the interrelations among those parts. If we are wholly physical things, then it's likely that we have the capacity for thinking in virtue of how the various parts of our brains are constituted and interrelated. But let me set aside issues of underlying ontology for now. I will return to them in section 2.4.

If at any given time a person's capacity for thinking is being exercised, then that person is thinking. But a person need not be thinking at a given time in order to have the capacity for thinking at that time. I, for example, had the capacity for thinking throughout the night last night even though, at times, I wasn't thinking. This is because my parts and their interrelations were such that I *could have* been thinking—I was *poised* for thinking. If I would've been appropriately stimulated—if an alarm would have gone off, for example, or if someone would have shaken me—then I would have been thinking.

To *appropriately stimulate* someone is to do whatever it takes to get that person thinking. But there are limits. We don't want a notion of appropriate stimulation that is so broad that *any* intervention, no matter how extreme, counts. Perhaps a lifeless lump of tissue could be modified, rearranged, and enhanced so that it thinks. But we would not want to say that, prior to this intervention, the lump of tissue had the capacity for thinking. Otherwise just about everything would count as having the capacity for

thinking. On the other hand, we don't want a notion of appropriate stimulation so narrow that only gentle taps and pleading voices count. After all, heavy sleepers have the capacity for thinking.

So we need to draw the line somewhere. Here's my suggestion: Appropriate stimulation may include any intervention that does not add to or subtract from the sum of the (token) parts directly responsible for one's thinking, and does not alter the structural arrangement of those parts except for minor alterations that are mere bi-products of stimulation (such as small changes in neuronal arrangement that inevitably result from any stimulation).[104] Thus, alarm clocks, vigorous shaking, splashes of cold water, CPR, and electrical shocks all count as appropriate stimulation, but frontal lobe transplants do not (if, that is, the frontal lobe is part of what is directly responsible for thinking).[105] This is somewhat vague. But I doubt there is any way for me to completely avoid vagueness here without saying something implausible or speculative.[106]

---

[104] I've added "directly responsible" here because chopping a person's arm off (for example) would change her structural features, but not in any way that's relevant to her capacity for thinking. Also, removing a tumor from a person's brain would change her structural features, but unless things go badly wrong, such an operation need not threaten that person's capacity for thinking.

[105] I have been assuming that the person in question is in a more-or-less normal environment with the right amount of oxygen, heat, gravity, etc. But cases in which a person is in an abnormal environment may be covered by including interventions that normalize a person's environment as part of what counts as appropriate stimulation.

[106] This is not to say that there is no fact of the matter as to whether an intervention counts as appropriate stimulation. Though I will take not official position on the matter here, I am inclined to say that there is always a fact of the matter as to whether some thing is a person (whether or not we know what the fact of the matter is), as to whether it is thinking, and as to whether some thing has the capacity for thinking. Thus, I am inclined to say that there is always a fact of the matter as to whether an intervention merely stimulates (rather than creates) a capacity for thinking, and thereby counts as appropriate stimulation.

Still, there will be borderline cases of capacities for thinking. Issues concerning borderline cases and, more generally, vagueness when it comes to personal persistence are certainly not unique to my view. Almost any theory of personal persistence will bump up against these issues (Consider: being caught up in a life, being an animal, having the same mental states or characteristics as, and various other conditions that feature in theories of personal persistence, all admit of borderline cases). How one handles these issues will depend on one's view about vagueness and, I suspect, on one's view about the underlying ontology of people. I will not defend a view about vagueness here. And I will say all that I have to say about the underlying ontology of people in the next section.

So a person has the capacity for thinking at any given time if and only if, at that time, she is thinking or would be thinking if she were appropriately stimulated. Now we need to determine what makes for *continuity* (or sameness) in a person's capacity for thinking across time. Again, I assume that it has something to do with continuity in that person's parts and the interrelations between those parts. It might seem natural to discuss our underlying ontological makeup at this point. But, again, let me save that for section 2.4. For now, I'll put things more abstractly. First, here's what I suggest makes for sameness of capacities for thinking at *adjacent* times (which I will call 'CT-Connectedness'):

> *CT-Connectedness*: Person P's capacity for thinking C at time t is identical to person P*'s capacity for thinking C* at temporally adjacent time t* if and only if the parts and interrelations between parts responsible for C* are causally dependent on the parts and interrelations between parts responsible for C, and are either token identical to the parts and interrelations responsible for C, or have sufficient overlap with them such that, if the *non*-overlapping parts and interrelations were removed from P and P*, C and C* would still be capacities for thinking, and, if P and P* were appropriately stimulated, P and P* would be continuous in thinking.

Persons P and P* at adjacent times t and t* have the same capacity for thinking—they are CT-Connected—just in case they satisfy the above description. First, P*'s capacity for thinking must be *causally dependent* on P's capacity for thinking. The main

point of this condition is to rule out scenarios like the following: All of P's parts are annihilated by an evil genius, and then, in the very next instant, recreated to constitute P*. P and P* both have the capacity for thinking, and the parts and interrelations between parts responsible for those capacities are qualitatively identical, but since the causal dependence is missing, P and P* have different parts and thus different capacities for thinking. The causal dependence missing here is the same relation of dependence that obtains in *any* case of something's persisting from one moment to the next. And it is necessary for CT-Connectedness.[107]

But it's not *sufficient* for CT-Connectedness. There must also be a certain degree of continuity (or overlap) in the parts and interrelations between parts that are responsible for P and P*'s capacities for thinking. Complete overlap obviously counts. That is, if *all* of the parts and interrelations between parts responsible for C and C* are *identical* (and C* causally depends on C), then P and P* are CT-Connected. But the overlap needn't be complete.[108] If the parts and interrelations between parts responsible for C and C* are the same except for one insignificant atom, and C* causally depends on C, then P and P* are CT-Connected.

---

[107] This sparsely characterized causal condition is standard. David Lewis (1976) puts the point as it relates to mental continuity as follows: "Such change as there is should conform, for the most part, to lawful regularities concerning the succession of mental states—regularities, moreover, that are exemplified in everyday cases of survival. And this should be so not by accident (and also not, for instance, because some demon has set out to create a succession of mental states patterned to counterfeit our ordinary mental life) but rather because each succeeding mental state causally depends for its character on the states immediately before it" (p. 17). Also see, for example, Rey (1976, p. 48), Grice (1965, section V), and Shoemaker (1985). Dainton and Bayne (2005) contend that this sort of dependence need not, and indeed should not, be construed as *causal*—that it can be based on a kind of *experiential* dependence. It's not clear to me why the experiential dependence that Dainton and Bayne describe shouldn't count as a kind of causal dependence, broadly construed. But not much turns on this point. At the end of the day, I am happy to be flexible about the precise nature of the dependence here.

[108] That is, unless the correct view turns out to be a version of substance dualism whereby the mind is a simple immaterial substance. On such a view, complete overlap is required for CT-Connectedness.

The question of where to draw the line between not-quite-enough overlap and just-enough overlap is settled through appeal to continuity in thinking itself. Consider the following case. Suppose that we are wholly physical things and that a person's capacity for thinking is instantiated in her brain. Now suppose that somehow, right between t and t*, Paula's cerebrum is swapped out for a new cerebrum. Both Paula at t and Paula* at t* have the capacity for thinking. However, Paula at t is *not* CT-Connected with Paula* at t*. Paula and Paula* don't share the same capacity for thinking. There isn't enough overlap. To see this, set aside their differences and consider only the parts and interrelations between parts that Paula and Paula* have in common. This includes a body and brain stem, but no cerebrum. This—what Paula and Paula* share—isn't sufficient for the capacity for thinking (assuming the capacity for thinking is instantiated in the brain). You can't think without a cerebrum. Thus, Paula and Paula* do not share the same capacity for thinking. They aren't CT-Connected.

Now consider a different case. Again, suppose that a given person's capacity for thinking is instantiated in her brain. But now also suppose that a person's occipital lobe is unnecessary for thinking. Finally, suppose that somehow, right between t and t*, Paula's occipital lobe is swapped out for a new occipital lobe. In this case, Paula and Paula* are CT-Connected, since what they share is enough, by itself, to generate a capacity for thinking such that, if Paula and Paula* were appropriately stimulated, they would be continuous in thinking. Thus, Paula and Paula* share the capacity for thinking. They are CT-Connected.

Whether or not Paula and Paula* are CT-Connected depends on whether Paul and Paula* are causally related in the right way and have enough overlap between them such

that they each the capacity for thinking and, if they were appropriately stimulated, they would be continuous in thinking. The same goes for anyone else at adjacent times. This is what makes for the sameness of capacities for thinking at adjacent times.

But now what about *non*-adjacent times? My capacity for thinking is the same capacity for thinking that I had five minutes ago, yesterday, and indeed, 20 years ago. It's like my body in this regard. My body has undergone all sorts of changes over the years, but it has remained *numerically* the same. Likewise, although the parts and interrelations between parts that are responsible for my capacity for thinking have changed over the years, I've maintained the same capacity for thinking. This, in itself, is neither a surprise nor a problem. Indeed, it follows from my characterization of CT-Connectedness. If person P1 at time t1 is CT-Connected to person P2 time t2 (which is adjacent to t1), who is CT-Connected to person P3 at time t3 (which is adjacent to t2), and so on, then by the transitivity of identity, P1's capacity for thinking is identical to P2's capacity for thinking as well as to P3's capacity for thinking as well as to any person's capacity for thinking that is in an unbroken chain of CT-Connections with P1.[109] Thus, so long as there is an unbroken chain of CT-Connections between my eight-year-old self and me (and there is), we have the same capacity for thinking. Again, this is neither a surprise nor a problem. However, it does raise the question of how to *talk* about sameness in capacities for thinking over extended periods of time. We cannot say that two people have the same capacity for thinking if and only if they are CT-Connected. For I am not CT-Connected

[109] One might deny that the relevant cross-temporal relation here is *strict identity*. A four-dimensionalist, for instance, might talk about persistence in terms of a series of non-identical *object stages* that are related in some way other than by strict identity. If this is one's view, then invoking the transitivity of identity won't work here. But one can get the same result by characterizing sameness in capacities for thinking over any period of time in terms of the *ancestral* of CT-Connectedness, whereby person stage P has the same capacity for thinking as person stage P* if and only if P is CT-Connected to P*, or P is CT-Connected to a person stage who is CT-Connected to P*, or P is CT-Connected to a person stage who is CT-Connected to a person stage who is CT-Connected to P*, and so on.

with my eight-year-old self—we probably do not share enough parts. So what we should say is that two people have the same capacity for thinking if and only if there is an unbroken chain of CT-Connections connecting them across time. Call this 'CT-Continuity'. Now we can say that, for *any two times* t and t*, person P at t has the same capacity for thinking as person P* at t* if and only if P and P* are CT-Continuous. So the parts and interrelations between parts that are responsible for a person's capacity for thinking may change over time, even perhaps quite substantially, while remaining the very same capacity for thinking, just so long as that change is gradual enough that this person never lacks CT-Connectedness from one moment to the next.

So sameness in the capacity for thinking is CT-Continuity. Thus, when I say that personal persistence consists in the capacity for thinking—that continuity in the capacity for thinking is necessary and sufficient for personal persistence—I am saying that personal persistence consists in CT-Continuity. Person P is identical to person P* if and only if P and P* are CT-Continuous. That, together with a few more details, is my view.[110]

2.4 So What Are We?

Thus far I have described, in rather abstract terms, what the capacity for thinking is. And I have described, also in rather abstract terms, what makes for continuity in the capacity for thinking. I have not, however, described which specific structures—which

[110] Notice that what I've said rules out "gappy" existence; that is, it implies that people cannot come back into existence after having gone out of existence. I embrace this implication. However, those who believe in the possibility of gappy existence may amend my view as follows: Allow that CT-Connectedness can obtain between people at non-adjacent times so long as some to-be-specified kind of causal dependence obtains between their capacities for thinking and so long as they would have been continuous in thinking if their capacities for thinking had existed over the temporal gap in question. I leave it to believers in gappy existence to fill in the details.

parts of the brain or mind—are necessary for the capacity for thinking. That is, I have not described the underlying *ontology* responsible for the capacity for thinking. And so I have not taken any stand on issues regarding the ontology of personal persistence.

This is precisely as it should be. For we are ignorant of what's required, ontologically, for the capacity for thinking. We are ignorant of what kind of stuff could give rise to thinking and could thus ensure one's survival. I, for instance, have no idea which parts or interrelations between parts—whether in the brain, or in some immaterial mental substance, or whatever—are necessary for thinking. For I have no idea what *thinking*—the sort of phenomenal consciousness of which we are subjects—depends on, ontologically.[111] And it would be foolish of me to pretend otherwise. For this reason, I believe that the correct position with respect to our underlying ontology is *agnosticism*.

You might think I am overstating our ignorance. You might think that it's obvious that the *brain* is responsible for thinking. After all, there are strong correlations between neural activity (or inactivity, as in the case of sleep or brain damage) and thinking. So you might think that this settles the matter.

But it doesn't. And here's why. First, that there are correlations between thinking and neural activity does not imply that thinking *is* neural activity, or even that thinking is *constituted* by neural activity. These correlations are consistent with a wide range of ontological views, even including substance dualism.[112]

And even if the correlations between thinking and neural activity did give us reason to believe that neural activity constitutes our thinking, or at least is somehow

---

[111] This is for introspective, philosophical, and scientific reasons. Introspection does not reveal the ontological nature of thinking or of the self. Philosophy has not delivered this bit of knowledge either. And, as should be clear from what I say below, science doesn't have an answer either (nor will it ever, at least on its own).

[112] See Lycan (2009) for a frank and illuminating discussion of this point.

directly responsible for our thinking, this still doesn't settle the matter. For even if neural activity is responsible for our thinking, it might still be *possible* for thinking to occur in some other, non-biological medium. It might be possible for a person's brain to be gradually replaced with silicon chips without her ever losing her capacity for thinking. So even if we grant that the brain is what's directly responsible for our thinking, this doesn't settle the issue of what's necessary, ontologically, for the capacity for thinking. Thus, it doesn't settle the issue of what's necessary, ontologically, for personal persistence.

And even if we suppose that thinking is necessarily biological, or that thinking is necessarily instantiated in the brain, still, that does not settle the issue. For it doesn't settle which *specific* structures or neural configurations are necessary for the capacity for thinking. Is the whole brain required? Or is it just the cerebrum? Or do we only need certain portions of the brain—perhaps the frontal lobe—in order to think? Could the biological material that makes up the brain support thinking in another arrangement? Or is there something special—and indeed *essential*—about the current arrangement of the brain?

We have no idea. Perhaps some day we will. But today is not that day. And we shouldn't pretend otherwise. Given our current state of ignorance, we should not take a stand on the ontology of thinking or the capacity for thinking. Nor should we take a stand on the ontology of personal persistence. We should remain agnostic.

2.5 We Are Thinkers

With that, here is my view of personal persistence:

THINKERS: Necessarily, person P at time t is identical to person P* at time t* if and only if P has the same capacity for thinking as—i.e., is CT-Continuous with—P*.

My view is that we are thinkers. We are thinking things essentially, both in and through time. Continuity in the capacity for thinking—i.e., the capacity for undergoing phenomenal experiences that include self-experiences—is necessary and sufficient for personal persistence. And my view—THINKERS—is neutral about matters of underlying ontology. I say that we are *thinkers*, rather than brains or minds or souls or thinking animals, so as to express agnosticism, and thus neutrality, about the underlying ontology of thinking, the capacity for thinking, continuity in the capacity for thinking, and indeed, personal persistence.

THINKERS has several immediate advantages. First, it is consistent with the results of the previous chapters. Specifically, it is consistent with what we learned from The Thought Claim and The Perceptual Claim, which is that nothing unnecessary for conscious thought or perception is necessary for personal persistence. THINKERS says that continuity in the capacity for thinking is sufficient for personal persistence. So it says that continuity in the capacity for thinking is all that is necessary for personal persistence. The capacity for thinking is necessary for thinking, which includes conscious thought and perception. Thus, THINKERS is consistent with what we learned from The Thought Claim and The Perceptual Claim.

Yet THINKERS allows that we can persist through periods of unconsciousness—periods during which we are not thinking. Since it is continuity in the *capacity* for

thinking, rather than continuity in thinking itself, that is necessary and sufficient for personal persistence, THINKERS doesn't run afoul of the obvious fact that I survived the night.

And THINKERS also avoids a different potential problem. Peter van Inwagen (1990) introduces the problem this way:

> To imagine whether a certain situation contains a continuous consciousness we have to find out first whether a certain situation contains a continuously existent thinker. We can't do things the other way around. We can't find out whether the situation contains a continuously existence thinker by first finding out whether it contains a continuous consciousness (p. 206).

The worry here is that thinking-based theories of personal persistence are *circular*, since in order to characterize personal persistence in terms of continuity in thinking, it seems one has to invoke the identity of the person who is thinking. But THINKERS isn't subject to this worry. I've characterized continuity in the capacity for thinking in terms of the parts and interrelations between parts responsible for the capacity for thinking. And, by doing it this way, I have avoided invoking the identities of any person in question. We can talk of person P at time t and person P* at time t*, and, without presupposing whether they are the same person, ask whether they are CT-Continuous in virtue of their parts being continuous in the way described above. And thus, THINKERS avoids the potential problem.

THINKERS also has the advantage of neutrality. We are ignorant of the ontology of personal persistence. This neutrality is built into the name, 'THINKERS'. This is not a cop out. Nor is it an unfortunate bit of squeamishness. For, as will become clear in the next

section, owning up to our ignorance about the ontology of personal persistence is crucial for defusing some of the most pressing puzzles in the literature on personal persistence.[113]

Finally, one immediate advantage of THINKERS is that it's, well, *plausible*. The notion that a person persists through time just in case she maintains her capacity for thinking, and that she goes wherever her capacity for thinking goes, is intuitively satisfying. That this is so will become more and more evident as we move forward and consider certain puzzles and challenges. So that is where I'll turn now.

3. Puzzles

In this section I will consider two philosophical puzzles: Bernard Williams' mind-transfer scenarios, and cases of fission. Each of these puzzles raises serious difficulties for many theories of personal persistence. But I will show that THINKERS can handle them both.

---

[113] Notice that my ontological neutrality also allows one who is sympathetic to my view to embrace the *too-many-thinkers arguments* that give many other mind-based theories of personal persistence such a hard time. Here's Eric Olson's (2003) version of that argument: "(1) There is a human animal sitting in your chair. (2) The human animal sitting in your chair is thinking … (3) You are the thinking being sitting in your chair. The one and only thinking being sitting in your chair is none other than you. Hence, you are that animal. That animal is you" (p. 326).

Given the ontological neutrality of my view, one who is sympathetic to my view can embrace Olson's argument. To see this, suppose that you are indeed an animal. This by itself isn't all that interesting. After all, you are a lot of things. You are a reader, a person, a son or daughter, a philosopher (perhaps), and so on. But this, in itself, says nothing about what you are *essentially*. And it certainly doesn't threaten the idea that there is only one thinker in your chair right now. If you are an animal, you are an animal only *contingently*. For you can survive the loss of your entire body minus your cerebrum (Ch. 3, §2). And cerebra are not animals. This is consistent with the claim that, right now, you are an animal that thinks and acts and reflects, and is capable of doing so in virtue of having a certain biological makeup.

Now, if thinking is essentially a biological process, then you are essentially a biological thing. So, on my view, some biological theory of personal persistence could end up being right. However, since we don't know whether thinking is essentially a biological process, my view not a biological theory *per se*. I say that if animals can't think, then you aren't an animal; if they can think, then you are an animal, but only contingently. What you are essentially is a thinker. And this is consistent with each premise of Olson's too-many-thinkers argument.

3.1 Mind Transfer

Bernard Williams (1973) describes two scenarios that, taken together, generate one of the most perplexing and widely discussed puzzles in the literature on personal persistence. Here is an adaptation of the first scenario:

> Imagine two people, A and B, who enter into a "mind transfer" machine such that they exchange all of their memories, beliefs, desires, personality and character traits, and other mental capacities and characteristics. A's "mind" goes into B's body, and vice versa. And suppose that one person is to receive a large sum of money after the transfer and the other is to be tortured. Now, if A (or B) were given a choice, based purely on self-interest, as to who gets what (the person in A's body gets the money and the person in B's body gets tortured, or vice versa), what should A (or B) choose?

Williams claims that A (or B) would, and perhaps *should*, choose for the money to go to the post-transfer person with her mental characteristics. Given that A and B both want money and no torture *for themselves*, it seems that A should want the money to go to the person who will come to occupy B's body, and B should want the opposite. What this shows, according to Williams, is that we are naturally inclined to believe that a person goes with her mind—that psychological continuity is sufficient for personal persistence, and everything else, including bodily continuity, is unnecessary for personal persistence.

But now to the second scenario, which starts like this:

> Someone in whose power I am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to tomorrow in great apprehension. He adds that when the time comes, I shall not remember being told that this was going to happen to me, since shortly before the torture something else will be done to me which will make me forget the announcement (p. 80).

Williams goes on to imagine that, in addition to having his memories erased, all of his beliefs, desires, personality traits, and other mental characteristics will be

exchanged for an entirely new set of mental characteristics. Williams then asks whether it would still be rational for him to fear the impending torture. His answer: Yes! He says,

> For I can at least conceive the possibility, if not the concrete reality, of going completely mad, and thinking perhaps that I am George IV or somebody; and being told that something like that was going to happen to me would have no tendency to reduce the terror of being told authoritatively that I was going to be tortured, but would merely compound the horror (p. 81).

Williams is inclined to think that he would survive complete psychological upheaval. He believes that it would still be *him* who is tortured. And what this seems to suggest, according to Williams, is that it is actually bodily continuity, not psychological continuity, which is sufficient for personal persistence.

So we have a puzzle. Williams' two scenarios seem to generate conflicting intuitions. It seems we cannot accept both intuitions. But we don't want to reject either one. What Williams tentatively suggests is that our intuition about the second scenario outweighs our intuition about the first scenario. Thus, Williams tentatively suggests that we should adopt a bodily criterion of personal persistence. Others respond differently. Some find the first scenario more compelling.[114] Others say that, if anything, Williams' scenarios just go to show that our intuitions about personal persistence cannot be trusted.[115]

But THINKERS allows us to see through all of this. We can see that Williams' tentative conclusion is wrong, we can see where he went wrong in drawing it, and, most importantly, we can solve his puzzle.

According to THINKERS, Williams would indeed survive the second scenario—he *would* be the one tortured. But this is not because his *body* persists; rather, it's because

[114] See, for example, Shoemaker (1985, Ch. 10) and Noonan (2003).
[115] See, for example, Rovane (1998, p. 44), Nozick (1981), and Gendler (1998, p. 604).

his *capacity for thinking* persists. This diagnosis is consistent with Williams' description of the scenario. And, in fact, it captures our intuitions better than does Williams' diagnosis. To see this, imagine that while Williams' mental characteristics are gradually replaced, so too are the parts of his body. His arms and legs are swapped out for new arms and legs. He gets new organs and a new face. Everything, except perhaps for his brain, is gradually replaced. Now, I take it that insofar as there was ever the intuition that Williams survives, that intuition remains. I suggest this is because, as a *thinker*, Williams can survive the loss of many of his mental characteristics, as well as the loss of much (if not all) of his body, so long as his capacity for thinking remains. So I suggest that Williams is wrong to conclude that his second scenario pushes us toward a bodily criterion of personal persistence. If anything, it pushes us toward a view like THINKERS.

Now consider Williams' first scenario. When we are asked to imagine A undergoing a "mind transfer", we are asked to imagine all of her mental characteristics being transferred to another body. This includes A's thinking and capacity for thinking. So THINKERS implies that A and B switch bodies in this scenario. Thus, THINKERS affirms the first-scenario intuition that people go where their minds go.

But there's more. We can tease apart the elements of the first scenario and create a *new* scenario that offers special support to THINKERS. Imagine the following. Like Williams' first scenario, A and B undergo a kind of "mind transfer". But this time, not all of A and B's mental characteristics are transferred. Their memories, beliefs, desires, personality, character traits and other dispositional states are transferred as before. But their thinking and capacities for thinking are not. Throughout the transfer, A's body and B's body are subjects of continuous streams of experiences that are not transferred along

with their other mental characteristics. Now, I suggest that in this scenario the intuition is that A and B do *not* switch bodies—they stay where they are (cf., Dainton and Bayne, 2005). Put yourself in A's shoes. First imagine yourself undergoing *complete* mind transfer, as in Williams' first scenario. When I do this, I imagine taking up a different *experiential perspective*. I imagine my *point of view* changing. But now imagine your body undergoing a continuous stream of (non-transferred) experiences. When I do this, I have the intuition that I do *not* transfer bodies. All of this suggests that what's really crucial here is the *thinking*. If A and B's thinking and capacities for thinking are transferred, then A and B switch bodies; if their thinking and capacities for thinking are not transferred, then they don't switch bodies. In other words, A and B go where their thinking and capacities for thinking go. That's the intuition. And it supports a view like THINKERS.

So THINKERS not only accommodates the intuitions from both of William' scenarios, and thus solves one of the most vexing puzzles about personal persistence, THINKERS also especially supports intuitions generated by a more nuanced look at Williams-style scenarios. Thus, at the end of the day, what we have is not a puzzle about personal persistence; rather, it's a positive argument for a view like THINKERS.[116]

3.2 Fission

[116] Might we go further and say that these scenarios *confirm* or *prove* THINKERS? I don't think so. These scenarios do highlight the intuitive plausibility of THINKERS. But we don't know what is necessary, ontologically, for thinking. So we don't know whether it is possible to "switch bodies". Nor do we know what alterations a person could undergo without losing the capacity for thinking. Thus, we should remain agnostic about the possibility of the above scenarios. So while these scenarios do highlight the intuitiveness of THINKERS, they do not prove it.

Fission occurs if a person is split amoeba-like into two distinct people. The possibility of fission generates a puzzle for various (criterialist) theories of personal persistence. For various theories of personal persistence seem to imply the impossible—namely, that one person could be identical to two distinct people.

For example, if person A's psychology could fission so that A at time t is psychologically continuous with two distinct people—B and C—at time t*, then various psychological theories of personal persistence imply that A is possibly identical to *both* B and C. But this is impossible, since B and C are distinct. Thus, the threat of contradiction arises. And the same applies to other theories of personal persistence. If *biological* or *physical* fission is possible, then biological or physical theories of personal persistence face the same problem. As do other theories of personal persistence. Hence, the puzzle of fission.

When it comes to THINKERS, the puzzle arises with the possibility of fission in one's capacity for thinking. This possibility is made vivid by cases of *commissurotomy*, wherein the connective tissue between a person's two cerebral hemispheres is severed. Some suggest that in such cases the brain's two hemispheres constitute distinct seats of consciousness that could in principle be fully separated into two autonomous mental lives.[117] Thus, some suggest that commissurotomy supports the notion that mental fission—including fission in one's capacity for thinking—is possible (perhaps even actual!).

But it's not quite that simple. The behavioral abnormalities of commissurotomy patients that are sometimes given as evidence for two distinct minds are in fact quite obscure and difficult to interpret. They usually only arise in certain highly artificial

[117] See, for example, Davis (1997), Gazzaniga and Ledoux (1978), Moor (1982), and Marks (1981).

experimental conditions. And they are consistent with a variety of interpretations. It's not obvious that the disruptions caused by commissurotomy ever reach the level of consciousness. And it's not clear how severe or deep or divisive these disruptions really are.[118]

Nonetheless, we can perhaps *imagine* cases like commissurotomy where one's mental life is split in two. And we can ask: How should we deal with such cases? So let's consider some options. One option is to just add a "no branching" clause to THINKERS as a necessary condition for personal persistence. This would not imply that fission in one's capacity for thinking is impossible. But it would imply that it's impossible for a person to survive such fission. So, on this option, fission-generated contradictions are ruled out by stipulation. One problem with this option is that it seems *add hoc*. Another problem is that, on this option, a person's persistence depends on facts *extrinsic* to her—it depends on facts about whether some *distinct* person is created as the result of fission. These problems may or may not be conclusive. But a further problem—one that *is* conclusive from my perspective—is that the no-branching clause is ruled out by the results of Chapter 3. In that chapter I showed that continuity in conscious thought or perception is sufficient for personal persistence. But given that a no-branching clause is meant to *add* a condition to one's theory, including such a clause would imply that continuity in conscious thought or perception is *not* sufficient for personal persistence, since doing so would imply that that conscious thought or perception *also* has to be non-branching.[119]

---

[118] For research suggesting that the effects of commissurotomy are limited, equivocal, restricted to certain conditions, and do not result in multiple streams of consciousness, see, for example, Zaidel (1994), Fergusen (1985), Akelaitis (1944), Bogen (1993), Bayne (2008), and Bayne and Chalmers (2003). For helpful reviews of the scientific literature on the effects of commissurotomy, see Gazzaniga (2000), Springer and Deutsche (1998), Seymour et al. (1994), Sidtis (1986), and Wolford et al. (2004).

[119] Here I mention that a non-branching clause is meant to add a condition to one's theory because one could think (and, as will become clear below, I *do* think) that, just given the nature of thinking, there

Thus, the sufficiency of conscious thought or perception for personal persistence, which was established in Chapter 3, rules out the no-branching clause. So that option is off the table.

Another option is to say that if a person's capacity for thinking were to fission, she would remain a single, albeit pathologically disjointed person with one mind and one capacity for thinking. But this option isn't very appealing. A certain amount of confusion or disunity may be permissible. But suppose that each of a fission patient's two capacities for thinking could be made truly autonomous from each other—perhaps separated into two distinct bodies so as to yield two completely different perspectives. In such a case, it seems implausible to insist that what we have is still one person with one mind.

A third option is to say that, from the start, each human body contains two distinct minds. We could say that normally these minds are well coordinated, but in cases like commissurotomy, that coordination begins to break down and reveal a fault line that was there all along. This is an option. But it's not a good one. For consider these potentially embarrassing questions: If there are two minds—two *thinkers*—in my body right now, then which one am I? Am I both thinkers? Or am I just one of them? It sure seems like there is just one thinker in my chair right now. Could I be wrong about that? When 'I' issues from my mouth, are two thinkers being mentioned? Or is it just one or the other? When I attend to my thoughts, am I the only one attending to them? Or is my bodily cohabitant able to peer into my mind? These questions reveal serious problems for the "two minds" solution to fission.

---

couldn't be a split or branching in thinking of the sort that would generate worries about fission. So one could think that such branching is impossible, and yet not add a no-branching clause to one's view because such a clause would be redundant—it wouldn't constitute a *further* condition on personal persistence.

So none of the above options are great. And there are even worse options out there.[120] That's what makes fission such a puzzle.

But consider one more option. It's the option that I prefer. And I think it's quite plausible, given THINKERS. It's this: Fission in a person's capacity for thinking is impossible. That is, thinking and the capacity for thinking are *indivisible*; they are essentially *unified*. Thus, the threat of fission does not arise for THINKERS.

Sounds easy enough. But it's not. It's easy enough to say that fission in one's capacity for thinking is impossible. In fact, THINKERS implies this. For it implies that continuity in one's capacity for thinking is sufficient for personal persistence. And it's impossible for one person to be identical to two people. So THINKERS implies that fission in one's capacity for thinking is impossible. But this hardly solves the puzzle of fission. For it looks like THINKERS *also* implies that fission in one's capacity for thinking *is* possible. This is because, given my construal of CT-Continuity, there doesn't seem to be anything to rule out the possibility that one person could be CT-Continuous with two people who each have their own distinct capacity for thinking. So the lingering worry is that THINKERS implies a contradiction because it implies *both* that fission is impossible *and* that it is possible.

So what I need to do is motivate the claim that even though it might initially seem like fission in one's capacity for thinking is possible given my description of CT-Continuity, in fact it's not. I will do this by giving independent reasons—that is, reasons other than "I say so"—for believing that what I say about CT-Continuity does not imply that fission in thinking is possible. This will motivate the claim that it's not possible for

[120] For a survey of the available options, see Nagel (1971) or Van Inwagen (1990, Ch. 16).

one person to be CT-Continuous with two people who have two distinct capacities for thinking.

I assume that the reason it might seem like my description of CT-Continuity implies that fission in one's capacity for thinking is possible is because one can imagine a scenario in which a person's brain is split in two, each half supports its own autonomous capacity for thinking all by itself, and so one person ends up being CT-Continuous with two distinct people who have two distinct capacities for thinking that give rise to two distinct streams of thinking. But, in order for my solution to the puzzle of fission to work, I must deny that such an outcome is possible. And I do deny this, even though that may seem strange at first. I claim (and will argue) that each one of our brain hemispheres cannot, all by itself, support its own autonomous capacity for thinking. Perhaps one of the hemispheres can. But if so, the other can't. Or if it's not the brain that is directly responsible for thinking, then whatever *is* directly responsible for thinking cannot divide so as to yield two autonomous capacities for thinking. No doubt there could have been creatures a lot like us who had two capacities for thinking per head. Maybe there *are* such creatures. But we are not they. We just don't have it in us. Each one of our capacities for thinking is unitary and indivisible—incapable of being split into two capacities for thinking.

That's my claim. Now for my justification: If our minds could divide so as to yield two autonomous capacities for thinking, then fission in thinking and the capacity for thinking would be possible; but we have independent reasons—reasons other than "I say so"—for believing that fission in thinking and the capacity for thinking is not possible, and thus, we have independent reasons for believing that our minds cannot divide so as to

yield two autonomous capacities for thinking. To begin with, we have independent reasons for believing that thinking is *essentially unified*, and thus, cannot possibly fission. Tim Bayne and David Chalmers (2003) describe the unity of thinking as follows:

> At any given time, a subject has a multiplicity of conscious experiences. A subject might simultaneously have a visual experience of a red book and a green tree, auditory experiences of birds singing, bodily sensations of a faint hunger and a sharp pain in the shoulder, the emotional experiences of a certain melancholy, while having a stream of conscious thoughts about the nature of reality. These experiences are distinct from each other: a subject could experience the red book without the singing birds, and could experience the singing birds without the red book. But at the same time, the experiences seem to be tied together in a deep way. They seem to be *unified*, by being aspects of a single encompassing state of consciousness (p. 23).

I claim that thinking is essentially unified in the way described above. I do not deny that one's thinking could be confused or disorganized. Nor do I deny that there could be disruptions in one's access to information that one normally gets by thinking.[121] My claim is that it's essential to thinking that the various elements of one's thinking at any given time are experienced together in a single total state of thinking. In other words, it's impossible for one's thinking to be divided up into multiple non-unified total states.

Hence, fission in thinking is impossible. This claim is supported by the fact that fission in thinking is *inconceivable*. I can conceive of all sorts of disunities in my experiences, such as my vision splitting in two, an oar looking bent but feeling straight, or my feeling that I'm flying (via virtual reality glasses) while I also feel like my feet are firmly on the ground. I can conceive of such things. But what I *cannot* conceive of is my

[121] Indeed, a breakdown in one's access to certain kinds of information—that is, a breakdown in what's sometimes called "access consciousness"—is a plausible way to understand the deficits of commissurotomy (see, e.g., Bayne (2008) and Bayne and Chalmers (2003). Such a breakdown doesn't imply any disunity in thinking.

thinking dividing into two distinct streams.[122] The best that I can do is imagine half of my experiences becoming very different from the others, as if I'm watching two TVs tuned to different channels. But, in such cases, I'm still experiencing all of my experiences as one total experience. What's inconceivable is an experiential disunity that would verify the claim, "These are all *my* experiences, but I am not the subject of any total state of thinking that contains them all".[123] Such disunity is inconceivable. And that's a reason to believe that fission in thinking is impossible.

So my solution to the puzzle of fission starts with the plausible claim that thinking is essentially unified, and thus, cannot fission. Next, I claim that fission in the *capacity* for thinking is also impossible. For, if thinking is essentially unified, then the capacity for thinking is also essentially unified. To see this, recall:

> Person P has the capacity for thinking at time t if and only if, at t, P's parts and their interrelations are such that either P is thinking or P would be thinking if P were to be appropriately stimulated.

Now consider capacity for thinking C of person P at time t. For C to count as a capacity for thinking, C must be such that if P were appropriately stimulated, P would be thinking. And since thinking is essentially unified, the thinking that C gives rise to cannot be

---

[122] Here's how Bayne and Chalmers (2003) put it: "… there seems to be something *inconceivable* about phenomenal disunity. It is difficult or impossible to imagine a subject having two phenomenal states simultaneously, without there being a conjoint phenomenology for both states. And there is a sense that there is something incoherent about the suggestion" (p. 41).

[123] Perhaps the idea isn't that I could become the subject of two distinct streams of experience; perhaps it's that I could be split into two distinct *subjects* of two distinct streams of experience. But this doesn't help on the conceivability front. I can imagine experiencing X at time t1 and then not experiencing X at time t2, while perhaps someone else experiences X at t2. But I have no idea what it would be to imagine myself splitting into two subjects of experience.

divided into multiple non-unified total states of thinking. In this way, C must be unified. And since C is such that it *must* be unified in this way, C cannot fail to be unified in this way. Thus, C cannot fission. Fission in the capacity for thinking is impossible.

Here's another way to look at it. On my view, in order for C and C* (P*'s capacity for thinking at t*) to be identical, P and P* must share enough parts such that if P and P*'s non-overlapping parts and interrelations were removed, C and C* would still be capacities for thinking, and if P and P* were appropriately stimulated, they would be continuous in thinking. Thus, in order for C and C* to be the same capacity for thinking, it must be that P and P* would be continuous in thinking if they were appropriately stimulated. But neither P nor P* can be continuous in thinking with something that is the subject of multiple, non-unified episodes of thinking. For this would imply that fission in thinking is possible. But fission in thinking is not possible. So in order for C to be identical to C*, neither C nor C* can be divided. Thus, fission in the capacity for thinking is impossible.

It should now be clear that my description of CT-Continuity does *not* imply that fission in one's capacity for thinking is possible. In fact, it implies the opposite. In order for person P at time t to be CT-Continuous with two people—P* and P**—who each have their own distinct capacity for thinking at time t*, P would have to be continuous in thinking with both P* and P** if they were all appropriately stimulated. But that would imply that P's thinking could be continuous with two distinct streams or episodes of thinking. In other words, it would imply that P's thinking could *fission*. But P's thinking cannot fission. Fission in thinking is impossible. Thus, it is not possible for one person to

be CT-Continuous with two people who each have their own distinct capacity for thinking.

Again, it might initially seem like fission in one's capacity for thinking is possible given my description of CT-Continuity. But I have argued that, in fact, this is not possible. So, again, I claim that whatever it is that is directly responsible for our thinking—whether it's the brain or something else—it cannot divide so as to yield two autonomous capacities for thinking. Each one of our capacities for thinking is unitary and indivisible—incapable of being split into two capacities for thinking.

At this point, one might worry that I've rigged the game—that the above results are merely an artifact of the way I have construed the capacity for thinking. I have tied the capacity for thinking to thinking *itself*. And, by doing this, the essential unity of thinking guarantees the essential unity of the capacity for thinking.

But I deny that this is an odd or unfair result. I have not rigged the game. I have made it exactly as it should be. I have characterized and individuated capacities for thinking by *what they do*. They give rise to thinking. And thinking is essentially unified. So it makes perfect sense that capacities for thinking are essentially unified.[124]

Or consider this: No *active* capacity for thinking can give rise to multiple non-unified total states of thinking. Such fission is impossible. This leaves open whether an *inactive* capacity for thinking can fission. But the only difference between an active and an inactive capacity for thinking is that one (but not the other) needs its possessor to be appropriately stimulated in order for it to give rise to thinking. And appropriate stimulation does not involve any addition to or rearrangement of the structural features

[124] The alternative is to ascribe to people multiple capacities for thinking. But this not only belies the true nature of our thinking; it also suggests that the various elements of our thinking form their own autonomous domains. And this gives rise to the very puzzle we wish to solve.

responsible for a capacity for thinking. So there is no reason to believe that the differences between active and inactive capacities for thinking are so great that the one cannot fission but the other can. If an inactive capacity for thinking can give rise to divided thinking, then so can an active capacity for thinking. But an active capacity for thinking cannot give rise to divided thinking. So neither can an inactive capacity for thinking. It may be tempting to worry that more is possible when one isn't on guard—when one is asleep and unaware of what's going on. But this temptation is no more rational than the temptation to be afraid of the dark. In both cases, nothing changes just because the lights go out; the darkness offers no additions or subtractions save for the light by which one sees.

So the claim that fission in one's capacity for thinking is impossible is just as plausible as the claim that fission in thinking itself is impossible. Thus, my solution to the puzzle of fission is to deny the possibility of fission in the capacity for thinking. There is good reason to believe that thinking and the capacity for thinking are essentially unified. And there is no compelling reason to believe otherwise. So I say such fission is impossible.

Now, notice that this solution to the puzzle of fission is only plausible given a view like THINKERS. It's plausible to say that fission in thinking and the capacity for thinking is impossible, because we have independent reasons for believing that thinking and the capacity for thinking are *essentially unified*. But this isn't true of other kinds of states or processes that feature in other theories of personal persistence. Take physical or biological processes, for example. It's not plausible to say that these processes are essentially unified. For one thing, physical and biological fission is clearly conceivable in

a way that fission in thinking is not. I can easily imagine a person's body gradually developing in such a way that it splits amoeba-like into two separate organisms. In fact, it doesn't require any imagination to see that physical or biological processes *in general* are divisible. Just consider amoebas, twins, or plant cuttings. So the motivation just isn't there for denying that fission in physical or biological processes is possible. The same goes for various non-conscious psychological states and processes. I can easily imagine a division in my brain activity, or I can imagine some of my memories, beliefs, desires, personality traits, or other mental states becoming isolated from other of my mental states. In fact, disunities of this sort *actually do* occur. Just consider Dissociative Identity Disorder, commissurotomy, or dementia. So there isn't the same motivation for denying the possibility of non-conscious psychological fission as there is for denying the possibility of fission in thinking. Thus, THINKERS supports a plausible solution to the puzzle of fission that is not available to the various other theories of personal persistence that center on the above processes.

4. Other Virtues

THINKERS helps solve two of the most important puzzles in the literature on personal persistence. This is a virtue. And it's a virtue that, together with the other virtues that come with the consistency and plausibility of the theory, is enough to make THINKERS a formidable theory of personal persistence.

THINKERS has other virtues. For example, THINKERS soothes some intuitions that may at first seem to support certain widespread (but false) claims about personal

persistence. Examples of these claims include: People are essentially *agents* (Velleman, 2006); People are essentially capable of *reflection* and *self-evaluation* (Taylor, 1976); People essentially craft *self-narratives* (Schechtman, 2011; MacIntyre, 1984); People are essentially bearers of *moral responsibility* (Locke, 1689/1975). These claims are false. A person can survive the loss of any part or property that isn't necessary for continuity in thinking (Ch. 3, §2). And this includes one's capacity for self-reflection, evaluation, and narration, and for responsible action. However, with THINKERS, we can at least see that these features are *connected* to our essence. It's plausible that the capacities for reflection, evaluation, narration, and action derive from the capacity for thinking. For these capacities seem to require the kind of self-awareness afforded to us by thinking; otherwise, we would expect to find other creatures, with other kinds of conscious experiences that don't amount to thinking, with the above capacities. So the capacity for having conscious experiences that contain self-experiences may be crucial (even necessary) for the development of many of the above and other distinctively human capacities. And so even if the possession of these capacities is not necessary for the capacity for thinking, and thus is not necessary for personal persistence, it nonetheless may be closely tied to the capacity for thinking. Thus, THINKERS can help disarm and pacify the intuitions behind claims like those mentioned above by showing that, while they are off the mark, they aren't too far off the mark—while they are wrong, their wrongness is understandable.

Another virtue of THINKERS is that it gives us a clear way forward in the study of personal persistence. Necessarily, person P at time t is identical to person P* at time t* if and only if P has the same capacity for thinking as—i.e., is CT-Continuous with—P*

(§2.5). But we are ignorant as to what's required, ontologically, for the capacity for thinking (§2.4). So we are ignorant as to the underlying ontology of personal persistence. In order to rid ourselves of this ignorance, we must discover the underlying ontology of the capacity for thinking. Thus, on my view, an important avenue for the further study of personal persistence is the study of thinking—of our distinctive form of consciousness. This diagnosis does not leave us with an easy task. But it does at least leave us with both a diagnosis and a task.

There are other virtues of THINKERS that I won't go into here. And there are other topics and problem cases that I have not mentioned. For example, I have not talked about vagueness, the Problem of the Many, or other issues having to do with composition, philosophy of time, or modality. A lot of this makes sense, since many of these issues concern the underlying ontology of personal persistence, and my view is that we are ignorant of this underlying ontology. The other issues will simply have to wait.

As with any new theory of personal persistence, THINKERS needs further attention and development. New implications and objections will inevitably arise. My goal here has not been to refute all competing theories, nor has it been to show that my theory is completely satisfying in every way. Rather, my goal has been to develop and motivate the kind of theory of personal persistence that I believe can deal with the major hurdles brought out in this dissertation and elsewhere. And I believe that THINKERS is up to that task.

CONCLUSION

In this dissertation I first argued that the self shows up in experience and that we can be directly aware of ourselves, and I then used these results to rule out various theories of personal persistence and to argue that we are thinkers whose persistence is a matter of continuity in the capacity for thinking. By way of conclusion, I will briefly suggest some directions for future research on these topics.

There are several very general topics brought up in this dissertation that deserve further attention, both because this would make my arguments richer and more compelling, and because my arguments may shed some light on these topics. For example, there is much more to say about introspection and acquaintance in light of this dissertation. And, as I've suggested, further research on thinking—especially the ontology of thinking—will be crucial to developing the ideas I've discussed. There is also more to say about personal identity, especially as it relates to philosophy of time, vagueness, modality, composition, and ethics. All of these very general topics deserve further attention.

But that goes without saying. So let me now suggest a few specific directions for further research. I will start with two larger projects that I see developing from this dissertation. The first is a full-fledged account of self-awareness. I have argued that we can be, and often are, directly aware of ourselves (Ch. 2). And I suggested that this thesis could be developed into a promising account of self-awareness (Ch. 2, §4). But I have not yet done that. I have not brought the above thesis to bear on many of the issues that are central to discussions of self-awareness. A theory of self-awareness should explain how it

is that I can be aware of myself in a way that no one else is aware of me, and how it is that I can entertain thoughts about myself in a way that no one else can. Yet it should also explain how it is that I can communicate to others what I am thinking about myself. This sort of communication involves indexicals like 'I' and 'me'. So a theory of self-awareness should explain certain facts about indexicals, such as their unique semantic and epistemic roles in self-reference and self-awareness, and their role in motivating action. A theory of self-awareness should also explain how it is that I can be certain that I exist right now and why I am immune to certain errors of self-misidentification. But it should do all of this without threatening the idea that my thoughts about myself are cognitively significant. These are the main explanatory criteria that drive discussions of self-awareness.[125] And I believe that this dissertation can shed light on many (if not all) of them. For example, since I am the only person who is acquainted with me, the fact that I am acquainted with myself can make sense of the fact that I can be aware of and think about myself in a way that no one else can be aware of or think about me. Also, in Chapter 1 I suggested how immunity to error through misidentification might be handled by an account like mine (§3.2), and Chapter 2 shows how the fact that I am self-acquainted can make sense of the fact that I can be certain that I exist right now. Yet there is also room here for a theory of indexicals—one whereby 'I' refers directly (and rigidly) to the person with whom one is acquainted—that I believe can explain the various unique roles of indexicals. This is just a sketch. But I am optimistic about the prospects of an account of self-awareness developing from this dissertation.

Another larger project that I see developing from this dissertation concerns the structure of thinking. Two specific features that I have in mind are *subjectivity* and the

[125] For an overview of these criteria, see Howell (2006).

unity of thinking. That consciousness is essentially subjective is not a new idea. But the results of this dissertation may help us better understand that idea by laying the groundwork for a model of consciousness according to which the subject is an essential part of consciousness. There's also more to be said on the unity of thinking. For example, although I discussed the *synchronic* unity of thinking—the unity of thinking at a particular time—I did not discuss the *diachronic* unity of thinking, which concerns the way in which a person's experiences form into a single, unified stream across time. Research on this phenomenon will help in developing the notion of continuity in thinking introduced in Chapter 5.

In addition to the two larger projects just mentioned, there are some other specific topics that deserve further attention in light of this dissertation. First, in Chapter 1 I briefly discussed the character of self-experience. More could be said on this topic. One way to further explore the character of self-experience is to look at research on other mental disorders that may affect self-experience, such as Dissociative Identity Disorder, Anarchic Hand Syndrome, Cotard Syndrome, and Depersonalization Disorder.

Second, in Chapter 2 I made use of The Doubt Test to argue that we are acquainted with ourselves. And one of my goals in doing so was to make the features of and rationale for this commonly used test explicit in a way that many authors who use the test do not. There are other arguments for acquaintance that could also use some fleshing out. One example is Bertrand Russell's (1912) regress argument for the claim that we are acquainted with our experiences. It goes something like this: We must be directly aware of *something*. Otherwise we would face an infinite regress. For suppose that I am not directly aware of anything. Then when I am aware of some x, it is in virtue of being

aware of some distinct y. And since I am not directly aware of y either (and presumably not aware of y in virtue of being aware of x), there must be some distinct z in virtue of which I am aware of y. But then of course there has to be something else in virtue of which I am aware of z … and so on. Thus, if I am not directly aware of anything, then whenever I am aware of something, I am the subject of infinite awarenesses of infinitely many things. But that's absurd. So I must be directly aware of something. And what is that something? Well, the best candidate seems to be my mind—specifically, my *experiences*. Thus, I must be directly aware of my experiences. As stated, this argument is flawed, since it's not obvious that indirect awareness must involve an intervening *awareness* of some distinct thing (rather than just an intervening causal process, for example). Nonetheless, it would be interesting to see whether this argument can be revived in some form.

Finally, in Chapter 3 I broached various topics having to do with the relationship between experience and time. There are two particular points about this relationship that strike me as especially worthy of further discussion. One has to do with *experience* as it relates to time. The question is: How are we to understand the experience of temporal passage? One view is that the experience of time is the experience of a series of instants. I've argued that this view is mistaken (Ch. 3, §2.2). Another view—the retentional model—says that each experience contains two components: an instantaneous present component and a component that represents the recent past. This view also has problems. One problem that I introduced in Chapter 3 (§2.2) stems from the fact that since your seeming to think '2+2=4' *just is* your thinking '2+2=4', there isn't any distinction between your thinking '2+2=4' and your entertaining a representation of the thought,

'2+2=4'. So, on the retentional model, it would seem that when you think '2+2=4' you actually think *two* thoughts—one spread out in time and one instantaneous. Or, in fact, it seems that you think *many* thoughts, since each successive instant over a certain period of time will contain a representation of your thought which itself counts as a thought. As I suggested in Chapter 3, this feels like a bad result. One final view—the extensional model—says that each experience is temporally extended and experienced as a unified whole. This is the view that I tentatively favor. But there's a lot more work to be done here.

The other point about experience and time that strikes me as especially noteworthy has to do with *time* as it relates to experience. One argument for The A-Theory of time, which is the view that there is an ontologically privileged present, is based on our experience of temporal passage—that is, our experience of a present coming and going.[126] It is not obvious to me that anything I've said in this dissertation provides any new support or trouble for this argument. Nonetheless, I believe that further research on this topic in tandem with research on the nature of the experience of time may pay dividends here.

And there's a lot more. My dissertation is at a very philosophically fertile nexus of metaphysics, philosophy of mind, and epistemology. The above research ideas hardly scratch the surface. There's plenty left to explore. I hope to have made a modest contribution to these topics. And I relish the prospect of doing more. Working on this dissertation has caused me to consider and wrestle with many interesting ideas, and it has raised many new questions in my mind. I hope reading it has done the same for you.

---

[126] For discussions of this argument, see Paul (2010), Maudlin (2007, p. 135, 142), Skow (2009), and Hare (2010).

REFERENCES

Akelaitis, A. (1944). "A Study of Gnosis, Praxis and Language Following Section of the Corpus Callosum and Anterior Commissure," *Journal of Neurosurgery*, 1, p. 94-102.

Alston, William (1971). "Varieties of Privileged Access." *American Philosophical Quarterly*, 8, 3, p. 223-241.

Anscombe, G. E. M. (1994). "The First Person" in Q. Cassam (ed.) *Self-Knowledge*. Oxford University Press, p. 140-159.

Armstrong, David (1968). *A Materialist Theory of Mind*. New York: Routledge and Kegan Paul.

Armstrong, David M. (1980). "Identity Through Time." In Peter van Inwagen (ed.), *Time and Cause.* Dordrecht: D. Reidel, p. 67-78.

Ayer, A. J. (1956). *The Problem of Knowledge*. London: Macmillan.

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

Balog, Katalin (2012). "Acquaintance and the Mind-Body Problem" in S. Gozzano and C. Hill (eds.) *New Perspectives on Type Identity: The Mental and the Physical*. New York: Cambridge University Press, p. 16-42.

Bayne, Tim (2004). "Self-Consciousness and the Unity of Consciousness." *The Monist*, 87, 2, p. 224-241.

Bayne, Tim (2008). "The Phenomenology of Agency." *Philosophy Compass*, 3, p. 1-21.

Bayne, Tim (2008). "The Unity of Consciousness and the Split-Brain Syndrome," *The Journal of Philosophy*, 105, 6, p. 277-300.

Bayne, Tim and Chalmers, David (2003). "What is the Unity of Consciousness?" in A. Cleeremans (ed.) *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford University Press, p. 23-58.

Bermúdez, José Luis (1999). "Self-as-Subject and Self-as-Object: Reply to Brook." *The Field Guide to the Philosophy of Mind* (symposium on Bermúdez, 1998). Available at http://host.uniroma3.it/progetti/kant/field/bermudezsymp_replytobrook.htm.

Billon, Alexander (2014). "Why Are We Certain that We Exist?" *Philosophy and Phenomenological Research*, 88, 3, p. 1-37.

Blakemore, Sarah-Jane (2000). "Monitoring the Self in Schizophrenia: The Role of Internal Models" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 185-202.

Bleuler, E. (1950). *Dementia Praecox or the Group of Schizophrenias*. International University Press.

Bogen, J. E. (1993). "The Callosal Syndromes" in K. H. Heilman and E. Valenstein (eds.) *Clinical Neuropsychology*. Oxford University Press, p. 337-407.

Bonjour, Laurence (1985). *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.

Bonjour, Laurence (1999). "Foundationalism and the External World." *Philosophical Perspectives*, 13, p. 229-249.

Brentano, Franz (1973). *Psychology from an Empirical Standpoint*. A. C. Rancurello, D. B. Terrell, and L. McAlister (trans.). London: Routledge.

Butler, Joseph (2008). "Of Personal Identity." in J. Perry (ed.), *Personal Identity*. Los Angelis: University of California Press, p. 99-105; originally an appendix to *The Analogy of Religion* (1736).

Campbell, John (2002). "The Ownership of Thoughts." *Philosophy, Psychiatry, and Psychology*, 9, 1, p. 35-39.

Carruthers, Glen (2007). "A Model of the Synchronic Self." *Consciousness and Cognition*, 15, p. 533-550.

Carruthers, Peter and Veillet, Benedicte (2011). "The Case Against Cognitive Phenomenology" in T. Bayne and M. Montague (eds.) *Cognitive Phenomenology*. Oxford University Press, p. 35-56.

Carruthers, Peter. (2011). *The Opacity of Mind*. Clarendon: Oxford University Press.

Cassam, Quassim (1994). *Self-Knowledge*. Oxford University Press.

Casullo, Albert (1988). "A Fourth Version of the Bundle Theory." *Philosophical Studies*, 54, 1, p. 125-139.

Chalmers, David (1996). *The Conscious Mind*. Oxford University Press.

Chalmers, David (2003). "The Content and Epistemology of Phenomenal Belief." *Consciousness: New Philosophical Perspectives*. Oxford University Press, p. 1-54.

Chalmers, David (2006). "The Foundations of Two-Dimensional Semantics." *Philosophical Studies*, 118, p. 153-226.

Chisholm, Roderick (1957). *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.

Chisholm, Roderick (1969). "On the Observability of the Self." *Philosophy and Phenomenological Research*, 30, p. 7-21.

Chisholm, Roderick (1973). "Parts as Essential to Their Wholes." *Review of Metaphysics,* 26, p. 581–603.

Chisholm, Roderick (1975). "Mereological Essentialism: Further Considerations." *Review of Metaphysics,* 28, p. 477–84.

Chisholm, Roderick (1976). *Person and Object: A Metaphysical Study*. La Salle, Illinois: Open Court Publishing Co.

Chisholm, Roderick (1989). *Theory of Knowledge, 3rd ed*. New Jersey: Prentice Hall.

Coliva, Annalisa (2002). "Thought Insertion and Immunity to Error Through Misidentification." *Philosophy, Psychiatry, and Psychology*, 9, 1, p. 27-34.

Dainton, Barry (2008). "The Experience of Time and Change," *Philosophy Compass*, 3, 4, p. 619-638.

Dainton, Barry (2008). *The Phenomenal Self*. Oxford University Press.

Dainton, Barry and Bayne, Time (2005). "Consciousness as a Guide to Personal Persistence," *Australasian Journal of Philosophy*, 89, 4, p. 549-571.

Davis, L. (1997). "Cerebral Hemispheres," *Philosophical Studies*, 87, p. 207-222.

Dent, Kevin, Catling, Jonathan C., and Johnston, Robert A. (2007). "Age of Acquisition Affects Object Recognition: Evidence from Visual Duration Thresholds." *Acta Psychologica*, 125, p. 301-318.

Descartes, Rene (1643/1993). *Meditations on First Philosophy, 3rd ed*. D. A. Cress (trans.). Indianapolis: Hackett Publishing Company.

Efron, Robert (1970). "The Minimum Duration of a Perception." *Neuropsychologia*, 8, 57, p. 57-63.

Eilan, Naomi (2000). "On Understanding Schizophrenia" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 97-114.

Eklund, Matti (2004). "Personal Identity, Concerns, and Indeterminacy." *The Monist*, 87, 4, p. 489-511.

Evans, Gareth (2001). "Self-Identification" in A. Brook and R. DeVidi (eds.) *Self-Reference and Self-Awareness*. Philadelphia: John Benjamins Publishing Company, p. 95-142.

Ewing, A. C. (1980). *The Fundamental Questions of Philosophy*. Routledge.

Fergusen et al. (1985). "Neuropsychiatric Observation on Behavioral Consequences of Corpus Collosum Section for Seizure Control" in A. G. Reeves (ed.) *Epilepsy and the Corpus Callosum*. New York: Plenum, p. 501-514.

Fish, F. J. (1984). *Fish's Schizophrenia, 3rd ed*. M. Hamilton (ed.). Wright.

Fish, F. J. (1985). *Clinical Psychopathology: Signs and Symptoms in Psychiatry*. M. Hamilton (ed.). Wright.

Frankfurt, Harry (1988). *The Importance of What We Care About*. Cambridge University Press.

Frith, Christopher (1992). *The Cognitive Neuropsychology of Schizophrenia*. New Jersey: Lawrence Erlbaum Associates.

Frith, Christopher and Done, D. J. (1989). "Experiences of Alien Control in Schizophrenia Reflect a Disorder in the Central Monitoring of Action." *Psychological Medicine*, 19, 2, p. 359-363

Frith, Christopher and Johnstone, Eve (2003). *Schizophrenia: A Very Short Introduction*. Oxford University Press.

Fulford, K. W. M. (1989). *Moral Theory and Medical Practice*. Cambridge University Press.

Fumerton, Richard (1995). *Metaepistemology and Skepticism*. Lanham, MD: Rowman and Littlefield.

Fumerton, Richard (2005). "Speckled Hens and Objects of Acquaintance," *Philosophical Perspectives*, 19, p. 121-139.

Gallagher, Shaun (2000). "Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error Through Misidentification" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 203-241.

Gallagher, Shaun and Marcel, Anthony J (1999). "The Self in Contextualized Action" in S. Gallagher and J. Shear (eds.) in *Models of the Self*. Imprint Academic, p. 273-300.

Gallagher, Shaun and Shear, Jonathan (1999) (eds.). *Models of the Self*. Imprint Academic.

Gasser, Georg and Stefan, Matthias (2012). "Introduction" in G. Gasser and M. Stefan (eds.) *Personal Identity: Complex or Simple?* Cambridge University Press, p. 1-18.

Gazzaniga, M. S. (2000). "Cerebral Specialization and Interhemispheric Communication: Does the Corpus Callosum Enable the Human Condition?" *Brain*, 123, p. 1293-1336.

Gazzaniga, M. S. and LeDoux, J. (1978). *The Integrated Mind*. New York: Plenum.

Gendler, T. S. (1998). "Exceptional Persons: On the Limits of Imaginary Cases," *Journal of Consciousness Studies*, 5, p. 592-610.

Gertler, B. (2001). "Introspecting Phenomenal States." *Philosophy and Phenomenological Research*, 63, 2, p. 305-328.

Gertler, Brie (2011). *Self-Knowledge*. New York: Routledge.

Gertler, Brie (2012). "Renewed Acquaintance" in D. Smithies and D. Stoljar (eds.) *Introspection and Consciousness*. Oxford University Press, p. 89-123.

Gibbs, Paul J. (2000). "Thought Insertion and the Inseparability Thesis." *Philosophy, Psychiatry, and Psychology*, 7, 3, p. 195-202.

Goldman, Alvin (1997). "Science, Publicity, and Consciousness," *Philosophy of Science*, 64, p. 525-545.

Graham, George (2002). "Recent Work in Philosophical Psychopathology." *American Philosophical Quarterly*, 39, 2, p. 109-126.

Grice, Paul (2008). "Personal Identity," in J. Perry (ed.) *Personal Identity*. Berkeley: University of California Press.

Hacking, Ian (2000). *The Social Construction of What?* Harvard University Press.

Hamilton, William (1860). *Lectures on Metaphysics and Logic*. H. L. Mansel (ed.). Oxford University Press.

Hare, Caspar (2010). "Realism About Tense and Perspective," *Philosophy Compass*, 5, 9, p. 760-769.

Hinton, J. M. (1973). *Experiences*. Oxford University Press.

Hoffman, R. (1986). "Verbal Hallucinations and Language Production Processes in Schizophrenia." *Behavioral and Brain Sciences*, 9, p. 503-517.

Howell, Robert (2006). "Self-Knowledge and Self-Reference." *Philosophy and Phenomenological Research*, 72, p. 44-70.

Howell, Robert (2010). "Subjectivity and the Elusiveness of the Self." *Canadian Journal of Philosophy*, 40, 3, p. 459-484.

Hume, David (1739/1975). *A Treatise of Human Nature, 2nd ed*. L. A. Shelby-Bigg (ed.). Oxford: Clarendon Press.

Jackson, Frank (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.

Johnston, Mark (1987). "Human Beings." *Journal of Philosophy,* 84, p. 59-83.

Kanzian, Christian (2012). "Is 'Person' a Sortal Term?" in G. Gasser and M. Stefan (eds.) *Personal Identity: Complex or Simple?* Cambridge University Press, p. 192-205.

Keesey, U. T. (1972). "Flicker and Pattern Detection: A Comparison of Thresholds." *Journal of the Optical Society of America*, 62, p. 444-448.

King-Smith, P. E. and Kulikowski, J. J. (1975). "Pattern and Flicker Detection Analyzed by Subthreshold Summations." *Journal of Physiology*, 249, p. 519-548.

Kriegel, Uriah (2004). "Consciousness and Self-Consciousness." *The Monist*, 82, 2, pp. 182-205.

Kulikowski, J. J. and Tolhurst D. J. (1973). "Psychophysical Evidence for Sustained and Transient Detectors in Human Vision." *Journal of Physiology*, 232, p. 149-162.

Langsam, Harold (2001). "Pain, Personal Identity, and the Deep Further Fact." *Erkenntnis*, 54, p. 247-271.

Langsam, Harold (2002). "Consciousness, Experience, and Justification." *Canadian Journal of Philosophy*, 32, p. 1-28.

Legge, Gordon E. (1978). "Sustained and Transient Mechanisms in Human Vision: Temporal and Spatial Properties." *Vision Research*, 18, 1, p. 69-81.

Leibniz, G. W. (1765/1953). *New Essays on Human Understanding*, P. Remnant and J. Bennett (trans. and eds.). Cambridge University Press.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. Le Salle: Open Court.

Lewis, David (1976). "Survival and Identity" in A. O. Rorty (ed.) *The Identities of Persons*. Los Angeles: University of California Press, p. 17-40.

Linn, E. (1977). "Verbal Auditory Hallucinations: Mind, Self, and Society." *Journal of Nervous and Mental Disease*, 104, p. 8-17.

Locke, John (1689/1975). *An Essay Concerning Human Understanding*. P. H. Nidditch (ed.). Oxford University Press.

Lowe, E. J. (2009). *More Kinds of Being: A Further Study of Individuation, Identity, and the Logic of Sortal Terms*. Oxford: Wiley-Blackwell.

Lycan, William (2009). "Giving Dualism Its Due," *Australasian Journal of Philosophy*, 87, 4, p. 551-563.

MacIntyre, Alasdair (1984). *After Virtue, 2nd ed*. University of Notre Dame Press.

Madell, Geoffrey (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.

Malbranche, Nicolas (1674/1997). *The Search After Truth*. T. M. Lennon and P. J. Olscamp (eds.). Cambridge University Press.

Malcom, Norman (1975). *Knowledge and Certainty: Essays and Lectures*. Cornell University Press.

Malenka, R. C., Angel, R. W., Hampton, B., and Berger, P. A. (1982). "Impaired Central Error-Correcting Behavior in Schizophrenia." *Archive of General Psychiatry*, 39, 1, p. 101-107.

Marcel, Anthony (2003). "The Sense of Agency: Awareness and Ownership of Action" in J. Roessler and N. Eilan (eds.) *Agency and Self-Awareness: Issues in Philosophy and Psychology*. Oxford University Press, p. 48-93.

Marks, C. (1981). *Commissurotomy, Consciousness, and Unity of Mind*. Cambridge: The MIT Press.

Martin, M. G. F. (2002). "The Transparency of Experience." *Mind and Language*, 17, p. 376-425.

Maudlin, Tim (2007). *The Metaphysics Within Physics*. Oxford University Press.

McDowell, John (1994). *Mind and World*. Cambridge: Harvard University Press.

McGinn, Colin (1983). *The Subjective View*. Oxford: Clarendon Press.

Mellor, C. H. (1970). "First Rank Symptoms of Schizophrenia." *British Journal of Psychiatry*, 117, p. 15-23.

Merricks, Trenton (1998). "There Are No Criteria of Identity Over Time." *Noûs*, 32, 1, p. 106-124.

Merricks, Trenton (2001). "How to Live Forever Without Saving Your Soul" in K. Corcoran (ed.) *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*. Ithaca: Cornell University Press, p. 183-200.

Merricks, Trenton (2001). *Objects and Persons*. Oxford: Clarendon Press.

Modell, A. H. (1960). "An Approach to the Nature of Auditory Hallucinations in Schizophrenia." *Archives of General Psychiatry*, 3, p. 259-266.

Moor, J. (1982). "Split-Brains and Atomic Persons," *Philosophy of Science*, 49, p. 91-106.

Nagel, Thomas (1971). "Brain Bisection and the Unity of Consciousness," *Synthese*, 22, 3, p. 396-413.

Nagel, Thomas (1986). *The View From Nowhere*. Oxford University Press.

Noonan, Harold (2003). *Personal Identity, 2nd ed*. New York: Routledge.

Nozick, R. (1981). *Philosophical Explanations*. Oxford University Press.

O'Brien, Lucy (2007). *Self-Knowing Agents*. Oxford University Press.

Olson, Eric (2003). "An Argument for Animalism" in R. Martin and J. Barresi (eds.) *Personal Identity*. Blackwell, p. 318-334.

Olson, Eric (2007). *What Are We?* Oxford: Clarendon Press.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Parnas, Josef (2000). "The Self and Intentionality in the Pre-psychotic Stages of Schizophrenia: A Phenomenological Study" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 115-148.

Paul, L. A. (2010). "Temporal Experience," *The Journal of Philosophy*, 106, 7, p. 333-359.

Perry, John (1976). "The Importance of Being Identical," in A. O. Rorty (ed.) *The Identity of Persons*. Los Angeles: University of California Press, p. 67-90.

Perry, John (2008). "Personal Identity, Memory, and the Problem of Circularity," in J. Perry (ed.) *Personal Identity, 2nd ed*. Los Angeles: University of California Press, p. 135-155.

Price, H. H. (1932). *Perception*. London: Methuen.

Prinz, Jesse (2011). "The Sensory Basis of Cognitive Phenomenology." in T. Bayne and M. Montague (eds.) *Cognitive Phenomenology*. Oxford University Press, p. 174-196.

Quinton, Anthony (1962). "The Soul," *The Journal of Philosophy* 59, p. 393-409.

Radden, Jennifer (1999). "Pathologically Divided Minds, Synchronic Unity and Models of Self" in S. Gallagher and J. Shear (eds.) in *Models of the Self*. Imprint Academic, p. 343-358.

Reed, Baron (2002). "How to Think About Fallibilism." *Philosophical Studies*, 107, p. 143-157.

Reid, Thomas (1785/2002). *Essays on the Intellectual Powers of Man*. D. Brookes (ed.). Pennsylvania State University Press.

Reid, Thomas (2008). "Of Identity," in J. Perry (ed.), *Personal Identity*. Los Angelis: University of California Press, p. 99-105; originally in *Essays on the Intellectual Powers of Man*. (1855).

Rey, Georges (1976). "Survival" in A. O. Rorty (ed.) *The Identities of Persons*. Los Angeles: University of California Press, p. 41-66.

Rosenthal, David M. (2004). "Being Conscious of Ourselves." *The Monist*, 87, 2, p. 159-181.

Rovane, C. (1998). *The Bounds of Agency*. Princeton University Press.

Rudd, Michael E. (1996). "A Neural Timing Model of Visual Threshold." *Journal of Mathematical Psychology*, 40, p. 1-29.

Russell, Bertrand (1912). *The Problems of Philosophy*. Thornton Butterworth Limited.

Ryle, Gilbert (1949/2002). *The Concept of Mind*. The University of Chicago Press.

Sass, Louis (2000). "Schizophrenia, Self-Experience, and the So-called 'Negative Symptoms'" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 149-182.

Schechtman, Marya (2011). "The Narrative Self" in S. Gallagher (ed.) *The Oxford Handbook of The Self*. Oxford University Press, p. 394-418.

Shoemaker, Sydney (1970). "Persons and Their Pasts," *American Philosophical Quarterly* 7, p. 269-285.

Seymour et al. (1994). "The Disconnection Syndrome: Basic Finding Reaffirmed," *Brain*, 117, p. 105-115.

Shoemaker, David (2002). "The Irrelevance/Incoherence of Non-Reductionism About Personal Identity." *Philo*, 5, 2, p. 143-160.

Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness." *Journal of Philosophy*, 65, 19, p. 555-567.

Shoemaker, Sydney (1985). "Personal Identity: A Materialist Account" in S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Basil Blackwell, p. 67-133.

Shoemaker, Sydney (1994). "Introspection and the Self" in Q. Cassam (ed.) *Self-Knowledge*. Oxford University Press, p. 118-139.

Sider, Theodore (2000). "Recent Work on Identity Over Time." *Philosophical Books*, 41, p. 81-89

Sider, Theodore (2001). *Four-Dimensionalism.* Oxford: Clarendon Press.

Sider, Theodore (2013). "Against Parthood," in K. Bennett and D. Zimmerman (eds.) *Oxford Studies in Metaphysics, vol. 8*. Oxford University Press, p. 237-293.

Sidtis, J. J. (1986). "Can Neurological Disconnection Account for Psychiatric Dissociation?" in J. M. Queen (ed.) *Split Minds/Split Brains: Historical and Current Perspectives*. New York: NYU Press, p. 127-148.

Skow, Bradford (2009). "Relativity and the Moving Spotlight," *The Journal of Philosophy*, 106, 12, p. 666-678.

Snyder, S. (1974). *Madness and the Brain*. McGraw-Hill.

Springer, S. P., and Deutsch G (1998). *Left Brain Right Brain, 5th ed*. New York: Freeman and Co.

Stephens, Lynn G. and Graham, George (1994). "Self-Consciousness, Mental Agency, and the Clinical Psychopathology of Thought Insertion." *Philosophy, Psychiatry, and Psychology*, 1, 1, p. 1-10.

Stephens, Lynn G. and Graham, George (2000). *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge: MIT Press.

Strawson, Galen (1999). "The Self and the SESMET." *Journal of Consciousness Studies*, 6, 4, p. 99-135.

Strawson, Galen (2000). "The Phenomenology and Ontology of the Self" in D. Zahavi (ed.) *Exploring the Self*. Philadelphia: John Benjamins Publishing Company, p. 39-54.

Swinburne (1985). "Personal Identity: the Dualist Theory" in S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Basil Blackwell, p. 1-66.

Swinburne, Richard (2012). "How to Determine Which is the True Theory of Personal Identity" in G. Gasser and M. Stefan (eds.) *Personal Identity: Complex or Simple?* Cambridge University Press, p. 105-122.

Taylor, Charles (1976). "Responsibility for Self" in A. O. Rorty (ed.) *The Identities of Persons*. Los Angeles: University of California Press, p. 281-300.

Thompson, Judith Jarvis (1997). "People and their Bodies." in J. Dancy's *Reading Parfit*. Oxford: Blackwell, p. 202–229.

Unger, Peter (1979). "I Do Not Exist" in *Perception and Identity*, G. F. MacDonald (ed.). Cornell University Press, p. 235-251.

Unger, Peter (1990). *Identity, Consciousness and Value*. Oxford University Press.

Unger, Peter (2006). *All the Power in the World*. Oxford University Press.

Van Cleve, James (1985). "Three Versions of the Bundle Theory." *Philosophical Studies*, 47, p. 95-107.

Van Cleve, James (1986). "Mereological Essentialism, Mereological Conjunctivism and Identity Through Time." in Peter French, Theodore E. Uehling, Jr. and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy XI: Studies in Essentialism.* Minneapolis: University of Minnesota Press, p. 141-156.

Van Inwagen, Peter (1990). *Material Beings.* Ithaca: Cornell University Press.

Velleman, J. David (2006). "The Self as Narrator" in *Self to Self: Selected Essays*. Cambridge University Press, p. 203-223.

Warren, Clive and Morton, John (1982). "The Effects of Priming on Picture Recognition." *British Journal of Psychology*, 73, p. 117-129.

Wilkes, Kathleen (1994). *Real People: Personal Identity Without Thought Experiments*. Oxford University Press.

Williams, Bernard (1973). *Problems of the Self.* Cambridge: Cambridge University Press.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.

Wing, J. K. (1978). *Reasoning About Madness*. Oxford University Press.

Wolford et al. (2004). "Split Decisions" in M. S. Gazzaniga (ed.) *The Cognitive Neurosciences III*. Cambridge: The MIT Press, p. 1189-1200.

Zaeidel, D (1994). "A View of the World from a Split-Brain Perspective" in E. M. R. Critchley (ed.) *The Neurological Boundaries of Reality*. London: Farrand Press, p. 161-174.

Zahavi, Dan (2000) (ed.). *Exploring the Self*. Philadelphia: John Benjamins Publishing Company.

Zimmerman, Dean (1998). "Criteria of Identity and the 'Identity Mystics'," *Erkenntnis*, 48, p. 281-301.

Zimmerman, Dean W. (1995). "Theories of Masses and Problems of Constitution." *Philosophical Review* 104, p. 53–110.

Zubin, J. (1985). "Negative Symptoms: Are They Indigenous to Schizophrenia?" *Schizophrenia Bulletin*, 11, p. 461-469.