PERSONAL IDENTITY AND MORAL CHARACTER: HOW A NECESSARY MORAL PERSISTENCE CONDITION MIGHT BE POSSIBLE

by

CHRIS LAY

(Under the Direction of Rene Jagnow and Sarah Wright)

ABSTRACT

When discussing how it is that we persist through time, most philosophers have, at best, an incidental place for things like moral character traits; if these traits have any part to play in determining whether one of us persists at all, it is as just one among a whole host of mental features. However, there is a tension between the sparse constitutive role of the moral and the fact that moral concerns are nonetheless regularly invoked both as motivating factors in formulating persistence theories and as practical implications that result from those theories. In other words, persistence thinkers find themselves in a strange position whereby the moral is inextricably connected to the project of persistence but is not a part of the actual metaphysics of persisting. In this dissertation, I investigate if and how a moral persistence condition might be possible. Based on empirical evidence that non-philosophers tend to intuitively prioritize moral features in determining when one of us persists, I argue that a necessary moral persistence condition is plausible. I then propose a novel ontological foundation that could support a moral persistence condition and posit a modification of Parfit's Narrow Psychological View of persistence that takes continuity of moral features as a necessary condition for persistence.

INDEX WORDS: Personal identity; Persistence; Personal ontology; Philosophy of mind

PERSONAL IDENTITY AND MORAL CHARACTER: HOW A NECESSARY MORAL PERSISTENCE CONDITION MIGHT BE POSSIBLE

by

CHRIS LAY

B.S., The University of the Cumberlands, 2008

M.A.T., The University of the Cumberlands, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

PERSONAL IDENTITY AND MORAL CHARACTER: HOW A NECESSARY MORAL PERSISTENCE CONDITION MIGHT BE POSSIBLE

by

CHRIS LAY

Major Professors: Rene Jagnow
Sarah Wright

Committee: Beth Preston
Piers Stephens

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2018

DEDICATION

To Ashley, for her years of support and love and for her willingness to uproot our lives just so that I could write this.

ACKNOWLEDGEMENTS

Many thanks to my advisors Rene Jagnow and Sarah Wright, without whose help I certainly could not have completed this dissertation, but who also have supported me in offering exceedingly helpful comments on other articles for publication and conference presentations and given me constant direction in my academic career. My gratitude also goes to out the University of Georgia Philosophy department generally, but especially to Beth Preston, Thanassis Samaras, and Piers Stephens, each of whom has been an incredible resource for advice and encouragement.

TABLE OF CONTENTS

Page

CHAPTER 1

NO ROOM FOR A MORAL PERSISTENCE CONDITION

This dissertation is about personal identity. More specifically, I am concerned in this dissertation with laying out a theory of identity over time, or diachronic identity. This topic has enjoyed a good deal of scholarly attention since the middle of the twentieth century or so, but nearly everyone who talks about it has a different way of analyzing just what ‘identity over time’ means. Eric Olson (1997) cashes out ‘identity over time’ as ‘persistence’ and Marya Schechtman (1996) as ‘reidentification’. That is, in the former case, ‘identity over time’ comes to mean ‘the conditions under which we can say that a subject persists’. In the latter case, it means ‘the criteria used to properly reidentify some subject as the same subject’. For Richard Swinburne, when I ask ‘what is identity over time’, what I want to know is “the logically necessary and sufficient conditions for a person P_2 at a time t_2 [to be] the same person as a person P_1 at an earlier time t_1, or, loosely, what does it mean to say that P_2 is the same person as P_1?” (1984, 3). John Perry says that the issue is about specifying “the unity relations between...nonsimultaneous person-stages” (1976, 11). And of course, this is hardly a problem limited to professional philosophers. Taking note of the core of the problem, a distraught Alice—fresh down the rabbit-hole and now growing exceedingly tall—wonders, “Let me think: was I the same when I got up this morning? I almost think I can remember feeling a little different. But if I’m not the same, the next question is, Who in the world am I? Ah, that’s the great puzzle!” (Carroll, 1998, 19).

Each of these descriptions of identity over time amount to slightly different ways of describing the same problem, but none of them seems to me to be especially straightforward. And I think that this in part accounts for the confusion surrounding identity over time as a subject of philosophical discourse. It is not just that each philosopher has a different means of solving the problem of identity over time; rather, they all seem to have a subtly different way of stating the problem, too. So, for my project to make any kind of sense, it would be good to formulate the problem of identity over time in a way that both is uncomplicated and captures the common element about identity over time that each of these varied descriptions tries to represent. To my mind, the clearest way of describing the problem is as the need to identify that something—whatever it is—that, in Schechtman's words, "*makes* someone the same person at...two times" (1996, 8).

This description might not actually be as simple as it at first appears. It might be thought that Schechtman's formulation blends personal identity with what Judith Jarvis Thomson (1997) calls 'personal ontology'—the identification of what kind of things we are. So, if Schechtman's formulation is to apply to *us*, it already declares that *we* are a certain sort of thing: persons. Olson (1997) contends that framing the question as 'what does it take for a person to be the same person over time' seems to assume 'person essentialism', the view that we are most fundamentally persons. In other words, it takes for granted that something which is a person will *always* be a person. To see why assuming person essentialism is problematic to giving a clear and non-question-begging description of identity over time, consider the traditional Lockean definition of 'person'. On Locke's account, a person is something that is rational, self-conscious, and feeling. Once we note that these features of personhood are all psychological features, Olson's worry should become apparent. If our description of identity over time assumes person essentialism, and if persons are defined according to psychological features, then ‘what makes someone the same at two different times’ will surely be a matter of strictly psychological features.

Of course, one could resist this conclusion by just denying the Lockean definition of persons. One could, say, follow David Wiggins in thinking that persons must be animals with certain attributes (1980), or accept his stronger claim that the individuation criteria for ‘person’ and ‘human being’ coincide (2001)—meaning that all and only humans are persons. But the trouble here is that an assumption of person essentialism stacks the metaphysical deck against a number of possible answers to the problem posed by identity over time. One could not, for instance, hold both the Lockean definition of person and that the conditions for identity over time are biological instead of psychological. And again, this is because what makes something a person on the Lockean view is a matter of psychological features, not biological ones. (For the same reasons, the assumption of person essentialism would also preclude some separately existing substance—like a soul—from being the condition of identity over time.) Given these considerations, a more ontologically neutral description of identity over time would be best.

The upshot is that perhaps part of the reason for such a range of descriptions of identity over time is worry over illicitly front-loading one’s statement of the problem. It thus looks like one’s take on personal ontology—and if we are essentially or only contingently persons—bears directly on identity over time. This makes sense. What kind of things we are ought to determine the conditions for our identity over time. As Olson (2007) suggests, it would then behoove the philosopher to first take seriously her position on personal ontology before putting together a theory of identity over time. Since I have *not* yet considered personal ontology at this point, I will hereafter in this chapter use the ontologically neutral term ‘subject’ instead of person. I say more about the personal ontology / personal identity relationship in Chapter 3. Introducing the notion here serves only to observe that one must be careful in laying out what identity over time is to mean.

Now, there are other questions about identity than just identity over time, and these are regularly conflated. In particular, the question of what is to count as evidence of identity often gets subsumed into the question of the conditions for identity over time. This is unfortunate, as these questions are quite distinct. To ask what counts as evidence of identity is to ask an epistemological question: *how can we know* that someone is one and the same subject at two different times? On the other hand, the question of identity over time as I have formulated it is about *what makes* someone the same subject at two different times. Rather than the epistemic, the philosopher interested in identity over time is concerned, as Harold Noonan points out, “with the constitutive, the metaphysical-cum-semantic, not the evidential, criterion of personal identity” (1991, 2). It might turn out that what makes someone the same subject at two different times is not something easily verifiable, but this is a problem of evidence and not with the conditions of identity over time themselves.

Taking all of this together, I will in this dissertation consider identity over time in the following way. I propose an account that takes moral character to be necessarily partly constitutive of identity. For structural purposes and because I think this way of cashing out identity over time is fairly uncomplicated, I help myself to the term ‘persistence’ to hereafter express the metaphysical problem of identity over time. In what follows, I develop an account of the necessary and sufficient conditions by which one of us persists over time as one and the same numerical subject (whatever sort of thing this subject is—and so remaining neutral in this definition over whether or not we are or must always stay persons). Since I argue that among the aforementioned conditions is a necessary moral persistence condition, this means is that I think a necessary condition for one of us to persist over time as the same subject is sameness of moral character traits (with certain exceptions). One significant advantage of this account, and a central focus of this project, is that it satisfies widespread folk intuitions about the connections between moral character and identity.

First, however, I will in this chapter situate my account in the persistence literature. A substantial reason for this is to show that the current literature cannot properly accommodate a necessary moral persistence condition. But I also want to use this chapter to call attention to a version of the psychological approach—Derek Parfit's—that *can*, I later argue, be appropriately modified to accommodate such a condition. Following a longer section on the psychological approach, I briefly present several alternative persistence theories: the somatic approach, the substance approach, and nihilism about persistence. Then, in a final section, I identify a gap in the literature concerning moral persistence conditions and argue that my project would fill this gap in an important way. Among the dominant theories in the field, only the psychological approach gives any weight to moral character as at all constitutive of persistence. However, even the psychological approach only lets in moral character as an accidental feature. The account I develop instead posits moral character as a necessary persistence condition (though not the sole condition) and is thus a novel contribution to the literature.

1.1 The Psychological Approach

The purpose of this section is to establish the basic contours of the psychological approach. This section will need to be somewhat robust to allow me to later argue both (a) for a psychological account of persistence with a necessary moral persistence condition and (b) that the current state of the psychological approach cannot sustain such an account. In addition, I use this section to introduce a general version of Parfit’s psychological view, which becomes my basis for (a) in Chapter 4. As such, the present section is in three parts. First, I give the basic historical background of the psychological approach found in Locke’s theory of identity, including several important historical criticisms posed against him. Next, I explore several contemporary responses to these criticisms that perhaps let the psychological approach enjoy its current dominance in personal identity theory. Finally, I consider the most important approaches to cashing out ‘psychological continuity’, the principal persistence relation of the psychological approach—including Parfit’s, which I later adopt and modify. Before this, I offer a few general comments about the approach and its appeal.

Broadly, the psychological approach states that persistence of one of us is a matter of persistence of certain psychological states. I say ‘broadly’ because what exactly ‘persistence of certain psychological states’ means will vary among distinct psychological theories, and I will explore the differences between these theories in detail below. The psychological approach is probably the most widely-held persistence theory among Western philosophers. There is also empirical data that suggests it is the more intuitively plausible theory among Westerners more generally (Nichols & Bruno 2010).

This seems to be in large part because of the attractiveness of thought experiments like the transplant case. If my brain/head/cerebrum is transplanted to another body (and another brain/head/cerebrum is transplanted to mine), there is something straightforward and accessible about saying that ‘who I am’ after the operation is the being that first-person remembers my twelfth birthday, recognizes and is worried about my family, and has both my love of cheesy science fiction and my disgust at millipedes (assuming I had all of these things before the operation, of course). In other words, ‘who I am’ is the subject that knows the things I know, cares about the things I care about, and acts like me. For many of us, it is natural to assume that ‘who we are’ has something substantial to do with psychological features that we bear, such as memories, beliefs, desires, and character traits. And so the transplant case brings out such intuitions nicely.

1.1.1 The Traditional View and Some Important Criticisms

While there are many versions of the psychological approach, all owe something to Locke (1975).[1] In the ‘Identity and Diversity’ section of his *Essay*, Locke characterizes persistence as continuity of consciousness, saying:

> [C]onsciousness, as far as ever it can be extended, should it be to Ages past, unites Existences, and Actions, very remote in time, into the same Person, as well as it does the Existence and Actions of the immediately preceding moment: So that whatever has the consciousness of present and past Actions, is the same Person to whom they both belong (340).

Locke is commonly interpreted here as arguing for a *memorial* account of persistence.[2] What makes a subject persist as one and the same subject over time is continuity of consciousness, and Locke most often cashes this out in terms of a subject who remembers some prior actions and endorses them as his own. He illustrates this with the ‘Day Man / Night Man’ example. Starting from the certainly acceptable premise of sleepwalking, Locke has us suppose that a man has two consciousnesses—one active during the day and another only at night. As the Day Man is in no way conscious of the Night Man’s activities (and likewise, for the Night Man) and so neither can remember what the other has done, how can either one be held morally to task for the actions of

[1] Indeed, many adherents of this approach call their theories, collectively, neo-Lockean for this very reason.

[2] Though this interpretation is disputed. See Schectman (1996: 105-12) and Strawson (2011).

the other? Locke thinks it obvious that they cannot. More formally, we can express the Lockean relation as: a subject S_1 at time t_1 persists at t_2 if S_2 at t_2 remembers S_1's action at t_1 as his own.

Perhaps somewhat radically for his time—and certainly breaking ranks with the Cartesians—Locke argues that consciousness and memory have nothing whatsoever to do with immaterial substances (or souls). Although more or less a dualist himself, Locke is quite adamant that one could persist without having one and the same soul. Suppose we have souls and that a man has such a soul that was once in the body of one of the Greek heroes at Troy, Nestor or Thersites.

> But he, now having no consciousness of any of the Actions of either *Nestor* or *Thersites*, does, nor can he, conceive of himself as the same Person with either of them? Can he be concerned in either of their Actions? Attribute them to himself, or think them his own more than the Actions of any other Man, that ever existed? So, that this consciousness not reaching to any of the Actions of either of those Men, he is no more one *self* with either of them, than if the Soul or immaterial Spirit, that now informs him, had been created, and began to exist, when it began to inform his present Body...(339).

A given subject persists as one and the same subject so long as she is self-conscious in the sense of ascribing current and past actions to herself (primarily by remembering *doing* those actions). For Locke, this is true whether the subject has a single soul that has been placed in different bodies throughout her life or whether the subject's consciousness has been realized in multiple different souls. This is why Locke is quick to distinguish between 'person' and 'man'. The former is "a thinking, intelligent Being, that has reason and reflection, and can consider itself as itself" (335) and is "capable of Happiness and Misery" (341). Man, on the other hand, has 'animal identity', which is persistence in virtue of having the same functional organization or the same life. This consists in having "a Body so and so shaped...the same successive Body not shifted all at once" and with "the same immaterial Spirit" (335). So, for Locke, 'Men' and the conscious subjects sometimes realized in them have entirely different persistence conditions.

Joseph Butler and Thomas Reid both raised pointed objections to the Lockean position that shaped later psychological theories. Butler argued that Locke's memory criterion of persistence was circular:

> [O]ne 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 (1736, reprinted in Perry 1975, 100).

What Butler means is that memories presuppose a subject that remembers—a rememberer. So, memories cannot be constitutive of persistence because there must first be a persisting rememberer (someone who was present at the time of the initial experience and who now remembers). The problem is, as Parfit puts it, that "[it] is part of our concept of memory that we can remember only *our own* experiences" (1984, 220).

Reid (1785, reprinted in Perry 1975) presents a different problem: that of the 'brave officer'. We are given three time-slices of a man's life: a boy who is flogged for robbing an orchard, a young officer who performs bravely in battle, and finally an old man reflecting on his life. The young officer remembers the flogging, and the old man remembers being decorated for his valor. However, the old man *does not* remember the childhood flogging. Reid asserts that this generates a contradiction on Locke's memorial account. Given the transitivity of identity, the old man must be the same subject as the boy: the boy is identical to the young officer (as the officer remembers the flogging), and the old man is identical to the officer (as the old man remembers his distinction in battle), so the old man must be identical to the boy, too. But since the old man does not remember his childhood flogging, he cannot on Locke's account be identical with the young boy. Thus Locke's account absurdly both permits and denies that the boy and the old man are identical.

1.1.2 Contemporary Responses to Criticisms of the Traditional View

Neo-Lockeans have offered a few different sorts of solutions to these problems. Taking Butler first, advocates of the psychological approach have tried to define memory in a non-circular way—that is, in a way that does not take for granted the requirement of a persisting rememberer. Perry (2002b) attempts to analyze memory like this and draws from H.P. Grice's (1941) memorial account of persistence, which replaces 'persons' with 'total temporary states' (TTSs), or sets of simultaneous experiences of some one subject. For Grice, a persisting person is just a set of TTSs (not necessarily chronological) in which each TTS contains either (i) a memory of an experience contained in the next TTS, (ii) an experience of which there is a memory contained in the next TTS, or (iii) would contain (i) or (ii) if certain conditions obtain.[3]

But Perry thinks Grice's account still needs some modification, as just switching things to talk about TTSs instead of persons does not save the Gricean from a charge of circularity, on three counts. First, we can imagine a case where two men, Smith and Jones, have a memory of looking at a green cube. Smith actually looked at the cube, but Jones's memory was implanted through hypnotic suggestion. Smith, therefore, is really remembering, but Jones is not. What makes the difference between an apparent memory and a genuine one? Based on the Smith/Jones example, it might seem like the difference is that actually remembering requires that the same subject had the experience in the first place. But if memories of this sort are supposed

[3] Perry adds that Grice's system of TTSs, due to the conditional in (iii), is immune to Reid's 'brave officer' case and would thus be a good answer to that problem, too.

to constitute a person, it is surely circular to presuppose that very person in order to have memories!

A second circularity problem results from the subjunctive ‘would’ in (iii). Any example meant to satisfy this claim would require both some actual TTS and a counterfactual TTS. So, Smith is currently in a dreamless sleep and his TTS contains neither any memories whatsoever nor any experiences that he can later remember. Yet, if we woke Smith up and asked him to recount his trip to Wrigley Field earlier that day, he certainly could do so (he might also be able to later remember the present experience of being woken up in the middle of the night by some rude person curious about his day at Wrigley Field!). Hence, the counterfactual is supposed to show that Smith is presently *capable* of having a TTS that satisfies (i) or (ii), but certain circumstances are preventing that present satisfaction. Perry says that countenancing this conditional in any coherent way would require something constant to hold between the actual event and the counterfactual as a point of comparison. And the only constant that makes sense is Smith—the TTSs in question are both Smith’s: the one he actually had, and the one he *would* have had if woken up. Again, though, we see that the argument presupposes an existing person and so is circular.

Lastly, Perry argues that the phrase ‘certain conditions’ is too vague to be of much help in Grice’s formulation. Without specifying further what sort of conditions he is talking about, Grice nonetheless thinks that there *are* particular conditions that could satisfy (i) and (ii). However, Perry claims that he cannot imagine any sort of condition that would do this without having to invoke the persisting subject as a constant, as in the above objection. The gist of the argument seems to be that the onus would be on Grice to give such a condition. Since he does not do so, Perry sees no way to avoid circularity.

Despite these objections, Perry believes that Grice's idea can be rescued from circularity by a specific analysis of memory. He defines an instance of remembering in the following way (where Perry stipulates that *A* and *B* are 'human beings', but not necessarily persons, as this is to be determined):

A remembers *e* if and only if
(1) *A* represents the past occurrence of an event of type *E*
(2) *B* witnessed *e*; and
(3) *B*'s witnessing of *e* is *M*-related to *A*'s representation of the past occurrence of an event of type *E*

The *M*-relation described in (3) is, for Perry, "a process, the nature of which we do not know" (100)—though it is somehow causal—that leads from an experience (or 'event witnessing') to the representation by a subject of an event. Note that the basic formulation does not stipulate that *A* represent *e* accurately—that is, as the same *type* of event as what *B* witnessed. This allows for the possibility that we may still 'remember' even if the memory and the thing remembered do not always match up exactly. But memory is frequently considered a source of knowledge just because it is often accurate. So, a 'paradigm' instance of remembering, or an instance of remembering from which we can derive knowledge, involves two additional premises:

(4) *e* is of type *E*; and
(5) *A* believes (1)-(4).

By reapplying this to the Gricean argument, Perry thinks that the argument can avoid his three charges of circularity. Recall that the first objection turns on the fact that, supposedly, the only way to account for the fact that Smith actually remembers seeing a green cube and Jones apparently (but not actually) remembers it is that Smith *witnessed* the cube and Jones did not. This is, of course, circular. Perry's analysis of memory accommodates this witnessing condition, but in a noncircular way—it requires only that *someone* (not necessarily the same person) saw the cube and that this witnessing is *M*-related to the later representation of the cube by another someone. In Jones's case, no one saw a cube and this impersonal witnessing condition is not satisfied, so Jones does not remember.

As for the other objections, Perry proposes that we can formulate Grice's argument without the subjunctive conditional at all by using his analysis of memory. This would defang both objections. To do this, we must first recognize that cases in (iii) are cases of possible memory and then cash out 'possible memory' as something like dispositions to remember. Perry analyzes possible memory in this way:

> *A* possibly remembers *e* if and only if
> (1) *A* is disposed to represent the past occurrence of an event of type *E*;
> (2) *B* witnessed *e*; and
> (3) *B*'s witnessing of *e* is *M'*-related to *A*'s being disposed to represent the past occurrence of an event of type *E*

The *M'*-relation in (3) does the same work as the *M*-relation in remembering. It is a nebulous and partly causal process that appropriately links the witnessing condition with the representation condition. As there is no conditional involved, there is no need for a constant that can be a point of comparison between actual and counterfactual events. In addition, Perry's analysis of possible memory—just like his analysis of memory—does not presuppose that the witnessing subject and the subject disposed to represent are the same. Given his analysis of memory and its application to Grice's formulation, Perry contends that it is possible to overcome Butler's objection and make this version of the psychological approach defensible, though he admits that there are probably other psychological relations that matter rather than just memory alone.

Another well-known response to Butler's criticism is offered by Shoemaker (1970). Shoemaker recognizes that memory seems to be a relation between some current cognitive state and a past cognitive and sensory state of some event. In cases of ordinary memory, this relation is a causal one. My past experience causes my current memory of the event. Clearly, this implies the circularity that Butler argues for: if *my* memory is caused by *my* past mental states, then it follows that ordinary memory requires that there must be a single persisting subject that has (or had) both the past and present states involved.

To overcome this, Shoemaker posits the logical possibility of ‘quasi-memories’. As Shoemaker puts the difference, “Whereas someone’s claim to remember a past event implies that he himself was aware of the event at the time of its occurrence, the claim to quasi-remember a past event implies only that someone or other was aware of it” (271). In other words, while my quasi-memory *can* be of an experience *I* had, quasi-memories also allow for the possibility that my quasi-memory is of an experience that is not mine. Like ordinary memories, quasi-memories still stand in an ‘appropriate causal relation’ to some past mental state or another.[4] Shoemaker calls this relation an ‘M-type causal chain’, which he stipulates ought to be as close as possible to the typical causal chain in ordinary memory production while still being compatible with cases in which the mental states involved are not states of the same subject. In this way, ordinary memories are just a subset of quasi-memories demarcated by a specific causal relation: the relation that holds between experiential states and subsequent memories of the same subject. So, memorial accounts of persistence can avoid the circular ‘same subject’ presupposition of ordinary memory by appealing to quasi-memories instead.

Now, a potential worry with substituting quasi-memories for ordinary memories is that quasi-memories may undermine the reasons that philosophers have for proposing memory as constitutive of persistence in the first place. Memories are often thought to give privileged access to experiences ‘from the inside’. It is just this privileged access that makes memory a

[4] Shoemaker also considers the possibility of a weaker sense of quasi-memory that requires only 'correspondence' between memories and past experiences. However, the version of quasi-memory with the causal requirement is the one that Shoemaker takes up for most of the later conclusions he makes in his paper. So, the 'causal' version is the only type of quasi-memory I deal with here.

suitable persistence condition. If I remember something ‘from the inside’, it seems to confirm that I am the same persisting subject (at the very least, such a memory looks like a partial condition of my persistence). At the same time, this fact appears to be reliant on the presupposition of persistence—that it is necessarily one and the same subject who remembers as who did the remembered action. Quasi-memoires introduce the possibility of remembering (more accurately: quasi-remembering) experiences ‘from the inside’ that are not necessarily of one and the same subject. While this distances us from the ‘same subject’ presupposition that regular memory has, it also weakens the privileged access memory gives us to our own persistence. That is, quasi-memories are also ‘from the inside’, but this does not guarantee persistence (even as a partial condition) because I might be quasi-remembering *someone else’s* experience ‘from the inside’.

Shoemaker suggests that this is not as much of a problem as it seems. He argues that the only instances in which quasi-memories could not be constitutive of persistence are in counterfactual ‘branching’ cases—that is, in any case in which a single past mental history can be logically connected through M-type causal chains to two or more simultaneous present states. For example, consider Chisholm’s (1969) ‘amoeba man’. Suppose there were a man capable of dividing into two qualitatively identical parts, like an amoeba. Both resultant men would quasi-remember experiences of the original, pre-fission man. Shoemaker says that we cannot affirm that the original man persists as the two duplicates without giving up some bedrock philosophical assumptions. On the one hand, this scenario violates Leibniz’s Law. The original man has a property that the duplicates lack: *the property of having divided*. On the other hand, this scenario conflicts with the transitivity of identity. The original man is identical with two things that are not identical with each other (as they are two things, not one). Since it is extremely unpalatable to abandon either of these principles, and since choosing one or the other of the duplicates sounds arbitrary, it seems better to say that the original man does not persist—despite the fact that both duplicates share quasi-memories of his experiences. Even if we clench our teeth and call only one of the duplicates identical with the original man, this would still leave one duplicate that has quasi-memories of the original man's experiences but is not identical with him. So, in branching cases—like the fission case—quasi-memories cannot be constitutive of persistence. And while the amoeba man is certainly a bit of science fiction fancy, it is nonetheless logically coherent (and perhaps suggestive of other examples that may, in the near future, strike closer to reality).

The reason that Shoemaker does not see this as much of a problem for quasi-memories and memorial accounts of persistence is that he thinks quasi-memories *are* constitutive of persistence in all non-branching cases.[5] Without branching, both past mental states and current quasi-memories necessarily belong to the same subject. After all, this is just what 'branching' means here: to 'branch' off into another subject like the way a single river splits into two or more distributaries. In other words, non-branching cases of quasi-remembering would also be cases of ordinary remembering. Since ordinary remembering carries with it privileged access to a persisting subject, this access is also preserved for all non-branching cases of quasi-remembering. If Shoemaker is right, the upshot is this: if we substitute as a condition of persistence quasi-memories without branching in place of memory, we can avoid Butler's circularity claim.

Parfit (1984) supposes that there is another way to formulate the circularity objection and, in turn, another solution appropriate to this alternate formulation. It is possible that by his

[5] Moreover, Shoemaker argues that quasi-memories serve as good evidence for persistence. Here again I must distinguish between the metaphysical persistence question and the epistemic question of evidence. I reiterate that my purpose in this project is with the former and not the latter.

objection, Butler meant that memory—and perhaps all experience—presupposes even more than just a persisting thing (the rememberer or experiencer). Rather, one reading of Butler's objection is as a claim that experience presupposes the persistence of an entity that exists apart from the mental and physical facts at a given time. That is, experience presupposes a separate *subject of experience*. This way of thinking is very much in line with the Cartesian cogito. From the proposition 'I think', I can derive that I exist as a sort of subject that has certain experiences: namely, thoughts. Other philosophers deny that there is such a thing as a separately existing subject of experiences but still maintain that talk of experiences must involve something—like reference to a person—that unifies experiences. P.F. Strawson (1966) takes a stance like this, attributing it to Kant. Whether a persisting subject of experience is a separately existing entity or a convenient means of linking mental events, Butler's objection holds firm if all descriptions of experience actually presuppose a subject of experience.

Parfit thinks that what he calls an 'impersonal description' of experience would circumvent the circularity objection by giving a subject-less description of experience. His argument here comes in two parts. First, building on both Locke (1975) and Kant (1964), Parfit argues that an immediate awareness of a subject of experience does not imply the *continued* existence of one and the same subject. As a counterfactual, he imagines a machine that produces an exact Replica of himself. Parfit, on Earth, begins the thought 'Snow is falling' before pushing a button on his Replica Machine and producing an exact Replica of himself on Mars. The Replica continues the thought "It is cold", built from the apparent memories he has of beginning this thought with the phrase 'Snow is falling'. This is although the Replica—who only just came into existence—could not possibly have actually had the previous thought 'Snow is falling'. To Parfit, this shows that our supposed awareness of a single subject of experience is merely awareness of the continuity of certain psychological states in a stream of consciousness. Psychological continuity does not presuppose a single subject, as the 'Replica' example makes clear.

Parfit's second argument comes from Lichtenberg's (1971) objection to Descartes. According to Parfit, Descartes claimed too much in even saying 'I think, therefore I am'. All that really needs to be said is 'It is thought: thinking is going on' (or something like this). Such a description does not (and need not) involve ascribing these thoughts to a thinker. In Parfit's words, "Persons must be mentioned in describing the *content* of countless thoughts, desires, and other experiences. But...such descriptions do not claim that these experiences are *had* by persons" (226). If descriptions of experiences like thoughts do not attribute those experiences to subjects, then it is surely not permissible to derive from these descriptions a separately existing subject doing the thinking, *à la* a Cartesian ego. Taking both of Parfit's arguments together here, an impersonal description of experience neither presupposes a single, continuous subject of experience nor any sort of subject at all (if presented in the right way), and Butler's charge fails.

Those who favor the psychological approach also offer a number of responses to Reid's 'brave officer'. Most of these turn on drawing a distinction between 'connectedness' and 'continuity'. Shoemaker (1984) brings in the terminology of person-stages to explicate this distinction. 'Person-stages' here means something like a 'snapshot' of a person at a single point in time. It is rather like the distinction between diachronic identity—identity over time—and synchronic identity—identity *at* a single time. So, persons over time are composed of various individual person-stages, each of which contains a profile of the individual's psychological character, including memories. Now, let two person-stages be memory-connected if a later person-stage contains first-person memories of experiences from an earlier person-stage. Memory-continuity is achieved across chains of memory-connected person-stages:

> This comes to saying that two stages belong to the same person if and only if they are the end-points of a series of stages such that each member of the series is memory-connected with the preceding member (1984, 81).

In terms of Reid's brave officer, the flogged boy (a person-stage) is memory-connected with the brave officer (another person-stage), who is in turn memory-connected with the old man (a third person-stage). Though the old man is not memory connected with the flogged boy, there is memory-continuity here (and thus persistence) because each person-stage is memory-connected with the subsequent stage, creating an overlapping chain of memory-connected stages. This blunts the force of Reid's objection. On Shoemaker's view, an advocate of the psychological approach is no longer committed to the contradictory position of affirming both that the young boy is and is not identical with the old man.

Parfit (1984) gives a similar answer in terms of chains of psychological connectedness and psychological continuity. Like many other philosophers who adopt the psychological approach, Parfit widens a subject's psychological persistence conditions to include a whole host of psychological features, of which memory is but one among many (and with no real privileged place, at that). 'Connectedness' refers to direct psychological connections between subjects and 'continuity' to overlapping chains of 'strong' connections. Parfit suggests that these connections must be 'strong' in this way because he thinks it would be silly to consider something a persisting, continuous subject who just had a single connection across each overlapping chain. Of strong connections, Parfit says:

> For X and Y to be the same person, there must be over every day *enough* direct psychological connections. Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of connections, over any day, is *at least half*

> the number of direct connections that hold, over every day, in the lives of nearly every actual person (206).

Just as before, this side-steps Reid's objection and possibly in a different way than Shoemaker does. The flogged boy has direct psychological connections to the brave officer. The officer, too, has direct psychological connections with the old man. Assuming there are enough *other* psychological connections between each subject than either a single memory of being flogged as a boy (for the officer) or a single memory of valor on the battlefield (for the old man), the old man can be considered the same persisting individual as the boy. Parfit, though, gives us another way to respond to the objection. Since persistence is more than just a matter of memory connections, it is plausible to say that the old man *could* be directly connected to the flogged boy if the two were strongly connected through other psychological features—say, beliefs, desires, or character traits—whether the old man remembers that he held these features as a boy or not. If the old man and the flogged boy were in this way connected (and thus also continuous), Reid's objection would again be defused.

The point of the previous two subsections has been, respectively, to provide the historical backdrop for contemporary psychological accounts and to highlight how the traditional argument has evolved to meet its objectors. Because these arguments can often be difficult or abstract, I think that this context is important as a way of acquainting the reader with the vocabulary of persistence (and of the psychological approach in particular, which is especially important to my project). With this context behind us, I might now say something about how this traditional argument has splintered off into a plurality of contemporary accounts.

1.2.3 What Psychological Continuity Means

As the psychological approach is probably the most dominant theory of persistence in philosophical circles, it should come as little surprise that there is no ‘one way’ to run a psychological account of persistence. One thing that psychological accounts tend to have in common, though, is the claim that persistence is achieved in virtue of some sort of psychological continuity. Ascertaining the cash value of ‘psychological continuity’ and how this brings about persistence is where the variance in psychological accounts comes in. While I in no way give anything like a full survey of psychological accounts here, I think that it is important to call attention to a range of psychological views to better illustrate that no present psychological account can accommodate a necessary moral persistence condition.

I begin with Parfit’s view, since this is the account I later choose to modify to accommodate a moral persistence condition. Even so, the below is a broader picture of Parfit’s view meant to place it among other theories in the literature, and the brevity shown here harmonizes with the way other theories in the present chapter are presented. In Chapter 4, I present this view in greater detail appropriate to the goals of that chapter, but I did want to here alert the reader to this view’s eventual role in my project and familiarize the reader somewhat with Parfit’s position.

Parfit gives a reductionist account: he thinks that persons just are sets of psychological and possibly physical features—the reason for the ‘possibly’ before ‘physical’ will become clear in a moment. Much in the way that nations are nothing over and above the people and governments that compose them and clubs are nothing over and above their members and the rules that bind them, persons are nothing over and above these features. In part because of this, Parfit argues that persistence is indeterminate. In other words, there will be ‘borderline cases’ when it may be impossible to determine if one and the same subject has persisted or not. This indeterminacy is evident in his definition of psychological continuity, which he calls 'the psychological criterion' of persistence.

> *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4) (1984, 207).

(4) is to protect against the branching involved in, say, fission cases—like the amoeba-man. The interpretation of (3) is where greater or lesser degrees of determinacy can be found. Parfit proposes three ways of taking (3). The Narrow version accepts that the 'right kind of cause' just is the normal cause, much like Shoemaker's M-relation. This typically means that the type of continuity described in the Psychological Criterion is caused through the normal functioning of an ordinary brain. Since 'normal functioning' generally precludes things like swapping out pieces of one brain for another or 'transferring' consciousness to a synthetic brain surrogate, the mental states implicated in this Narrow continuity are states of the *same* brain. On the Wide version, any reliable cause constitutes the 'right kind of cause'. And for the Widest version, any cause whatsoever is the 'right kind of cause'.

Parfit believes that there is no good reason to choose either the Narrow or Wide view over the Widest largely due to the indeterminacy of persistence.[6] He thinks that since persistence is reductionistic, we do not actually *need* to state whether I persist through something like Teletransportation. Rather, giving a description of the facts ought to be enough. So, we can say only that one subject was destroyed, another created, and that the two are psychologically continuous (with or without the right kind of cause). Asking whether the Replica is the same as

[6] Parfit recants from this in his 2012, where he adopts something much closer to the Narrow view.

the Original or merely someone else who is exactly like the Original is to offer two different descriptions of the same set of facts. It is not to describe two different cases. If this is true—and Parfit definitely thinks it is—the Narrow version of the Psychological Criterion is too restrictive and perhaps draws only arbitrary determinations on persistence. Hence, Parfit's argument ends up being that subjects persist in virtue of non-branching continuity of psychological features (defined as overlapping chains of direct psychological connections) with any cause whatsoever.

Another version of the psychological approach from Shoemaker shares some of the bones of Parfit's account but makes very different metaphysical commitments. As observed by Dean Zimmerman (2009), Shoemaker's treatment of persistence is complex, systematic, and has been the product of the decades-long development of an interconnected metaphysic. The result is a functionalist account of persistence that is deeply tied to a coincidence or constitutionalist ontology of persons.

Shoemaker (2011) argues that persons and human bodies or human animals must be numerically distinct but coincident things.[7] This should be thought of in the way that a statue and a lump of clay, though different things, share the same material—that is, they are coincident—and so some of the statue's qualities and the lump's qualities are instantiated in the same physical realizer.[8] Properties shared in this way are 'thin' properties. Conversely, 'thick' properties are those properties specific to particular kinds of thing, and so will not be shared by coincident entities. Since persons and human bodies or human animals are distinct but coincident, they will each have some thin properties and some thick properties. Shared thin properties—like size, shape, and mass—determine what sort of thick properties can be

[7] This 2011 paper is just a recent version of an argument that Shoemaker has presented and refined through many earlier works. See also his 1984, 1999, 2003, and 2008.

[8] This is commonly said another way than 'coincidence': some say that the statue is *constituted by* the lump of clay. But this comes to the same thing as 'coincident with', so I will stick with that phraseology going forward.

instantiated but without determining the particular kind of thing that has the instantiated properties.

To put it another way, only thick properties determine the persistence conditions of things. On Shoemaker's account, the persistence of anything (not just persons) "is constituted by...a set of property instances the simultaneous members of which are united by relations of synchronic unity and the non-simultaneous members of which are united by relations of diachronic unity" (Shoemaker 2011, 357). That is, something persists in virtue of the way a set of properties hangs together either at a time (synchronically) or over time (diachronically). The 'relations' by which properties achieve synchronic or diachronic unity are causal. In its current form, Shoemaker puts the argument in terms of the 'causal profiles' of property instances. The sum of the effects of a property and its causal history—what sorts of things caused or could cause its instantiation—are a property's causal profile. The 'thickness' and 'thinness' of property instances refers to the richness of the property's causal profile. Thick properties have causal profiles that contain synchronic and diachronic unity relations; thin properties do not. This is how thick properties determine something's persistence conditions.

Although he does not always use this terminology, Shoemaker's is a *functionalist* account of persistence. Property instances or states are defined by how they relate to other property instances or states, and this relation is usually causal. Such property instances are synchronically unified when it is part of a property's causal profile to bring about certain characteristic effects when coinstantiated with other properties. To borrow an example from Shoemaker's 1984, the belief that it is raining and the desire to remain dry only regularly bring about the effect of bringing an umbrella along when the belief and the desire are coinstantiated in the same subject. In the same way, property instances are diachronically unified when it is part of a property’s causal profile that it bring about certain ‘successor states’—straightforwardly, later states of one and the same thing.

In earlier versions of the account, Shoemaker (1984, 1997) suggests that functionalism implies that psychological continuity is both necessary and sufficient for the persistence of persons. While Shoemaker never explicitly spells out his reasons for making this claim, here is what I think he means, as applied to the more recent characterization of the argument that we have been considering in this section. That psychological continuity is sufficient follows from functionalism because it is part of a mental property’s functional characterization (or causal profile) that its characteristic effects occur in the same subject over time. Since persons are subjects of mental properties, an account of the functional relations between mental properties over time just is an account of a persisting person. To show that psychological continuity is also a necessary consequence of functionalism, we start with the same premises as before: it is part of a mental property’s functional characterization (or causal profile) that its characteristic effects occur in the same subject over time, and persons are the subjects of mental properties. We then add that because functionalism allows for a mental property to be realized in just about anything that lets the property fulfill its functional role, only the relation between successive mental properties really matters for persistence and not how the successive property instances are realized.[9] So, functionalism also implies that psychological continuity is necessary for persistence.

[9] Some support for this explanation can be found in Shoemaker (1984), where Shoemaker challenges the idea that the physical realizers of psychological continuity need to be biological—that is, that psychological states must be realized in some sort of organic body. He imagines a 'Brain-State Transfer Device' (BST) that scans and stores the total brain states of a subject, completely destroying the scanned brain in the process. These total states can then be 'written' onto a new brain. Shoemaker supposes that the 'person' could persist through this process by way of psychological continuity, which would demonstrate that no *particular* sort of physical body is necessary for persistence. Shoemaker has since abandoned the idea that the BST is 'person-preserving', jokingly attributing the claim to one of his 'past selves' (2004).

Given Shoemaker’s functional formulation of psychological continuity, some—like Gary Fuller (1992) and Lawrence Davis (1998 and 2001)—have taken to referring to Shoemaker’s account as one based on *functional continuity* (although they dispute, to varying degrees, Shoemaker’s claims that functional continuity is necessary and sufficient for persistence). While Shoemaker has never personally adopted this usage, it does seem to *prima facie* capture the unique character of Shoemaker’s account. Put concisely, we might summarize the argument in the following way: persistence is constituted by the functional roles of successive mental properties or states. This means that, per the goal of this section, we can also say that Shoemaker cashes out psychological continuity as functional continuity of mental states.

David Lewis (1983) proposes a third major approach to psychological continuity that exploits the definition of person-stages as time-slices of a single continuant person. His theory is a sort of four dimensionalism (see, for example, Sider 1997), a view which posits that things do not exist through time as simple whole objects. Rather, things are extended through time in a way analogous to the way that things are extended through space. The region of space that my couch occupies can be divided into smaller sub-regions that contain only particular parts of the couch—say, the right armrest. In just the same way, we can also say that my couch occupies a certain region of time throughout its career as a couch. (Presumably, it’s ‘career’ would be the region of time beginning when the couch was fully put together and ending when the couch is destroyed, however that comes about.) During its career, there is a sub-region of time where my couch acquires an unsightly stain on the right cushion, and there is another sub-region wherein I in vain attempt to clean this stain with a chemical that actually just leaves behind a larger stain. So, we have specified two *temporal parts* of my couch that relate to the whole career of the couch analogously to the way that a spatial part—like an armrest—relates to the whole spatial object. Following from this, four dimensionalism invites an account of persistence wherein something persists as the sum of its temporal parts. This is called perdurance.

Lewis sets out to provide something like a four dimensionalist account of psychological continuity. For Lewis,

> something is a continuant person if and only if it is a maximal R-interrelated aggregate of person-stages. That is: if and only if it is an aggregate of person-stages, each of which is R-related to all the rest (and to itself), and it is a proper part of no other such aggregate (1983, 61).

What Lewis means is that there is a relation—call it the I-relation—between person-stages that makes them the stages of one continuant person. He argues that this I-relation obtains when person-stages are R-related to one another, where being R-related is cashed out in terms of Parfit's mental continuity and connectedness. In other words, Lewis says that this version of psychological continuity is what unites person-stages into a continuant person.

What sets Lewis's account apart is his allowance that continuant persons can overlap. Suppose there is amoeba-man-like fission of a subject. Ordinarily, we would say here that one subject split into two. This would mean that fission is not a person-preserving process. The identity of persistence is a one-one relation; two distinct persons cannot both be identical to the same pre-fission person. However, Lewis thinks that we should instead say that there were two overlapping continuant persons all along who just shared pre-fission person-stages. On this view, fission is not a case of one subject splitting in two but of two overlapping subjects failing to overlap anymore. At the same time, this solution raises other problems. It seems to frustrate our ordinary attempts to count how many persons there are at points where persons share stages. If there are two persons, how does it make sense to say what seem to be perfectly natural things like "on the day before the fission only one person entered the duplication center; that his mother did not bear twins; that until he fissions he should only have one vote; and so on" (Lewis 1983, 65)?

To counter this, Lewis suggests that we do not count by identity *simpliciter* but by a weaker relation: identity-at-a-time, or tensed identity. Since the post-fission subjects were identical yesterday, it makes perfect sense to say that *one person entered the duplication center* (qua identity-yesterday), even if two persons entered in terms of identity *simpliciter*. Thus, when counting according to tensed identity, we give the natural answer to 'how many persons entered the duplication center yesterday?' In this way, Lewis's definition of psychological continuity amends Parfit's 'continuity and connectedness' approach with the notion of overlapping temporal person-stages in order to overcome the fact that fission is not person-preserving. Whether Lewis's amendment is an improvement over accounts like Parfit's or Shoemaker's depends on how much tolerance the reader has for four dimensionalism more generally, as well as the often strict mereological and compositional commitments the view implies.

I close this section with a novel and more recent take on the psychological approach from Barry Dainton and Tim Bayne (2005). Dainton and Bayne do not embrace ordinary neo-Lockean psychological features as constitutive of persistence. Instead, they construe psychological continuity as continuity of phenomenal experience. As Dainton and Bayne see it, describing persistence in phenomenal terms turns the relation between earlier and later mental states from a causal to an experiential one. They argue that phenomenal experience is conceptually distinct from the ordinary set of psychological features found in the psychological approach, and so a subject's stream of consciousness could come apart from her beliefs, desires, memories, etc. Hence, the phenomenal stream of conscious experiences is the focus of Dainton and Bayne's view.

From here, the argument takes on a more recognizable neo-Lockean structure. One can have simultaneous conscious experiences—say, seeing a bird fly by while also hearing a car horn—that together form synchronically connected phenomenal experiences. Dainton and Bayne argue that our streams of consciousness are ordinarily more seamless than this, however. Rather, much of the time, streams of consciousness are made up of many successive synchronically connected experiences that cannot be picked apart discretely. One's experience of synchronic consciousness as persisting through time gives us access to another form of phenomenal connectedness: diachronic phenomenal connectedness. While synchronic and diachronic phenomenal experiences are themselves too brief to sustain a model of persistence, overlapping chains of such phenomenal connectedness can achieve streams of phenomenal continuity. Thus, a subject persists in virtue of having a continuous stream of consciousness. Dainton and Bayne call this the Inseparability Thesis. In their words: "self and phenomenal continuity cannot come apart: all the experiences in a single (non-branching) stream of consciousness are co-personal" (2005, 557).

1.2 Other Approaches

Although the psychological approach is the most popular persistence theory among Western philosophers, there are a number of competitors to this view. I now turn to briefly consider several of these alternatives: the somatic approach, the substance approach, and nihilism about persistence. I also remind the reader of my overall goal in digging deep into the persistence literature in this way—to better demonstrate that there is no good framework for a necessary moral persistence condition in the existing literature.

1.2.1 The Somatic Approach

We start with the somatic approach. What I am calling the somatic approach (after Olson 2009) is actually an umbrella term for a few different positions that share belief in a physical persistence condition. Though Bernard Williams (1970) argues for a physical persistence condition more generally, the somatic approach typically falls within one of two specific views whose relation to one another is not immediately evident. The first of these—which enjoys less support in the literature than the other—is that there is a bodily persistence condition.

Judith Jarvis Thomson (1997) says that people just are their bodies and calls this the ‘ordinary view’ of persistence. That is, she seems to think that most common-sense approach to persistence and the most natural answer to the question of what makes someone persist over time, at least for those who have not been spoiled by far-fetched thought experiments meant to prompt psychological intuitions. As is the case for most versions of the somatic approach, Thomson’s argument largely turns on casting doubt on the reasons we might have to accept the psychological approach and the purportedly attractive intuitions that support its adoption. Namely, she thinks that the psychological approach gives us implausible results. For one, it forces us to say that feeding someone a drug that completely changed her psychological features would bring about a new person (and the death of the old one).

A more popular version of the somatic approach—and the view in the recent persistence debate to most seriously contest the psychological approach—is animalism. Rather than some sort of bodily continuity, the animalist grounds persistence in *biological continuity.* In other words, what it takes for a subject to persist over time is a functional organization of material components that include such life processes as metabolism, respiration, and circulation. Paul F.

Snowdon (1989), Michael Ayers (1990), and Olson (1997) give standard accounts of animalism.[10]

Snowdon and Olson agree that the animalist view harmonizes nicely with most of our natural, everyday reactions to persistence questions. That is, when we recognize someone as the same individual we saw the day before, it is not on grounds of matching beliefs, desires, and other psychological features, but because the individual seems to be one and the same animal.[11] In this way, what we ordinarily take to be the persistence conditions of each other happens to be the same condition by which we approach the persistence of dogs, cats, and horses. Said differently, we regard ourselves as having the same persistence conditions as other *animals*. If we carry this notion with us as we reevaluate things like brain transplant/switch scenarios, Olson thinks it ought to be clear that what happens is not a case of body-swapping but of one animal donating an organ to another animal. This is especially true if we consider that the transplant scenario could be framed in such a way as to permit the donor animal to continue to survive in a persistent vegetative state, a condition where higher cognitive functions are impossible but ‘vegetative’ functions like respiration, circulation, and even digestion continue unimpaired.[12]

Due to widespread philosophical acceptance of the psychological approach, animalists generally spend much of their time arguing why we ought to resist intuitions favoring psychological persistence conditions that cases like brain transplants/switches are meant to

[10] There are also more divergent approaches to animalism. For instance, van Inwagen (1990) argues for a biological persistence condition on the basis of his metaphysical commitment to a sparse ontology in which the only material things that actually exist are either 'physical simples'—things that have no proper parts—or organisms. All other sorts of material thing are merely 'virtual objects' that, while linguistically convenient, do not actually exist.

[11] This example seemingly includes questions about two sorts of identity claim: persistence and evidence. What matters for the example, though is not what *counts* as evidence of identity, but what sort of persistence conditions the evidence supports. Of course, I could misjudge the evidence; suppose I see your twin and think it is you. But I do not think this undermines the animalist's point: in commonsense cases, we evaluate persistence based on appearance because appearance is generally a good indicator that something is the same human animal it was before.

[12] This could be induced in the patient by perhaps removing only the cerebrum but preserving the brain stem and midbrain organs.

evoke. Perhaps the hardest animalist critique for the advocate of the psychological approach to overcome is what Olson (1997) calls the 'thinking animal argument'; alternatively, this is known as the 'too many minds objection' to Shoemaker (1999) and the 'too many thinkers problem' to Parfit (2012). It also has its roots in Snowdon (1990). Snowdon argues that Locke's definition of person is meant to contrast with animal, but the definition is so loose that 'person' and 'animal' seem perfectly compatible. In fact, certain animals—say, humans—look like the perfect candidates to be persons! Now, if we have psychological persistence conditions, we cannot be animals, as no animal persists based on its psychology. If human animals are conscious and intelligent—if they can be persons—and we are not those human animals, then it seems that every time a human being thinks some thought, there are *two* possible thinkers of that thought. Bizarre though that may seem, it also implies that there is no way to tell which of these thinkers *you* are at any given moment. Olson writes:

> If you believe you are a person, the animal connected with you thinks it is a person as well. It thinks so for the same reasons that you think so; it has the same evidence as you have. (Moreover, the relevant belief-forming processes are the same for the animal and for yourself.) It is mistaken, however, for it is not a person. But if it is so easy to believe that one is a person and be wrong, how do you know *you* aren't mistaken about this (1997, 106)?

In other words, if both you and the human animal share thoughts, then you are psychologically indistinguishable. On Olson's argument, not only does the psychological approach invariably make it that each individual thinker is actually *two* psychologically indistinguishable thinkers, but it also gives us no discernible means to pick out which of those thinkers we are. If this is right, instead of making our persistence conditions clearer, the psychological approach actually makes them more complicated!

1.2.2 The Substance Approach

The common through-line among somatic approach arguments is that our persistence conditions are in some way physiological. In contrast to this, other views insist that persistence is a matter of the continuity of an immaterial substance—say, what Parfit (1984) calls a 'featureless Cartesian ego'. I call this the substance approach. Such a view is of course based on Descartes's claim in his *Meditations* that

> …I know certainly that I exist, and that meanwhile I do not remark that any other thing necessarily pertains to my nature or essence, excepting that I am a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing [or a substance whose whole essence or nature is to think]…it is certain that I [that is to say, my soul by which I am what I am], is entirely and absolutely distinct from my body, and can exist without it (1911, 190).

This is what Swinburne (1984) calls the 'classical dualist' position. Descartes argues from the fact that the only thing he cannot epistemically dismiss by way of his method of doubt is the knowledge that he is a thinking thing that exists to what he most fundamentally must be is a substance with the capacity for thought (a soul). Further, Descartes thinks it is logically possible to imagine this thinking thing existing apart from any sort of material extension. This has the consequence that, though each of us is attached to some sort of body, there is no sense in which this body has anything to do with what we really are (and hence how we might persist over time)—even if a human being, or "complete man" (Descartes 1934, 207), has both a soul and a body. So, on the classical dualist's position, persistence is just continuity of the same immaterial substance (or soul).

The dualist approach to persistence can take two general forms. The first is the 'simple' view. This mirrors the Cartesian account above, asserting that what matters in persistence is the continuity of the same immaterial substance only. Chisholm (1991) gives such an account, as do Swinburne (1984) and John Foster (1991). The second version of the dualist approach to persistence is the ‘complex’ view. On this view, continuity of a union between immaterial substance and some particular body is needed to satisfy persistence. St. Thomas Aquinas (1956) is an example of someone who makes the case for a complex dualist view of persistence (this is probably more accurately called a form of hylomorphism). Though neither the simple nor the complex dualist views are especially popular in contemporary philosophical discussion of persistence, the simple view seems to be the more commonly adopted.

1.2.3 Nihilism About Persistence

A final alternative to the prevailing psychological and somatic approaches is what might be called nihilism about persistence. For the nihilist, there are no persistence conditions for subjects just because *there are no subjects*. We can see this sort of nihilism in Hume’s brief remarks on the impossibility of locating the self in Book I, Section VI of his *Treatise on Human Nature*:

> 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 any thing but the perception. When my perceptions are remov’d for any time, as by sound sleep; so long am I insensible of myself, and may truly be said not to exist (1896, 142).

His point is that we seem capable only of an awareness of the experiences themselves and never a subject that has the experiences. Since we do not have access to a ‘self’ over and above some particular perceptions that we experience at a given time, Hume argues that ‘self’ is nothing but a “bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement” (134). This is another way of saying that there is no self at all, as there is nothing that unifies these varied and inconstant bundles into what could be called a persisting thing.[13]

The Buddhist tradition accepts a similar nihilism about persistence. The *Milina Panha* (quoted in Collins 1982 and Parfit 1984), records a dialogue between a king and a monk. When asked his name, the monk replies that it is Nagasena, but cautions the king that "it is just an appellation, a form of speech, a description, a conventional usage. 'Nagasena' is only a name, for no person is found here" (Collins 1982, 182-3). Likewise, Stcherbatsky (1919b) presents a conversation between Vasubandhu and Vatsiputriya about the existence of persons as cognizing agents. When Vasubandhu asks what the agent is, Vatsiputriya attempts to reply that it is a human being—for instance, a man named Devadatta. Vasubandhu counters:

> this does not represent any unity whatsoever. It is a name given to such elements (of which a man is composed)...It is an unbroken continuity of momentary forces (flashing into existence), which simple people believe to be a unity, and to which they give the name Devadatta (937).

Indeed, the Buddhist term for the individual is *santana*, which means stream (Stcherbatsky 1919a, 853).

These examples suggest a picture of persistence very much like the Humean view.[14] The Buddhist grants that there are perceptual experiences but denies that these link-up to form a persisting subject in anything more than name. Of course, such a name would be empty to the Buddhist—that is, it would have no real referent, but is instead a matter of convention and convenience in language. There can be no persisting subject because the perceptions of which such a subject would be comprised do not themselves persist. Hence, there would be nothing in virtue of which the subject could be said to persist. This is what is meant by calling an

[13] But see Parfit (1984) and Olson (2007) for examples of interpretations of Hume's 'bundle theory' that do not regard it as so nihilistic.

[14] See Giles 1993 for a good summary of where Humean and Buddhist views of persistence might overlap.

individual a stream; the would-be contents of an ostensibly persisting subject, like the waters of a stream, are never the same.

Overt nihilism about persistence is not exactly a popular position. At the very least, it is quite controversial to say that no subjects exist, and few philosophers do it. Nonetheless, there are contemporary accounts of persistence that take the nihilistic route. For instance, Peter Unger (1979a, 1979b, 1979c) has developed a broad nihilistic metaphysic around mereological simples—things with no proper parts—on which neither he, nor anyone else, nor any other ordinary object exists.

1.3 Why a Moral Condition Does Not Fit Into the Literature (And How It Could)

In this section, I argue that the foregoing reveals two things about the persistence literature: an observation (for now) and a problem. (To make it easier to talk about these, I will refer to the former as The Observation and the latter as The Problem.) First, The Observation: it is clear to me that a necessary moral persistence condition is not supported by the existing literature. More specifically, I make the claim that any account that includes moral character as even partly constitutive of persistence must be a psychological account. But, as I will go on to show, even current psychological theories include moral character as at best an accidental persistence condition. So, no present account of persistence can support a necessary moral persistence condition. If we end up with good reason to suppose that there *is* a necessary moral persistence condition, then the inability of the structure of the persistence literature to cope with such a condition *would* be a serious problem. Yet, as I will argue in Chapter 2 that we *do* have good reason to suppose that there is a necessary moral persistence condition, I think that the literature's incompatibility with such a condition will be a problem that the persistence theory I argue for in Chapter 4 might solve. Part of my dissertation, then, will be in showing that The Observation is actually a problem.

Second, The Problem: following Marya Schechtman, I argue that the existing persistence literature does not properly account for the moral considerations of its authors. Many philosophers writing on persistence are motivated by moral concerns and see moral consequences from their persistence theories. However, since their persistence theories do not take a moral character to be necessary for persistence, the actual metaphysics has little to no connection to the moral at all. Hence, despite their moral concerns, these philosophers allow no real relationship between the moral and metaphysical facts of persistence. This is a real gap in the literature, and I think that the necessary moral persistence condition that my project advances could help to close it.

To begin, though, I would like to speculate as to why the current literature is in many ways silent on the role of moral character in persistence. Frankly, the persistence literature just does not have much to say about moral character in the context of persistence. What attempts *are* made to include moral character are not optimistic as to its role in persistence. Parfit, for one, mentions moral character traits as one of myriad psychological features that constitute persistence. But for Parfit and those like him, persistence is a matter of the subject's entire psychological profile; character traits are just one psychological feature among many and have no special importance for persistence. Swinburne, though a substance-persistence theorist, nonetheless regards psychological features like character traits as good evidence of persistence. Yet he is quite dismissive of character traits even in this limited, evidential capacity: "Character continuity is a minor kind of evidence, hardly of great importance on its own" (Swinburne 1984, 24). So, even when moral character is acknowledged as contributing in some way to persistence—whether constitutive, evidential, or otherwise—the contribution is taken to be minimal.

I believe that part of the reason for this is that these philosophers see moral character traits as too general to fix individual subjects. That is, they think that character traits cannot individuate subjects very well. Suppose you want to pick out a present subject that corresponds to some past subject who had the following traits: generally courteous, mildly selfish, and a lover of animals. This character description is far too vague to be helpful in differentiating one present subject from another, as these are fairly common traits that probably fit *lots* of subjects. Now, it might be said that this description just is too coarsely-grained. If we were to specify a complete character profile for all past and present subjects, then we could find *the* present subject who is one and the same as a past subject (if there is one). Ignoring that constructing such a fine-grained character profile would be immensely complicated,[15] it still seems plausible—if rather unlikely—that two subjects could share identical character profiles. Hence, taking moral character traits alone to be the conditions for persistence makes the job of identifying *who* persists that much harder. And this is certainly not what persistence conditions should do.

1.3.1 The Observation: No Existing Account Can Accommodate a Moral Persistence Condition

Whatever their reasons, persistence theorists largely leave moral character out of their theories. More than this, it seems to me that the prevailing theories in their current states could not give moral character a necessarily constitutive role in persistence. This is because what theorists identify as the persistence conditions in each case have nothing (or almost nothing) to

[15] The rabbit hole of possible character traits can be as deep as you like. Consider peculiarly specific traits along the lines of *loves animals mostly but is exceptionally cruel to squirrels* or *kind only to convenience store clerks*. It seems next to impossible to classify such traits under broader headings just because they are built around exceptions.

do with character traits. I will consider the somatic approach, the substance approach, and the psychological approach each in turn.[16]

Looking to the somatic approach first, the condition for persistence is continuity of a certain functional organization, where this organization is defined either as the body or the whole organism, whether or not this organism can be pared down to something smaller—say, a brain. Persistence is then a physiological matter. It is about having the right sort of physiology structured in such a way as to permit certain self-sustaining life functions. Where this functional organization persists, so too does the subject. But no body or organism maintains its physiological structure by virtue of its character traits. This is true even if, following van Inwagen (1990), we say that a naked brain can count as the organism in question. Further, it seems possible for bodies and organisms to survive absent *any* character traits whatsoever (consider fetuses and humans in a persistent vegetative state). Clearly, then, moral character cannot be in any way constitutive of persistence on the somatic approach.

Now consider the substance approach. For the substance theorist, persistence amounts to continuity of some immaterial substance, such as a bare Cartesian ego or a soul. Since moral character traits can be characterized as mental dispositions, and since things like Cartesian egos are mental substances, it might look like the substance theorist is in a good position to account for character traits in a constitutive way. This is not the case, however. To the substance theorist, a mental substance like a Cartesian ego is irreducible to some set of mental features. In other words, a Cartesian ego would persist even if it had no *particular* mental features at all. Therefore, continuity or discontinuity of moral character traits—or any other mental features—cannot constitute the persistence of a mental substance. Since continuity of such a mental

[16] I here omit from discussion nihilism about persistence. This view quite obviously could not assign a constitutive role to moral character, as there is no persisting subject to the nihilist.

substance is the persistence condition of the substance approach, it follows that moral character cannot be constitutive of persistence on the substance approach.

Lastly, we have already seen how the psychological approach treats moral character. Persistence comes down to continuity of psychological states. Though exactly what this means varies considerably from theory to theory, the continuity the psychological theorist is interested in is usually continuity of a subject's overall psychological profile. This is certainly true of Parfit, Shoemaker, and Lewis. Moral character traits would definitely count as part of this profile and thus play a constitutive role in persistence, but not an especially strong one. After all, there are many different psychological features at play here, and character traits are no more important to persistence than memories, desires, beliefs, and personality traits—among others. Indeed, most psychological accounts are neutral on which psychological features are more important to persistence.

Given that *overall* psychology is what really matters, it seems entirely plausible for a subject to persist with complete discontinuity of character traits so long as a sufficient number of other psychological features remain the same. In many cases, this is regarded as an advantage of psychological accounts. No single psychological feature is privileged above any other, so subjects can persist through a wide range of changes. So, the best that can be said for the psychological approach here is that it *can* support moral character as an accidental persistence condition. But psychological theories can also get along just fine without continuity of character at all.

Dainton and Bayne are perhaps an exception here, but not one that really challenges my point. Their approach to psychological continuity—phenomenological continuity by way of overlapping streams of consciousness—is decidedly unique. At the same time, it is apparent that moral character traits are no better off in such an account. Actually, character fares *worse* on Dainton and Bayne's account. If persistence is a matter of phenomenological continuity, character cannot be even partly constitutive of persistence. This is because a subject could persist through complete discontinuity of moral character under the condition that the subject has continuous phenomenological awareness of this shift in character. So, although they offer a psychological account where continuity is not about the whole psychological profile, Dainton and Bayne's theory still comes to the same conclusion about moral character. That is, since psychological continuity can obtain in all of these cases through complete discontinuity of moral character, moral character is at best an accidental persistence condition but never a necessary one.

It might be thought that we could resist this claim in the following way: perhaps we could say that moral traits are too tightly bound-up with other mental features, and so there is no real way for both psychological continuity and moral *discontinuity* to simultaneously obtain. For instance, if I am a generous person, surely this implies that I have certain beliefs and desires, like the belief that it is good to give of my resources to others and the desire to do just that. If there is continuity among the relevant beliefs, desires, and other mental states that are deeply tied to moral character traits, there cannot also be discontinuity in the traits that rely on those states. I consider this objection in Chapter 4.2, where I think it is more relevant to that chapter's argument. There, I think that I sufficiently defuse the objection. For present purposes, though, let me point out that even if moral traits cannot be disentangled from the majority of our other mental states, my primary claim would still stand. That is, it would still be true that present psychological continuity theories need to be modified in order to accommodate a necessary moral persistence condition, as accounts like Parfit's and Shoemaker's do not view character traits in this interconnected way. Of course, *how* they need to be modified would change. Yet, the prevailing view is explicitly that no one type of mental state has more weight in constituting persistence than any other, and so modification would still be required. And this is true even if the 'modification' is just to specify that moral traits are interdependent with many other mental states and thus necessary to persistence because an entire web of mental states will collapse without them.

In spite of its problems with a moral persistence condition, the psychological approach is the only approach to persistence that allows for moral character to be even partly constitutive of persistence. If continuity of moral character turned out to be a necessary persistence condition, it may be possible to modify some versions of the psychological approach to reflect this. (In fact, I argue in Chapter 4 that we can do exactly this with a version of Parfit's view.) The same, of course, could not be said for the other approaches. This leaves the psychological approach uniquely able to accommodate the sort of moral persistence condition I will argue for later in the dissertation, though not without alterations.

This, then, is The Observation about the literature that my survey reveals: since no current persistence theory can accommodate the claim that moral character is a necessary persistence condition, no current persistence theory could give an accurate description of the metaphysical facts *if* moral character turned out to be such a persistence condition. And because only the psychological approach is compatible with moral character as even partially constitutive of persistence, *if* moral character were a necessary persistence condition, it looks like only a modified version of the psychological approach could adequately capture this idea.

1.3.2 The Problem: The Strained Relationship Between Persistence and Morality

The Observation is a problem only conditionally—it is only a problem if moral character actually is necessarily constitutive of persistence. There is a second problem, though, that does not depend on whether or not there is a necessary moral persistence condition. Indeed, I argue that it is a genuine gap in the current literature. Schechtman (1996) observes that many of the intuitions primed by persistence thought experiments are about practical, moral features. There are four features in particular that Schechtman finds consistently in the persistence literature: survival, moral responsibility, self-interested concern, and compensation. These features are what prompt many philosophers to tackle questions of persistence in the first place, and they are just as often considered as immediate implications of taking such-and-such to be constitutive of persistence. Locke is worried about persistence so that we can determine who is accountable for specific actions (and thus who ought to be compensated—or punished—accordingly). Parfit starts *Reaons and Persons* with a close look at theories about rational self-interest and devotes considerable space in the back half of the book to desert, responsibility, and survival. The relationship between survival and identity is also treated thoroughly in Shoemaker (1971) and Lewis (1983). And many of the thought experiments that underpin the persistence discussion frame the intuitions they generate in terms of special concern we have with subjects who we identify as ourselves. Chisholm (1969) and Williams (1970) certainly do this with their examples, and so do responses from Parfit (1984) and others. So, what Schechtman calls the 'four features' are tied up in the persistence debate at nearly every turn.

The trouble with this is that ordinary accounts of persistence end up abandoning these features as bearing any sort of relation to persistence. Though there are many others, the best example of this is probably Parfit (1984), who claims that identity is not really even what matters for our practical concerns.[17] The reason the loss of connection between the four features and persistence is a problem is that it cannot make sense of the way that we think identity matters to us. Typically, philosophers of persistence take on this problem not by denying that the four features are important, but by accepting that identity may not ultimately have the intuitive significance we thought it did. Schechtman believes that this conclusion is a mistake and that we ought not cast-aside our earlier intuitions about the tight relationship between identity and practical, moral problems. The mistake is that most philosophers do not recognize other identity relations than persistence. To Schechtman, the problem is not that identity cannot answer these practical intuitions, but that *persistence* cannot. Schechtman thus abandons persistence altogether for another aspect of identity that she calls ‘characterization’. She describes characterization in the following way:

> Most simply put, this question asks which actions, experiences, beliefs, values, desires, character traits, and so on...are to be attributed to a given person...characterization theorists ask what it means to say that a particular characteristic is that of a given person (1996, 73).

It is not, then, a question of how one and the same numerical subject persists but how we can distinguish to whom certain attributes belong. Schechtman sees the distinction between persistence and characterization as the difference between, respectively, regarding persons as objects or as subjects. It is only from the point of view of persons as subjects that we can begin to understand the relation between identity and the four features.

This is The Problem. Persistence thinkers are unable to reconcile their moral concerns with the metaphysics of persistence. Whether or not there turns out to be a necessary moral persistence condition, I think the gap of The Problem remains. This is because the typical response is to either (a) dismiss the importance of persistence to morality or (what comes to

[17] See also Parfit's 1995, tellingly titled "The Unimportance of Identity".

much the same thing) (b) to keep moral concerns and the metaphysics of persistence apart because they just do not play nicely together. To me, neither response is especially compelling. Nor does Schechtman's response work for me, as she is basically committing herself to another version of (a) by ditching persistence-identity for characterization-identity. I think that the metaphysics of persistence still could be importantly related to the moral. Even without a necessary moral persistence condition, it is possible that this relation could obtain in some other way. And, even without a necessary moral persistence condition, the typical responses to The Problem would seem to side-step the gap rather than bridge it.

1.4 Addressing Problems and Closing Gaps

In this chapter, I conducted a comprehensive survey of the persistence literature in order to reveal two important things regarding how persistence is talked about. What I called The Observation is the idea that current approaches to persistence are not equipped to deal with the possibility of moral character as a necessary persistence condition (though the psychological approach might be modified to be compatible with this possibility). Moreover, I argued that only the psychological approach is compatible with even the weaker claim that moral character is partially constitutive of persistence. Then there is The Problem, a gap in the literature that arises because persistence theorists have trouble squaring their moral concerns with the metaphysical facts of persistence. I claimed that the typical responses—keeping the two separate or outright disregarding the importance of persistence—are not satisfactory.

The reason that I think my project is valuable to the persistence debate is that it can on the one hand show that The Observation is actually a problem and on the other hand work to solve The Problem. If The Observation is really a problem in disguise, an account of persistence with a necessary moral persistence condition would obviously resolve it—it is uniquely fitted to it. But discussion of The Observation has already shown us how such an account would begin to take shape. That is, a necessary moral persistence condition would have to be part of a modified psychological theory of persistence, as this is the only approach amenable to a moral persistence condition at all. I develop this bare sketch of a theory into something more robust in Chapter 4, using a version of Parfit's psychological view that I introduced in this chapter.

Regarding The Problem, while a necessary moral persistence condition may not directly engage with the 'four features',[18] such a condition could meet the spirit of Schechtman's objection, if not the letter—i.e. that the moral concerns of persistence philosophers seem to have no place in their actual persistence arguments. In other words, the possibility of moral character as a necessary persistence condition would tie the moral to the metaphysical facts of persistence. Schechtman solves this problem by abandoning persistence altogether for another relation, 'characterization'. Conversely, a moral persistence condition could properly link the moral and the metaphysical while still putting forth a proper persistence theory. So, an account of persistence based around a necessary moral persistence condition might turn The Observation into a problem—and simultaneously resolve it. At the same time, such an account could fill the gap in The Problem—or at least reduce the size of the gap. It seems to me that an account of persistence based around a necessary moral persistence condition would thus have much to contribute to the persistence debate.

I also suggested in this chapter that some part of the reason that moral character is under-represented in the persistence literature is because of an individuation problem. Namely, since it is plausible for two or more subjects to have an identical character profile, a subject's moral

[18] I leave it open that a moral persistence condition *might* or *might not* be able to do so—but the purpose of this dissertation is not to argue against Schechtman, so I set aside this problem for another time.

character alone does not seem to be sufficient to distinguish one subject from another over time. This points to an important difficulty that my account will have to contend with in later chapters.

CHAPTER 2

MORAL PERSISTENCE INTUITIONS AND WHY THEY MATTER

In this chapter, I argue that folk intuitions about persistence make the notion of a necessary moral persistence condition plausible. However, the move from discussing intuitions to drawing conclusions about the actual metaphysics of persistence is complicated by antagonism within the field toward the methodology of arguing from persistence intuitions, including opposition to the use of the elaborate thought experiments designed to elicit those intuitions. For instance, Bernard Williams argues that exactly which intuitions are elicited depends largely on the way a thought experiment is framed. This means that thought experiments can often produce contradictory persistence intuitions, which makes it very difficult to determine which—if any—of our intuitions actually correspond to any metaphysical facts. In a similar way, Kevin Tobia observes that psychological persistence intuitions in particular may be susceptible to bias. He thinks that we are more apt to consider a subject persisting through even severe psychological changes if we assess those changes as 'for the better' of the subject. This, too, would rightly bring us to question the credibility and usefulness of these intuitions in making persistence claims.

If either Williams or Tobia is right, my attempt to motivate an argument for a moral persistence condition by way of our persistence intuitions would face substantial challenges—assuming that their criticisms would not block the move entirely. I argued in the previous chapter that only a psychological account of persistence could accommodate a necessary moral persistence condition. Since Tobia's objection targets psychological persistence intuitions more specifically, it sounds especially bad for my project. I do not think that either criticism succeeds in forcing us to dismiss our persistence intuitions, however. In spite of opposition like Williams and Tobia, I believe that we are justified in using our persistence intuitions in a limited, non-evidential capacity, and I will present arguments for how we can resist the opposition's worries. More pointedly, I argue that our persistence intuitions may be able to guide us to positive reasons for a plausible persistence account if those intuitions are (a) stable and (b) free from bias. Our intuitions are often reliable in general. So, as long as we do not take these intuitions as definitive proof of a given persistence account, then I think we may use intuitions in our theorizing.

This chapter has three main sections. In the first section, I consider empirical data that show that people intuitively take moral features to be constitutive of persistence—perhaps *the* most important constitutive feature. Then, in the second section, I introduce objections from Williams and Tobia to the well-trodden methodology of making arguments based on intuitions prompted by thought experiments. In the interest of clarity, this is prefaced by a short background on the way this methodology has been traditionally used in the persistence literature. Williams claims that our persistence intuitions are inconsistent and depend on how a thought experiment is framed. For his part, Tobia suggests that persistence intuitions that favor a psychological approach to persistence are biased. Taken together, these objections seem to demonstrate that our persistence intuitions make a poor guide to a workable persistence theory.

Lastly, I respond to Williams and Tobia in a third section. In countering their objections, I make the case that we can justifiably use the empirical results on folk persistence intuitions I present in the first section to support a necessary moral persistence condition. Against Williams, I argue from empirical data that the apparent conflict between persistence intuitions disappears when thought experiments are given a neutral frame of presentation—in fact, persistence intuitions resulting from a neutral frame tend to more consistently favor the psychological approach. In reply to Tobia, I argue that his conclusions go further than his data warrant. The bias he postulates, even if true, would not be devastating to my use of persistence intuitions because this bias at most advises care when drawing conclusions about persistence intuitions—it does not prove those intuitions cannot guide us in coming up with persistence theories that accurately describe the state of things. Further, I propose empirical data that suggest that Tobia's bias may be restricted to certain cases and may not actually have a meaningful impact on our persistence intuitions after all.

2.1 Empirical Evidence of Moral Persistence Intuitions

Strohminger and Nichols (2014) offer data that suggest subjects intuitively take moral features to be of high importance to persistence. In a series of studies, participants were given scenarios in which various mental features of a given subject were affected in some way, then asked about which of these features matter to the subject's persistence. For each case, participant responses heavily favored moral features as most persistence-preserving. (Examples included character traits like conscientiousness, honesty, tendency towards racism, and 'being moved by the suffering of others'.) Because each study presents unique circumstances that may be thought to elicit different responses from participants, I consider each study individually below.

The first study introduced participants to a hypothetical accident victim named Jim who underwent a surgery wherein pieces of his brain tissue were replaced. Each participant received one of a set of prompts to describe Jim's post-surgery condition. Either Jim exhibited no differences in his behavior whatsoever or suffered from one of the following: agnosia, desire apathy, complete amnesia, or loss of moral faculty. When asked to rate the extent to which the subject ‘Jim’ persisted in each case, the loss of moral faculty was rated as most impactful to Jim’s persistence. Moreover, when asked to explain their ratings, participants who rated loss of moral faculty to be identity-affecting considered it to be identity-affecting in and of itself and not because this loss would lead to the loss of other mental features. This was not true of, say, memory loss, which was reported by some participants to be identity-affecting only contingently. That is, at least some participants rated memory loss as highly identity-affecting because it would lead to loss of other features—like character and personality.

For the second study, participants were told of a drug that could selectively modify a single cognitive trait. Using a sliding scale, participants were asked to rate the degree to which using this drug to modify each of 62 different mental features would result in a subject ‘being the same person’ or ‘completely different’. As with the first study, participants rated change in moral features to be most identity-affecting. Since this study included a more specific set of mental features than the first, Strohminger and Nichols were also able to determine whether features like memory were regarded uniformly or if varying ‘types’ of memories elicited different responses from participants. Perhaps tellingly, episodic memories relating to social or intimate personal episodes were rated as substantially more identity-affecting than procedural memories like knowing how to ride a bike. Strohminger and Nichols argue that this is further evidence that memory is intuitively important because of its perceived relationship to social and moral features.

The third study extended questions about persistence to include folk-psychological notions of the soul. Participants were asked to both suppose that immaterial souls existed and could migrate to different bodies and to suppose that the soul of a man named Jim had moved to another body. Then, participants rated the extent to which they believed each of 66 features would transfer to the new body with the soul. Features were grouped among the following categories: moral features, desires and preferences, perceptual features, and somatic features (interestingly, this was the only study to include physical features).[19] Again, moral features were rated as most persistence-preserving—in this case, as most likely to remain with the soul through its new embodiment.

A fourth study centered around identifying possible differences in folk persistence intuitions about moral character traits as compared to non-moral character traits (personality traits). For this study, participants were told to suppose both that reincarnation is possible and that a subject has been reincarnated and that her 'true self'—those features most closely connected to her persistence—was preserved. Given 10 pairs consisting of a single moral trait and a single personality trait, participants were tasked with rating which of the members of the pair would be more likely to be preserved through reincarnation. Results showed that, without exception, moral traits were selected as more likely to be preserved in every pair. So, again, participants rated moral features as most persistence-preserving.

For the fifth and final study, Strohminger and Nichols described a more plausible situation prompt that contrasted sharply with the fanciful and supernatural scenarios from the first four studies: changing mental features due to old age. Using a sliding scale, participants were asked to evaluate the degree to which a change in each of 56 mental features of an old friend that the participant had not seen in 40 years would result in a change in persistence. Changes were split among both what may be considered improvements and decrements in

[19] Although 'somatic' features were rated as the least likely to follow the soul to a new body, it seems to me that it would be erroneous to interpret this response as evidence of intuitive preference for the psychological over the somatic approach. Since the terms of the study were to grant that immaterial souls exist, it is plausible that this predisposes the respondent to consider persistence in at least non-somatic terms. So, the deck is already stacked against the somatic advocate in this case.

mental features. Although the mental features surveyed were divided into broader categories than the fourth study—in addition to moral and personality traits, features here included memories, desires and preferences, perceptual features, and basic cognitive abilities—the results were quite similar. No non-moral feature was rated as important to persistence as any moral feature.

Strohminger and Nichols draw several important conclusions from these studies. Most obviously, they argue that the data support what they call the 'essential moral self hypothesis'. This is the claim that "moral traits are more essential [to identity] than any other mental feature, including those that provide functionality, distinctiveness, or personal narrative" (160). In all cases, folk intuitions about persistence are that moral features are more persistence-preserving than any other mental feature (and sometimes somatic features, too).

Though participants also have intuitions about the role other, non-moral mental features play in persistence, Strohminger and Nichols find that these features are often seen as identity-affecting because they are reducible to social and moral features. For example, certain types of memory—like episodic memories about intimate personal events—are prioritized over others in intuitions about persistence. And this is because these memories are specifically constitutive of, on the one hand, moral features (as moral principles learned and remembered). On the other hand, these episodic memories are often about social relationships. Though Strohminger and Nichols do not explicitly make this connection, it is not difficult to tie such relationships to moral features. One cannot be kind, generous, stingy, or discourteous without relating to others. In other words, it is impossible to be kind without being kind *to someone.* Moral features like a disposition toward kindness, then, certainly can be seen to involve social relationships, too. So, in both cases, those memories that are intuitively considered to be important to persistence are important because they bear on moral features.

Strohminger and Nichols consider the possibility that respondents might tend to favor moral features as persistence-preserving because changes in these features are seen as less common than changes to other mental features, like memories. Forgetting a memory or developing new preferences or desires seems to happen fairly often, but substantial moral change is less so. The idea is that less frequently-occurring types of mental change are seen as somehow 'special', and so when such a change *does* happen, it is taken to have a significant effect on persistence. If this is right, then the fact that responses leaned toward moral features as persistence-preserving would not reveal that moral features are intuitively important to persistence just because they are moral features (rather than some other type of mental feature). Instead, moral features would be persistence-preserving because change in moral features is less common.

However, a model created from the survey responses revealed the opposite: frequently occurring mental changes were regarded as more impactful on persistence than less frequent mental changes. So, not only do the data reveal moral features to be seen as most persistence-preserving, but the intuition that there is a significant relationship between moral features and persistence is an intuition about moral features qua moral features (as supported by responses from the first study). That is, moral features *simpliciter* are considered persistence-preserving; they are not persistence-preserving because of other factors—for example, that changes to moral features are regarded as uncommon and that uncommon changes to mental features are more identity-affecting (again, the Strohminger and Nichols data point to both of these notions as wrong).

A final relevant conclusion that Strohminger and Nichols make is that it is not intuitively seen as especially important that the features that are most persistence-preserving are distinctive to one particular individual. As argued in the last chapter, moral features do not individuate subjects well— in Strohminger and Nichols's words, such features are not 'distinctive'. It is entirely plausible that two simultaneously existing individuals could have identical character profiles, and so distinguishing between them by their unique moral features just would not work. Despite this, the data show that moral features are intuitively considered to be more important to persistence than other mental features that *do* individuate subjects. For instance, the particular experience I had visiting Disney World at age eight is not something shared by anyone else (barring the intervention of extreme science fiction). Thus, memories like this help to individuate me from other subjects. Strohminger and Nichols point out that, if individuation were intuitively especially important to persistence, then we should expect highly specific personal memories like the above to be intuitively seen as most persistence-preserving. Since the data show that folk intuitions about persistence favor moral features over these sort of more individuating ones (like memories), Strohminger and Nichols argue that whether or not the features that matter for persistence individuate the subject is just not intuitively taken to be all that important.

At first glance, the Strohminger and Nichols studies offer generous support for my project. From their data, it would seem that ordinary people do in fact regard moral features as an important—perhaps *the* most important—persistence condition. Moreover, they see moral features as constitutive of persistence in *their own right* and not because such features contingently provide the mental scaffolding for other features, like memories or beliefs. This was not the case for non-moral mental features. Respondents rated both beliefs and memories as important to persistence for the sake of *other* features. For instance, one respondent claimed that the loss of memory would be persistence-severing because it would result in "a different character/personality" (161). Strohminger and Nichols's findings also contain a possible response to my concerns from Chapter 1 about the failure of moral features to individuate subjects. That is, since it does not intuitively matter if the features that constitute persistence also individuate the subject from others, could it be that individuation is not as big a problem for persistence as I at first suspected? At the very least, this intuition suggests that there may be a way to reconcile what I have been calling the individuation problem with a necessary moral persistence condition. Altogether, I think that these are strong reasons to believe that moral features might have an important, constitutive relationship to our persistence.

But not so fast. The data offered by Strohminger and Nichols are only about what ordinary people *believe* constitutes persistence when presented with a series of sometimes far-fetched thought experiments. As what people believe does not always accord with the way things actually are, we must be cagey in deciding what these intuitions really show us. Strohminger and Nichols are in good company in the persistence-identity tradition, though. Dating at the very least back to Locke, talk of persistence has been heavy with thought experiments, 'puzzle cases', and fanciful stories meant to act as pumps to prime our intuitions about what constitutes identity. The purpose of these thought experiments is not to offer substantive proof of any one approach to persistence over others. After all, the fact that there is no such substantive proof is a prime reason why these thought experiments are invoked in the first place. Instead, what philosophers hope to show with these examples is that a particular approach seems more plausible than others because it can accommodate our intuitive beliefs about what persistence consists in (in most cases, this amounts to intuitions about who is who in instances where persistence is ambiguous).

Contrariwise, philosophers like Bernard Williams (1970) and Kathleen Wilkes (1988) argue that we ought not put so much stock in this kind of hypothetical scenario that they claim is probably *ad hoc* and may even elicit conflicting intuitions.[20] In other words, there is a question about the worth of not just some particular intuitions but of the entire methodology of drawing conclusions from intuitions prompted by thought experiments. If I am to use Strohminger and Nichols's results as an empirical basis for my argument for a necessary moral persistence condition, I will need to justify both the way I intend to make use of the Strohminger and Nichols data and that this methodology is even philosophically aboveboard at all.

2.2 Intuitions About Persistence and Why They Are Problematic

In this section, I consider the way in which intuition-priming thought experiments are regularly used to draw firm conclusions about persistence. I do this so that I may later argue, *pace* Williams, Wilkes, and others, that intuition-priming thought experiments are methodologically sound—when appropriately utilized (i.e. when not used as 'proof' of a metaphysical claim). This will allow me to justify my use of the Strohminger and Nichols data as a springboard to a more philosophically rigorous theory of a necessary moral persistence condition. In the present section, I first introduce a number of popular thought experiments important to the persistence debate. Then, I present arguments from Williams and Tobia that draw attention to potential methodological problems with the use of such thought experiments.

[20] I do not consider here another set of objections about our persistence intuitions made by thinkers like Paul Ricoeur (1992) and Marya Schechtman (2014). They argue that certain thought experiments—Ricoeur calls these 'technological' fictions—are simply inconceivable to us, though we *think* we can conceive of them properly and then draw intuitions from them. For a good discussion of these objections, see Beck (2016).

Both objections target possible bias in the way thought experiments are structured, which would then compromise the stability and credibility of the intuitions they elicit. Williams argues that the way a thought experiment is 'framed' leads the respondent to a particular intuition, resulting in inconsistent and even contradictory intuitions. Similarly, Tobia contends that people tend to regard positive psychological changes as person-preserving but think that subjects do not persist through negative psychological changes. If this is correct, then the 'direction' a change is given in a thought experiment—that is, whether the change is positive or negative—could have an unsavory influence on the intuitions it generates.

2.2.1 The Traditional Role of Thought Experiments and Intuitions

My purpose in starting the section with a series of important thought experiments is twofold. First, this will give the proper background needed to examine the objections raised by critics of intuition-based thought experiments. Second, examining thought experiments will help to demonstrate the central role such thought experiments have historically played in the arguments of persistence theorists.

Locke (1975) arguably begins this methodological tradition with the story of the Prince and the Cobbler:

> For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions (1975, 340).

Here, Locke thinks it obvious that most people would consider the prince-in-the-cobbler's-body to still be the prince. This is supposed to count as evidence for a psychological persistence condition, as it is the prince's psychological features—especially memory—that are in this case alleged to account for the identity claim that the subject in the cobbler's body is the prince. And this is true even if the fact that the prince no longer looks like himself (and instead looks like the cobbler) may be considered evidence *against* such a psychological persistence condition. So, the purpose of the thought experiment is to give the reader an intuition—that the prince-in-the-cobbler's-body is actually the self-same prince from before the body-swap—and the psychological approach is then invoked as the best explanation for this intuition.

More recent thought experiments advanced by advocates of the psychological persistence conditions sometimes shed the more fantastical elements of Locke's example but retain the 'body-swap' idea at its core. Probably the most well-known of these is the 'brain transplant' scenario (still fantastical, of course—at present, at least, brain transplants are not possible—but certainly more grounded than Locke's soul-transfer). Though the specifics sometimes differ—it may be the entire brain that is removed, or just the cerebrum (Olson 1997), or even the whole head (Parfit 2012)—all of these examples owe their start to Sydney Shoemaker's 'Brownson' (1963, 23-4). In this case, two men, Brown and Robinson, undergo brain surgery that requires brain extraction. However, a sloppy assistant mixes up the brains such that Brown's brain ends up in Robinson's body (and vice versa). The subject with Brown's brain and Robinson's body survives, while the other subject rejects the transplant and dies immediately. The remaining subject, whom Shoemaker calls 'Brownson', retains all of Brown's memories, but none of Robinson's.

Though Shoemaker himself is reluctant to pronounce that Brownson is identical with Brown, neither is he entirely resistant to the idea.[21] Others, though, were quick to take up brain transplants and their variations as the paradigmatic cases of intuitions about persistence. These philosophers maintain that most readers in transplant cases will affirm that the identity of the

[21] It should be said that Shoemaker later comes to accept that Brownson really is Brown (1970) and thereafter makes use of the transplant argument as strong intuitive evidence in favor of a psychological approach to persistence.

subject goes with the brain. Further, they hold that the reason for this is that the brain is the seat of psychological features like memory, character, and belief. So, as with Locke, transplant cases are meant to give intuitions about persistence that are best explained by a psychological approach.

Opponents of the psychological approach have their own set of thought experiments that generate contrary intuitions, however. Drawing upon a story attributed to C.S. Pierce (1935, 355), Roderick Chisholm (1969) suggests a counter to Locke's memory condition. He presents the case of a hypothetical and painful operation wherein you are offered two alternatives. The first is a simple anesthetic, but this is rather expensive. A much cheaper option would be to take a special drug just prior to the surgery that would induce full amnesia; a second drug taken afterwards would restore those prior memories but overwrite any memories of the operation (and, of course, the pain). Chisholm believes it ought to be obvious that, despite a lack of your prior memories, it would be *you* that suffers during the operation in the latter case. From this intuition, then, we should conclude that a psychological persistence condition just will not succeed, as it cannot do the work of explaining this contrary intuition.

Bernard Williams (1957) asks us to consider a bizarre situation in which a man named Charles apparently has all of Guy Fawkes's memories. This includes those memories that are historically verifiable and some other memories that, while not verifiable, seem appropriate to what is known of Fawkes's character and in fact offer plausible explanations to gaps in historical knowledge. Williams thinks that it might intuitively seem that we ought to say that Charles just *is* Fawkes. But we should resist this conclusion. The reason for this is that it seems perfectly possible that two men—Charles and his brother Robert—could both have all of Fawkes's memories and would thus have an equal claim to being identical with Fawkes. Certainly, both men cannot be Fawkes, who was just one man; and it would surely be arbitrary to pronounce one or the other alone to be Fawkes. For Williams, the most plausible explanation is that the two have just become qualitatively *like* Fawkes (and thus not identical with him). But "[i]f this would be the best description of each of the two, why would it not be the best description of Charles if Charles alone were changed?" (Williams 1957, 239). Williams's point is that intuitions about psychological continuity—memory in particular—do not apply in the Charles/Robert case. And so memory cannot be a persistence condition.

This argument, what has come to be called the 'Reduplication Problem', formed the template for a series of cases intended to elicit intuitions out of line with a psychological persistence condition. However else they may differ, all 'reduplication' cases share the same basic setup: a single subject is supposed to persist in virtue of relating in some way psychologically to a later subject, like Fawkes's relation to Charles. Then, a third subject is introduced—Robert, in the above—such that the original subject psychologically relates in the same way to both later subjects at the same time. Since one subject cannot later persist as two distinct and simultaneous subjects, reduplication cases are meant to show that the psychological relation in question, whatever it is, cannot be a condition of persistence.

A pair of representative examples of reduplication cases come from Wiggins (1967, 53) and Chisholm (1969). Wiggins reworks Shoemaker's Brownson so that, rather than a brain transplant with two patients, there is only a single surgery on Brown wherein the corpus callosum is severed and the two halves of Brown's brain are put into new bodies. Although Wiggins assumes that the two post-transplant hemispheres are equipollent and so both have access to the same memories (among Brown's other psychological features), we need not necessarily make this assumption. That is, we could say that the two hemispheres each inherit different psychological features, or perhaps some features are shared and some are not. Altering the example in this way does little to affect the result because what matters is that both post-transplant hemispheres are apparently psychologically continuous with Brown. But Brown I and Brown II are two subjects, not one—so it does not look like they can both be identical to Brown. Choosing one or the other to be Brown still seems unprincipled, too.

We can even take things a bit further. Suppose a careless doctor drops the Brown I hemisphere before it can be transplanted. Can we say that the Brown II hemisphere is just Brown now, since there is no competitor that shares Brown's psychological features? As Perry says, "It is natural to reply, on Wiggins's behalf...[w]hy should who I am be determined by what is going on elsewhere in the world—the presence or absence of a competitor to the identity of the person whose thoughts and actions I remember?" (2002a, 129). All of these puzzles apply equally to Chisholm's (1969) fission case, another form of reduplication which posits that the subject can split in two like an amoeba.[22] Wiggins and Chisholm here challenge the force of the psychological intuition by presenting a situation—the split-brain transplant—that generates a conflicting anti-psychological intuition. Like the transplant case, the split-brain and fission cases have become tried-and-true examples meant to generate intuitions about persistence—this time *against* a psychological persistence condition.

2.2.2 Why Our Intuitions Might Be Unstable and Biased

It seems then that our intuitions about persistence are not univocal and can even conflict—some thought-experiment-prompted intuitions suggest a psychological persistence

[22] Chisholm's case is somewhat different from later uses of fission in that he does not assume that both split subjects have the original subject's psychological features. He also argues, contra Williams and Wiggins, that it is perfectly plausible in reduplication cases to think that one but not the other of the reduplicated subjects could be identical with the original subject.

condition, while according to others a psychological persistence condition would lead to manifest contradictions. Williams (1970) argues that the source of these conflicting intuitions may be the way the cases are framed. Consider two subjects who are going to have their brains reprogrammed, each to match exactly the mental states the other had prior to the operation.[23] We might call this a mental state swap. After the mental state swap, one of the subjects will receive a large sum of money and the other will endure tremendous pain. Williams says that our intuition about who is who after the swap seems to depend on how the scenario is described to us.

For example, we might explain the case this way: Smith's mental states are copied onto Jones's brain (and vice versa). Effectively, this would mean that after the operation there is a person in Smith's body with Jones's mental states and a person in Jones's body with Smith's mental states. Before the swap, Smith is asked which of these post-swap persons she would prefer to get the money and which he would prefer to be painfully tortured. On this third-personal description, Williams believes that we would advise Smith to say that the person in Jones's body should get the money and the person in Smith's body the pain. In other words, Williams thinks we are inclined to answer that the person in Jones's body is actually Smith. This is because the third-personal description comes across as a case where Smith and Jones have swapped bodies, and this prompts the intuition that the subject goes wherever her psychological states do—in Smith's case, right into Jones's body. So, the third-personal description elicits the intuition that persistence is a matter of psychological continuity.

[23] Williams also supposes that this argument would apply to a more 'radical' Shoemaker-style brain swap instead of just copying mental states. His point is that the conflicting intuitions generated do not depend on any particular means of exchanging psychological states, just that two subjects apparently completely trade all of their psychological features.

On the other hand, if the case is described as the first-personal experience of a single subject, we get a very different result. What Williams has in mind here is this. Suppose you are told that you will endure an operation in which your memories, character, and other psychological features are wiped out. You will then receive the psychological features of another subject and will in fact think of yourself as this other subject. After this, you will undergo a procedure that will cause you incredible pain. Should you fear the pain? When presented with *this* version of the thought experiment, Williams holds that the clear intuition is *yes, I should fear the pain*. But a psychological persistence condition could not explain this intuition, as you have none of your original psychological features at the time the pain is caused. Indeed, this case counts as evidence against the psychological approach, as it conjures up an opposed intuition.

Although Williams's example is quite different from Wiggins's and Chisholm's, his comments on first and third-personal frames-of-reference can be (and have been) generalized to apply to the use of thought experiments full stop. *All* thought experiments are framed in such a way as to "produce a situation which would naturally elicit" the desired response from the reader (1970, 90). But Williams's point is that we might always describe the situation differently and get a conflicting answer. Suppose I redescribe Chisholm's case of the splitting of the amoeba-man as the production of two *clones*. This would suggest that perhaps the post-split subjects are not really psychologically continuous because a clone is already considered to be a numerically distinct being from the original. In this way, a thought experiment designed to produce intuitions against the psychological approach could be used to support it.

Even though Williams comes down on the side of the anti-psychological intuition as the stronger and more natural of the two, his argument here makes one rightly question just how far such intuition-priming thought experiments ought to count in determining persistence conditions. Kevin Tobia (2015) brings our attention to a related problem with thought experiment-based intuitions by pointing out another possible bias in the way certain thought experiments are presented. Taking as his starting point the case of Phineas Gage—the railroad worker who in 1848 suffered a tamping iron through his head which reportedly resulted in an extreme change (for the worse) in character and other psychological features[24]—Tobia conducts a series of empirical questionnaires about a Gage-like subject who experiences an accident and has resultant psychological changes. In one case, a once-kind subject becomes unfathomably cruel after the accident; in another, a cruel subject is found post-accident to be kind and considerate.

Tobia also considers one of Parfit's (1984) examples. Parfit tells the story of a Russian nobleman who fears the loss of his peasant-friendly political ideals as he ages. To this end, he drafts a legal document (which can only be revoked with his wife's consent) guaranteeing his inherited lands will be given to the peasants and asks his wife to never allow him to revoke the document. The nobleman believes that loss of these ideals will be tantamount to his death. Again, Tobia constructs a pair of questionnaires around this example. In one case, the nobleman is worried about losing his peasant-friendly ideals. A second case features a peasant-hating nobleman who fears he will lose his disdain for the peasants and the accompanying desire to keep his land from them. In both cases, the nobleman drafts a legal document expressing his wishes and exacts a promise from his wife that she will keep them, regardless of what the nobleman may say to the contrary at some later date.

[24] There is a question about whether Gage's purported changes in character, or the degree to which he was famously claimed to be "No longer Gage" actually held true (Macmillan & Lena 2010). Whether or not Gage's accident really caused profound psychological changes, one can imagine a case where a similar event *does* cause such changes. So, Tobia's use of the Gage example here seems perfectly principled.

In both the Gage-like and the Russian nobleman scenarios, questionnaire respondents 'agreed more strongly' with the idea that the subject did not persist over time when his psychological changes seemed to be degenerative. That is, when the Gage-like subject became cruel after his accident, or when the nobleman later came to despise the peasants, many respondents thought that the original subjects were not numerically identical with the subjects whose character traits were what we might consider 'worse'. On the other hand, respondents 'agreed more strongly' that the same subject *did* persist over time when his psychological changes seemed to be for the better. So, when the Gage-like subject's accident caused him to be kind, respondents saw this as merely a qualitative change in one and the same subject (likewise for the nobleman who later wants to provide his peasants with land). In both cases, it looks like respondents have an intuitive bias. Positive psychological changes are seen as qualitative only, while negative psychological changes are interpreted as indicative of genuine change in identity. There is thus a bias in what Tobia calls 'direction of change'.

Tobia argues that these data have important implications for our intuitions that result from evocative thought experiments about persistence. If direction of change bears no relation to personal identity, these data show that our intuitions in these cases—perhaps especially those that gesture at the psychological approach—are biased (and so probably are not a good indicator of the metaphysical facts of persistence). Conversely, if direction of change does meaningfully have something to do with personal identity, these data suggest what that relation may be. In other words, these data would be possible evidence of an asymmetry in direction of change that shows that identity can sustain a greater degree of positive change in psychological features than of negative change.

2.3 How We Are Methodologically Justified in Using Our Intuitions About Persistence

We have seen in the previous section both the vital part that the thought-experiment-to-intuition method plays in the persistence debate and some good reasons to think this method dubious. In particular, bias in the way problem cases are presented may undermine the stability of our persistence intuitions and their credibility and usefulness in guiding us to plausible accounts of the metaphysical facts of persistence. How then are philosophers interested in persistence to regard these intuitions and the thought experiments that give rise to them? We could do as Mark Johnston (1987) and Kathleen Wilkes (1988) urge and just abandon the intuitional methodology altogether. But I think Carol Rovane (1998) has a better answer for this. There is a temptation to take these intuitions as evidence for (and against) some persistence condition or another. That is, thought experiments that produce intuitions about persistence can sometimes be so persuasive that they seem like proof. Given the ways in which these intuitions conflict, however, it looks like each intuition can be met by a contrary intuition from a different case. The only way to let such intuitions count as proof is if some persistence condition can be posited that satisfies all or most of these varied intuitions, which seems unlikely. Rovane therefore argues instead of 'strict proof' for persistence derived from our intuitions, we ought to just be "seeking *positive* reasons to embrace one side over the other, and to revise our commonsense beliefs accordingly" (59).

In this section, I follow Rovane in holding that our persistence intuitions do not provide proof of the metaphysical facts of persistence. Yet, I also draw the preliminary conclusion that our intuitions about persistence—if not unstable or biased—should not be dismissed lightly and could lead us to 'positive reasons' to favor a particular persistence account. Quite generally, we put stock in our intuitions because they often precipitate success. So, while recognizing that our persistence intuitions do not count as hard evidence of some persistence condition or another, I do believe the bare fact that we have such intuitions in the first place makes them worth investigating. And since we are looking only for positive reasons from our persistence intuitions and not strict proof, all that needs to be shown is that our intuitions are stable and unbiased (and so might lead to a plausible persistence theory).

If I am right, this means that I would be justified in positing a persistence theory with a moral persistence condition based on the intuitions revealed by Strohminger and Nichols, *if* those intuitions are both stable and free from bias, *contra* Williams and Tobia. As such, the rest of this section consists in responses to Williams's and Tobia's objections. Based on data from Nichols and Bruno, I argue that our persistence intuitions are not as contradictory as Williams claims because they can be given a more 'abstract' frame of presentation that removes experimenter bias. Against Tobia, I argue that direction of change bias—even if true—does not disprove the credibility and usefulness of our intuitions in guiding us toward a theory of persistence.

2.3.1 Why Our Persistence Intuitions Are Not Actually In Conflict

Nichols and Bruno (2010) offer data that challenge Williams's claim that our intuitions about persistence are often contradictory due to the way problem cases are framed, and I think that their data and conclusions allow us to reject Williams's concerns. As others—like Rovane[25]—have pointed out, Williams's claim casts doubt on the usefulness of both our intuitions about persistence and, even moreso, the problem cases that prime those intuitions. Curious whether the evidence supports Williams's claim, Nichols and Bruno conducted a series of surveys based on the problem cases Williams uses in his argument. Beyond his specific talk

[25] See also Sider (2001).

of first and third-personal frames, Williams can be taken to argue the more general point that problem cases are designed to elicit a particular intuition or at least lean heavily in one direction over another. Nichols and Bruno call this problem ‘thought experimenter bias’ because philosophers usually introduce thought experiments to bank intuitive goodwill from the reader toward some argument the philosopher is making.[26] And the data appear to support Williams’s concern here. Respondents tended to favor a psychological explanation in cases that used psychological language to frame the scenario and a somatic explanation in cases that used physiological language.

That there is apparently bias in the construction of problem cases does not definitively establish that either the problem case methodology or our intuitions about persistence are faulty, however. Next, Nichols and Bruno conducted surveys around problem cases designed in an attempt to remove these biased frames of presentation as much as possible. First, they presented respondents with an abstract question about persistence that made no reference to psychological or physiological features, and a majority of respondents answered that persistence was a matter of the preservation of some psychological feature or another. For their experiment, Nichols and Bruno provided respondents with the following open-ended prompt:

> One problem that philosophers wonder about is what makes a person the same person from one time to another. For instance, what is required for some person in the future to be the same person as you? What do you think is required for that? (Please pitch your explanation at a very simple level—don’t use any words that might be unclear) (304).

[26] While Williams thinks that problem cases with a psychological frame—that is, that describe the scenario in terms of psychological features like memory—are especially guilty of thought experimenter bias, Nichols and Bruno argue that the opposite is true. Williams’s somatic frame asks the reader to determine if the subject will be the one to feel the pain of a medical procedure after being induced to lose all of her memories. For Nichols and Bruno, this stacks the deck in favor of a somatic-friendly response because “if [the subject] isn’t going to feel the pain, then *who* [is]?” (2010, 304). So, the very question ‘Will the subject feel the pain?’ invites a response that the subject persists through the procedure.

Anticipating that such open-ended questions cannot always be counted on to produce more accurate responses than concrete ones, Nichols and Bruno also conducted a final survey that counterbalanced the abstract question with a more concrete one based on Williams's somatic-framed case. Respondents were expected to draw conflicting intuitions about persistence from the two questions and so were asked to engage in reflective equilibrium to determine which of their two opposed responses they believed more strongly. At this point, the prompt plainly told respondents "that it wouldn't really be consistent to say both that you would feel the pain of the shots and also that in order for a person to be you, that person must have some of your [psychological features]" (306). Again, a majority of respondents favored a psychological explanation for persistence.

These data reveal that our intuitions about persistence from problem cases only seem to conflict when problem cases suffer from thought experimenter bias. Nichols and Bruno propose two ways to address this. First, a more abstract question might remove thought experimenter bias altogether by presenting a neutral frame. In the event that the abstract question still directs the respondent to a particular answer, the respondent can be deliberately presented with conflicting intuitions and forced to confront them in a state of reflective equilibrium. When bias is removed in this way, psychological persistence intuitions appear to dominate. To Nichols and Bruno, this suggests that our intuitions about persistence may not be contradictory and in fact tend toward the psychological approach. Based on Nichols and Bruno's data, then, it looks like our persistence intuitions are stable after all.

2.3.2 Why Direction of Change Bias Does Not Discount Our Intuitions

Even if our intuitions are stable, though, Tobia gives the separate criticism that they could be biased and thus not credible or useful indicators of an accurate persistence theory. Tobia examines folk responses to psychologically-framed cases that may seem especially relevant to my project, as these involve changes in the moral psychological features of subjects. In scenarios involving a Phineas Gage-like person who suffers a traumatic head injury and Parfit's Russian nobleman concerned about whether or not the serfs will inherit his lands, the subject either becomes cruel (when he was kind) or kind (when he was cruel). Whether the direction of change is cruel-to-kind or kind-to-cruel, both represent a significant change to psychological features, so both ought to evoke the same sort of psychological intuitions about persistence. Yet, the data show that a majority of respondents only regard change from kind-to-cruel as indicative of a change in the subject. That is, most respondents thought the subject persisted through positive changes to moral psychological features but did not persist through negative changes. Tobia's studies into direction of change problem cases end in a disjunction: either direction of change is important to persistence in some way, or our intuitions about persistence—and this is particularly true of psychological intuitions and, most problematically for my project, psychological intuitions having to do with moral character—are loaded with bias.

If Tobia is right, this certainly would seem to undercut any attempt to ground an investigation into the possibility of a moral persistence condition in our intuitions. However, I do not believe that Tobia's data reveal quite what he thinks. I remain largely reticent on the first disjunct—whether or not direction of change might affect the persistence of a subject—though I admit it looks *prima facie* metaphysically unlikely. Setting aside the first disjunct, I am still unsure that a direction of change bias keeps our persistence intuitions from being credible and useful indicators of the metaphysical facts of persistence.

Now, one answer to this issue would be to construct problem cases not susceptible to direction of change bias. This seems easy enough to do for a number of psychological features, including memories, beliefs, and desires, because those features can be qualitatively neutral toward one another. Suppose my overriding life's desire is to protect and conserve non-human animal life. Replacing this with an equally strong desire to develop a cure for cerebral palsy does not look to be qualitatively better or worse on most accounts. Likewise, consider memory-based problem cases like those invoked by Williams, Parfit, and others. A subject who loses all of her memories and acquires entirely new ones might seem to us to no longer be the same person, but this has nothing to do with traits we evaluate as good or bad. We could stipulate that both sets of memories are exclusively of happy events (or are, at least, equal on balance in terms of pleasant to unpleasant memories) such that there would be no way to deem one set qualitatively better than the other.[27] So, our intuitions about persistence here seem unrelated to direction of change, as there is really no qualitative movement from positive-to-negative or negative-to-positive. Rather, the intuition that the subject does not persist seems to come from the fact that the present subject is disconnected from the original subject in such a way as to make the two seem relevantly distinct. Therefore, it appears that we can construct problem cases that are not vulnerable to the problem of direction of change but still produce persistence intuitions.

But this avenue is not available to all types of psychological feature. In particular, moral features do not appear to be qualitatively neutral toward one another—which would be especially

[27] Strange though it may be, I think we *can* talk of memories this way. After all, most of us would probably prefer a memory of playing with a beloved pet to one of a man gunned down in the street.

damning for my project. Moving from kind to cruel or stingy to generous is obviously value-laden, but the same is also true of seemingly more benign moral changes. Perhaps someone is only slightly sensitive to the suffering of others, but them becomes merely indifferent to this suffering. This is not a major qualitative leap on par with going from sensitive to, say, a kind of schadenfreude. Nonetheless, there is a clear sense in which one trait is better or worse than the other. So, changes to moral features still appear to be subject to direction of change bias.

Despite this concern, Tobia's notion of direction of change seems undeveloped. It looks to me that direction of change bias is reflective of a psychological bias we have toward traits we find desirable. In other words, we have a tendency toward thinking subjects are the 'same' when they undergo moral change for the better—by this I mean that they develop traits we deem positive—because we want to believe that they can improve to what might be called the 'ideal state' we have in mind for them.[28] On the other hand, we see them as 'different' when they move away from this ideal and begin to express traits we consider negative. Of course, what traits are considered negative may vary between subjects. If the Russian nobleman's wife were as cruel as he is in some of Tobia's examples, she would probably think a 'positive' change toward kindness would be quite negative!

But even if it is true that what is considered 'better' or 'worse' might not be uncontroversial to all respondents, this is by no means a 'killing blow' to Tobia's argument. There could still be direction of change bias if the way we define positive and negative change is highly relative. At the same time, if what respondents take as desirable is relative in this way,

[28] 'Ideal state' may mean a good many things and incorporate features other than moral character traits; for instance, I may want my spouse to share my love of dystopian science fiction. The point here is simply that people do not regard such changes—however systemic—as affecting persistence. Perhaps this is because we already psychologically connect the subject as she is with the subject as we would ideally have her, but I would rather not speculate too much here.

that relativism could skew the data. At the very least, it would mean that we should expect different sample groups to produce extremely variable results depending on the traits a given group finds desirable. My point is just that to correctly interpret Tobia's data, I think direction of change needs to be better defined and more data ought to be obtained.

There is also the question of what to do when psychological features compete. We will use an example that Tobia cites as "a particularly striking demonstration of the [direction of change] effect" (2015, 401): Charlie Gordon from *Flowers for Algernon*. Citing Michael Shapiro (2005) and B.W. Cline (2012), Tobia claims that Charlie—a man with severe cognitive disabilities who undergoes a surgery that renders him exceptionally intelligent—is seen to persist through the substantial psychological changes of his surgery. However, the surgery is ultimately and tragically unsuccessful. Once his cognitive abilities begin to deteriorate to their former level, Tobia says that Charlie is regarded as 'dying' in a certain sense. That is, a new subject emerges due to discontinuity in Charlie's psychological features. This is an incredible asymmetry: Charlie is viewed as the 'same' through cognitive improvements, but when he deteriorates, he has somehow changed into a different subject—even though his 'deterioration' is just a return to the exact same features Charlie had before!

Now, Charlie's procedure makes him substantially smarter but also somewhat cold and selfish. If we grant that being more intelligent is indeed a boon, are we still to call post-operative Charlie an 'improvement' if he is morally a worse person than he was before? There is change happening in two conflicting directions in this case, so how are we to evaluate the sum direction of change? Again, just because Tobia's theory does not give us clear answers here does not mean that there is no direction of change bias. However, I do think it shows that, conceptually, direction of change needs to be further explored or elaborated upon.

As this relates to a possible moral persistence condition, I argue that the limitations of Tobia's theory should at least give us pause in how decisive we take his data to be. Of course, even if it is definitively true that our persistence intuitions are compromised by direction of change bias, there still could be a moral persistence condition. In such a case, our intuitions would be correct by happenstance, but correct all the same. Yet, my goal is to justify an investigation into a necessary moral persistence condition on the basis of folk psychological intuitions about such a moral persistence condition. So, the relevant question is not 'If Tobia is right, could there still be a moral persistence condition?', but 'If Tobia is right, does this invalidate moral persistence intuitions?' I still think the answer is 'no'. In order for direction of change bias to completely invalidate at least some of our persistence intuitions, the data would need to show that certain of our persistence intuitions—including moral persistence intuitions—are in the majority of cases the result of direction of change bias. But the data do not in fact show this.

From Tobia's data, we can at best extrapolate that there might be a direction of change bias that affects our persistence intuitions about particular problem cases. This urges us toward caution in the use of these intuitions in making philosophical hypotheses—and this is exactly what I do above. My claim is that our persistence intuitions may lead to positive reasons for the metaphysical facts about persistence. They are not proof that such-and-such is metaphysically true. Following our persistence intuitions could lead us nowhere at all. However, even if there is a direction of change bias in a small range of problem cases, this certainly does not encourage us to engage in a kind of blanket skepticism about all persistence intuitions (assuming those intuitions are otherwise stable). So, Tobia's data, even if correct, do not prove that our persistence intuitions are not credible or useful, but instead (and more weakly) suggest the very prudence with which I treat our persistence intuitions here.

There are other reasons to doubt direction of change bias and its impact on our persistence intuitions. For example, we might question just how widespread direction of change bias actually is. To return to *Flowers for Algernon*, I simply do not share the intuition that the ‘improved’ Charlie is the same subject. His entire relationship to the rest of the world has changed, including what place he occupies in it, his awareness of himself, and how he understands and relates to other people. On certain accounts of personhood, pre-operative Charlie would not even count as a person (or would only count marginally, at best). For instance, pre-operative Charlie could easily be argued to fall short of the rational, intelligent, self-aware, and feeling Lockean person. Thus, the asymmetry that Tobia argues for in what he seems to take as a paradigmatic case of this asymmetry strikes me as incredibly unnatural, and I suspect that there are many people who would agree with me—not least of all those who adopt the Lockean view of the person.

There is no need to just take my word for it, though. We already have empirical data that supports the idea that the direction of change bias may not be so widely shared as Tobia claims. Recall the Strohminger and Nichols (2014) studies from the first section of this chapter. In one study, respondents were presented with a hypothetical pill that would selectively remove psychological features, then tasked with evaluating to what degree removing particular features would result in a subject ‘being the same person’ or ‘completely different’. Removal of negative moral character traits—in particular psychopathy, pedophilia, and criminality—were considered by respondents to be more indicative of a persistence change in the subject than positive traits like empathy, conscientiousness, and virtuousness. Given Tobia’s data on direction of change, we should have expected the opposite. That is, since negative direction of change is more strongly associated with a severing of persistence than positive direction of change, a subject who lost one of these negative traits ought to have been viewed as staying the same—or at least of changing less than a subject who lost a positive trait.

It should also be noted that, even if Tobia is right and there is a direction of change bias favoring positive changes in moral character traits, this does not mean that positive changes are completely dissociated from a change in subject. Consider another of the Strohminger and Nichols studies from the first section, which presented respondents with a scenario in which they meet an old friend now in advanced age. Respondents were asked to indicate the degree to which changes in certain psychological features the friend now exhibits (but did not before) would most constitute a change in identity. Although negative moral character traits like racist, cruel, and rude were considered to have the greatest impact on the subject's persistence, positive traits, including forgiving, honest, and generous, were also assessed as having a substantial effect. These three positive traits were actually rated as more impactful on persistence than certain negative traits: namely, 'more likely to steal' and adulterous. And all moral character traits, both positive and negative, were considered to have a greater effect on persistence than any other psychological feature.

So, while there did appear to be a small bias in direction of change in this study—change from positive character traits to negative ones elicited more of a sense that the subject was no longer the same—respondents still regarded positive direction of change as highly persistence-severing to the subject. This speaks against Tobia's results, which suggest that positive direction of change should be seen as person-preserving. The fact that some cases of positive change were rated as more persistence-severing than certain negative changes also counts against Tobia's conclusions, as does the finding that both negative and positive changes in character traits were considered more persistence-severing than all other psychological features.

To recapitulate, I think there are three important reasons to be skeptical of Tobia's findings. First, the data are too limited to prove that a direction of change bias would invalidate our persistence intuitions. At most, the data support being careful about how we use persistence intuitions in making philosophical claims; I argue that my use in this project is perfectly in line with what he recommends. Second, the direction of change bias may not be especially widespread; there is empirical evidence of cases that should be vulnerable to direction of change bias that actually result in intuitions contrary to Tobia's claim: in these cases, positive direction of change is assessed as *less* person-preserving than negative direction of change. If it turns out that only certain cases evoke direction of change bias, then we can simply screen out those problematic case descriptions. That Tobia's expected results were inverted in the Strohminger and Nichols study hints at another possibility, too. Given a wider range of case descriptions and experiment population size, it may also be the case that this 'bias' evens out. Some cases and respondents may favor positive change, and others may favor negative change. But if we conduct more Tobia-style experiments with slightly different descriptions and larger pool of respondents, the average result may actually show no preference toward positive *or* negative change in character traits. Third, a bias in the way Tobia proposes may not ultimately sabotage our psychological intuitions about persistence. Empirical data show cases where a bias is present but does not skew the overall folk perception that changes in moral character traits, whether positive or negative, are meaningfully persistence-severing. That is, in such cases, even if there is a bias, the distance between positive and negative is not all that great. If all of this is right and there really are conflicting data about direction of change bias, then the real conclusion we should draw is that a verdict here is an empirical matter that demands further study.

2.4 Toward a Moral Persistence Condition

The purpose of the foregoing chapter is to justify the use of folk intuitions about persistence as a basis for my argument for a necessary moral persistence condition. To this end, I presented data from Strohminger and Nichols that suggests that people intuitively believe that the most important constitutive features of persistence are moral character traits. Importantly, Strohminger and Nichols conclude from this data that these features are both valued as person-preserving for their own sake and that other person-preserving features are often valued because of the way they contribute to moral features. To me, these intuitions suggest that a necessary moral persistence condition is plausible. Given concern in the literature about the acceptability of the use of intuition-based thought experiments, I also considered objections from Williams and Tobia to the thought-experiment-to-intuition methodology that has long been a staple of persistence-philosophers. But I did not find these arguments convincing.

Against Williams and Tobia, I argued that it is perfectly justifiable to use intuitions about persistence to inform and ground my investigation into a necessary moral persistence condition. Based on the findings of Nichols and Bruno, it seems reasonable to say our intuitions about persistence are not actually contradictory. Indeed, when triggered by problem cases presented from an unbiased frame, these intuitions consistently tend toward the psychological approach. So, our persistence intuitions are stable. Likewise, the current data appears too limited to support Tobia's claim that our intuitions about persistence (especially psychological-leaning intuitions) are truly biased in a way that undermines the credibility and usefulness of our persistence intuitions. Since I argued in Chapter 1 that a moral persistence condition must be part of a psychological theory of persistence, this is good news for my necessary moral persistence condition.

Now, none of this means that our intuitions about persistence prove that there must be a moral persistence condition, nor that a persistence account that includes a moral persistence condition is the right account of persistence. I think I have closed off some relevant objections, however. As far as our intuitions go, my goal with this chapter was to show that methodological worries like Williams's and Tobias's do not give us good reasons to do away with our persistence intuitions; in my case, these are intuitions about a moral persistence condition. Because moral persistence intuitions—like those from the Strohminger and Nichols studies—are in fact both stable and free from bias, we can justifiably pursue those intuitions further and look for positive reasons for a persistence theory that includes a moral persistence condition.

I move on in the next chapter to make good on a promise from Chapter 1 to connect personal ontology with persistence. This may appear to the reader to be an odd tangent to take at this point in the project. However, following Olson (2007), I think that a proper account of our persistence must first say what kind of thing we are before it can feasibly hope to determine how we persist. So, at the risk of being like the man who tries to get himself out of a hole by digging it deeper, I trade identity for ontology—shovel in hand.

CHAPTER 3

A MORAL PERSONAL ONTOLOGY

In Chapter 3, I take a detour from the work of investigating the possibility of a moral persistence condition in order to lay what I argue is the proper metaphysical groundwork for any kind of persistence account that could sustain a moral persistence condition. Specifically, my goal in this chapter is to propose an account of personal ontology that can accommodate a moral persistence condition—as suggested by the empirical evidence presented in Chapter 2. I call this the State-Realizer Account (SR). As the name might imply, SR holds that we are entities individuated both by mental states and the physical substrate that realizes those states. Though it is a novel view, SR is based on Parfit's Narrow Psychological theory of persistence.

This chapter comes in four sections. First, I justify why talking about personal ontology matters for a dissertation that is ostensibly about persistence. I show that persistence theories—theories of personal identity—tacitly presuppose accounts of personal ontology, even though persistence and personal ontology are conceptually separable. Therefore, it follows that clearly articulating one's personal ontology beforehand gets important metaphysical assumptions out of the way and keeps one from mistakenly conflating persistence and ontology. In the second section, I articulate SR. As part of this view, I add that the capacity for self-consciousness is a necessary condition for something to count as one of us. Together, I argue that state-realizer individuation plus self-consciousness implies that we are essentially persons on SR. In the third section, I distinguish SR from four competing accounts of what we are and give reasons for why we should prefer the SR for the current project. These competing accounts are: the Embodied

Parts view, the brain theory, material constitutionalism, and hylomorphism. Finally, I close the chapter with SR's implications for one of us coming into and going out of existence. Some of these implications are quite surprising and may be difficult for many to accept. However, I argue first that they are at least consistent with SR and are thus coherent if that view is. Second, I make the case that some of these implications are not limited to SR and are indeed more common than they first appear. As a result, I conclude that these implications may not be so surprising or deleterious to my position after all.

3.1 Who Cares About Personal Ontology?

In this section, I argue that there is a conceptual distinction that must be recognized between the persistence of persons and personal ontology. In short, I do not think that defining a person's persistence conditions will also double as an account of what we are. It might seem like an account of the persistence conditions of persons would also tell us what we are, as Shoemaker (2011) has argued, among others. Given the traditionally accepted Lockean definition of 'person', it is easy to see why. On this prevailing view, persons are intelligent, rational, sentient, and self-conscious; a person will therefore persist insofar as these features of personhood are preserved. Because these are all *psychological* features, it stands to reason that one would expect the persistence of persons to have a psychological explanation, too. So, if something is a person, it persists in virtue of the continuity of certain psychological features. In other words, whatever else we can say about them, persons are a psychological kind of thing—they are *psychological continuers*.

How does that tell us what *we* are? Well, if you are reading and understanding this chapter, I can assume that you are a person. And as the author of this chapter, it is likewise a safe bet that I am a person, too. Things like you and I are persons, and there are not many philosophers who would dispute that. (In fact, the presumed truth of that proposition is where much of the personhood discussion starts.) If *we* are persons, and persons are psychological continuers, then *we* are also psychological continuers. Hence, from our definition of person and an account of the persistence conditions of the person, we learn what we are.

There is more to it than this, however. You and I may be persons, but could either of us ever cease to be a person and still exist? That is, are we *essentially* persons? To answer this question, we can talk about which properties we have essentially and which we have only accidentally. Consider an example to illustrate the point. We could describe Dennett with the properties 'being a philosopher', 'being a sailor', 'being a husband', 'being a bearded man', 'being a person', and 'being a human animal'. Are any of these properties ontologically essential to Dennett? Presumably, Dennett was once a child and was not, at that point, a philosopher, a sailor, someone's spouse, nor at all bearded. Such properties then describe states that Dennett existed in temporarily. But Dennett the child was still an animal—and he may have been a person, as well. It thus appears that 'being an animal' and 'being a person' are more metaphysically essential to Dennett than 'being a philosopher' is.

Following Wiggins (1967, 1980), we can distinguish broadly between two types of concepts: substance sortals and phase sortals. Being a philosopher, sailor, husband, bearded man, and even a child are all clearly phase sortals for Dennett—they are accidental properties that could come and go without changing the subject's ontological nature. As Olson puts it, "[t]o come to be a philosopher is not to come into existence *simpliciter*, and to cease to be a philosopher is not necessarily to cease to exist altogether" (1997, 29). For instance, Dennett could decide to abandon philosophy for competitive dirt-bike racing, but this would not amount to his death. Clearly, the question of what we are is not about phase sortals like ‘philosopher’ and ‘child’. Rather, in trying to determine what we are, we are interested in what particular kind of thing we fundamentally are. This is the domain of substance sortals. In this way, we might rephrase the question as ‘which substance sortal correctly applies to us?’

Now, ‘being an animal’ and ‘being a person’ are properties that are more important to something being Dennett than are the properties ‘being a philosopher’, ‘being a child’, or ‘being a bearded man’. It seems less obvious that Dennett could cease to be an animal or person and yet continue to exist. We could consider each possibility separately. Suppose that ‘person’—in the Lockean sense described earlier—is a substance sortal and applies to Dennett. Then Dennett is essentially a person. There are a number of implications here: first, Dennett could not endure certain psychological changes. He was also never an embryo, as such beings lack all of the Lockean features of personhood. Moreover, Dennett could not survive loss of cerebral function and fall into a persistent vegetative state, as he would cease to exist when his relevant psychological features do (in this case, during brain death). Contrarily, suppose that ‘animal’ is a substance sortal and applies to Dennett. Then Dennett is essentially an animal, and no amount of psychological change will affect his biological persistence. He was once an embryo—because all animals are, early in their lives—and he could survive in a persistent vegetative state.

As far as substance sortals go, nothing can both essentially be an animal and essentially be a person, as the two can come apart. Suppose that Dennett is essentially a person. Then ‘animal’ must become a phase sortal, as it is conceptually possible for one person to be multiple animals throughout his life. To see why, consider the ‘brain transplant’ thought experiment. Here, Dennett’s brain is placed in a different body entirely but retains all psychological features. As Olson (1997, 2007) observes, no single animal would make the journey from Dennett’s skull to the new body. Instead, there would be two animals: Dennett’s former body, now brainless and sustained by life support machinery, and the new donor body into which Dennett’s brain is transplanted. We can slightly modify the experiment:

> *The Rejected Transplant*: There is a society in which brain transplants are fairly routine, but transplant technology is nonetheless highly (yet predictably) imperfect: the new bodies of transplant patients always reject the brain after a determinate amount of time, requiring a new transplant every two years. Dennett’s brain is transplanted into a new body, but must be continually transplanted every two years for the rest of his life.

Of course, *Rejected Transplant* would now mean that Dennett becomes a new animal every two years. So, although Dennett’s brain is transferred from animal to animal many times, Dennett the Lockean person survives throughout. Similarly, if Dennett is essentially a person, his brain could be installed into a *Robocop*-style cyborg body, whereby, as Obi-Wan says of Darth Vader in *Return of the Jedi*, he would become “more machine now than man”. Again, we see here that ‘animal’ is a phase sortal—Darth Dennett is no longer an animal of any kind, as he would no longer perform characteristic animal functions (metabolizing external matter, circulating blood and oxygen, etc.). However, Dennett (the person) continues life as a cyborg.

On the other hand, the inverse is true if Dennett is essentially an animal. That is, ‘person’ becomes a phase sortal and Dennett could continue to exist yet become multiple persons throughout his life. Consider a device that completely eradicates Dennett’s psychological features—a ‘mind wipe’ of sorts—but then implants new psychological features in place of the old. On the accepted Lockean view of the person, the discontinuity between the old and new features would constitute two distinct persons that, over time, inhabit one and the same animal. Again, we modify the experiment:

> *Mind Wipe Therapy*: There is a society in which mind wipes are used as a kind of therapy for individuals unsatisfied with their lives. For a small fee, your psychological features can be ‘overwritten’ by new features of your choosing—the catch is that no features can be retained at all. Fed up with a life of philosophy and sailing, Dennett pays for the mind

> wipe therapy and emerges as a raucous dirt-bike racer. Some time later, he tires of dirt-bike racing; yearning for a different direction, he again pays for the therapy. This process continues for the rest of Dennett's life.

In *Mind Wipe Therapy*, each use of the procedure would result in the emergence of a unique person. Despite this, there is only ever one animal involved. If Dennett is essentially an animal, 'person' is a phase sortal—a temporary state that Dennett (the animal) can live through. So, because the person and the animal seem to persist in different ways, nothing can be both essentially a person and essentially an animal.

While the above examples plainly show that there is tension between whether we are essentially persons or animals (or perhaps something else with a different metaphysical status), my point in this section is not to commit to an account of what we are. Right now, I simply aim to demonstrate that personal ontology is in fact separable from persistence. This does not mean that the two are completely independent, however. In the next section, I show that some personal ontologies can be a better or worse fit with certain persistence theories—for instance, a theory that includes a moral persistence condition.

For the moment, though, let us stick with the conceptual separability of personal ontology and persistence—and why noting this separability is important. If we try to answer the persistence question without first considering what we are, or if we try to answer the persistence question thinking it will answer the question of what we are for us, there is significant risk of question begging. For instance, to start—as I did in the first paragraph of this section—with the premises that we are Lockean persons with psychological persistence conditions might encourage an assumption of person essentialism. But, even if it is true both that we are persons and that persons have psychological persistence conditions, it might also be true that we are persons only contingently. That is, describing a person's persistence conditions still does not fix whether person is a substance sortal or a phase sortal—it is still possible that we might not be persons essentially.

This is true of other possibilities for what we are, too. Assuming person essentialism likewise seems to rule out our being immaterial Cartesian egos or souls. Why? As Locke points out in Chapter XXVII of the *Essay*, we can conceive of a single soul that once occupied the body of, say, Nestor at Troy but now resides in a person with a completely different psychology (and no memory of its previous 'life' as Nestor). As with 'animal', 'soul' essentialism looks to be incompatible with person essentialism—one soul could become multiple persons. So, considering personal ontology first gets these assumptions out on the table instead of illicitly sneaking them into our persistence theories, and it forces us to see that persistence and personal ontology are distinct concepts.

3.2 We Are Psychological Beings With a Constant Realizer

In this section, I state my account of personal ontology, SR. Again, this view holds that we are both sets of mental states and the constant physical substrate in which those states are realized. Although it is in many ways evidently similar to several existing theories—namely, the Embodied Parts approach, the view that we are brains, material constitutionalism, and hylomorphism—I think that my own view is importantly distinct from each of these. To this end, I believe it best to first state my view and its philosophical basis, then in the next section demonstrate what differentiates my view from these competitors. An important point: my aim in this section is not to argue that SR is *the* correct account of what we are. I only want to posit a plausible personal ontology that is compatible with a moral persistence condition. If I cannot formulate such an account, it will be clear right away that any intuitions about a moral persistence condition are unfounded simply because there could be no being that is ontologically such that it could persist in virtue of its moral character traits. But there may be other, better accounts that can meet the demands of a moral persistence condition. It is also important that an account of personal ontology that can accommodate a moral persistence condition not suffer from obvious defects. Clearly, if SR was manifestly sloppy metaphysics, that it satisfies a moral persistence condition would be little reason to prefer it to competing views about what we are. In this way, my reasons for positing SR will be (a) to formulate an account of personal ontology that is compatible with a moral persistence condition and (b) to formulate such an account that avoids some well-known problems of similar, competing views.

3.2.1 Basic Contours of SR

SR can be stated as follows:

> *The State-Realizer View*: we are (1) a suitably connected totality of psychological states that (2) are realized in a constant physical substrate. Further (3) for something to be one of us, there must exist as part of (1) a capacity for self-consciousness, or the capacity to self-ascribe psychological states.

(1) and (2) affirm that we are individuated in two ways: first, by mental states, and second, by the physical substrate in which those states are realized. (1) and (2) are necessary conditions for something to count as one of us. On the one hand, I think that the survival of *just* the physical realizer—sans mental states—does not in itself entail the survival of one of us. On the other, we are not *just* mental states of course, as states are instantiations. Hence, states must be instantiated in something in order to exist at all. Observe also that, on SR, states must be instantiated in the *same* something (a *constant* physical substrate)—even if SR does not stipulate what exactly that something is. Note that this formulation does not explicitly call us persons. Whether we are persons or not is part of what is up for grabs. SR provides the individuation conditions for one of

us. It might turn out that these individuation conditions fit our chosen definition of what counts as a person. Yet, we cannot assume that we are persons at the front end. Similarly, SR does not reduce what we are to human beings—we might not essentially be that, either.

Before considering (3), I need to acknowledge influences on my (1) and (2) in SR from another theory. My account builds on the groundwork of what Parfit calls the Narrow view of psychological continuity. As the Narrow view has even more importance to my arguments in Chapter 4, I will only offer a sketch of it here. Should the reader require a more complete explanation of Parfit's theory, please see section 4.1 of Chapter 4. Now, on the Narrow view, persistence is a matter of psychological continuity—cashed out in terms of overlapping chains of 'strong' psychological connections—without branching and with the 'normal' cause. By normal, Parfit has in mind *the ordinary functioning of the same brain*. Perhaps surprisingly, this implies that physical continuity of the realizer (the brain) still matters, even though the Narrow view is a psychological theory. Persistence is defined by continuity of mental states, but there would be no states nor continuity without their constant realization in the same physical substrate. Compared to the Narrow view, SR is not really concerned with the normal cause *per se*—just that a (constant, persistent) substrate realizer is a necessary condition for the manifestation of mental states. Still, there are clear structural parallels between Parfit's view and mine that need to be acknowledged.

What I am doing in defending SR is to essentially reverse engineer Parfit's Narrow Psychological account of persistence by asking what sort of personal ontology must be true in order to permit Narrow Psychological persistence conditions. Any personal ontology that could make possible an account of persistence that roots persistence in both psychological states and a common physiological causal component would have to allow that our metaphysical character is both psychological and physiological. To my mind, the simplest way to do this is by co-opting the components of the Narrow Psychological account and say that we are individuated by psychological states and constant physical realizers.

(3) in SR concerns self-consciousness. In addition to individuation by way of mental states and their physical realizers, SR states that a capacity for self-consciousness is a necessary condition for something to be one of us. Minimally, (3) takes this to be a capacity for the self-ascription of mental states. What this tells us is that something that lacks the particular realizer structure to bring about self-conscious states cannot be one of us—even it has an otherwise rich mental life. To be clear, I am not suggesting, like higher-order theories of consciousness,[29] that all consciousness is a kind of self-reflective behavior or awareness of one's own mental states. Indeed, beyond the minimal notion of 'self-ascription of mental states', I remain neutral here on what self-consciousness (or consciousness, for that matter) really consists in so that SR might be compatible with as many theories of self-consciousness (and consciousness) as possible.

The reader might ask what motivates the inclusion of (3) into SR, as a capacity for self-consciousness puts a heavy restriction on what counts as one of us. For it to be worth the ink spilled to write it, any account of what we are must capture what is distinct about *us*. And it is indisputably true that self-consciousness is a distinguishing feature about us (though this is not to say that we are the only sort of thing that might lay claim to this capacity). I follow Nagel (1986) in this belief, who argues that our 'true nature'—what we are—must be "what makes [us] capable of subjective, nonobservational self-knowledge extending over time" (42-43). So, since accounts of personal ontology should capture what is characteristic to the nature of the thing it is trying to define, and since self-consciousness is such a feature for *us*, a theory of *our* ontology must in some way account for self-consciousness as constitutively necessary. Nonetheless, I

[29] See, for example, Armstrong (1984), Carruthers (1996), Lycan (1996), and Rosenthal (2005).

admit that (3) is a *strong* condition. My claim that self-consciousness is a necessary condition for something to count as one of us has crucial implications. Namely, it changes our ordinary thoughts about when one of us comes into existence or out of it and how we relate to things like our bodies, the children who precede us, and the human animals deeply connected to each one of us. I consider these implications in the last section of this chapter.

Before moving on, though, an analogy may give some clarity to SR (it is one that I employ throughout the rest of the project). One helpful way to think about SR might be in terms of patterns and their instantiations. Patterns are abstract concepts that can be instantiated concretely in physical things. This instantiation gives the pattern a kind of spatiotemporal reality—a place and particularity—that the pattern would otherwise lack. The way in which the words are arranged on this page is one sort of pattern, and the specific features of the arrangement tell us which of many possible patterns has been instantiated. For instance, whether the sentences are single or double-spaced, the size of the margins, and the sort of font used collectively form the pattern of the page. Of course, this pattern could be instantiated elsewhere, on a different page, with different words, and in a different book altogether. As far as SR goes, we might say that each one of us at a given time has a ‘pattern’ that is the totality of our mental states at that time. Both the physical substratum, arranged in a way that realizes exactly those mental states, and the individual mental states themselves are the way this pattern has been instantiated. In the case of SR, instantiation is a necessary condition for there being a pattern at all. To SR, our psychological patterns are not something that exists separately from the individual mental states that comprise the pattern and the physical insantiations in which those states are realized. If we think of SR in this way, then SR says that we are instantiated patterns, where both the instantiation and the pattern in part constitute one of us. But unlike ordinary patterns, because our ‘pattern’ is just its states and their realizers, the pattern that is each one of us cannot be instantiated elsewhere. I say more on this in Section 3.2.3.

3.2.2 Dispositional and Occurrent States, and Humean Bundles

Above, I presented SR as a view on which we are individuated by mental states and a physical substrate that realizes these states. In this subsection, I argue that the mental states that in part individuate us must be dispositional states and not occurrent states. Otherwise, we would be something like Humean bundles of perceptions, and this result, I claim, would have many unwelcome consequences. To introduce the distinction between dispositional and occurrent states, I begin by discussing what is perhaps a common misconception about our mental states. Most of us probably think that our mental states are persistent. That is, we think that my memory today of visiting Disney World as a six-year-old and my memory yesterday of visiting Disney World as a six-year-old are the *same* memory. Similarly, we probably also think that things like beliefs and desires are continuously held. If I have the belief that ‘raccoons are pretty neat’ today and I also had the belief that ‘raccoons are pretty neat’ yesterday, that is because the very same belief I held yesterday is also held today. Yet, the vast majority of our mental states are not persistent in this way.

Most of our mental states are not occurrent—that is, actively thought of—but dispositional and unconsciously held. For example, until writing this sentence, I did not have the occurrent belief that opossum cannot breathe fire. Yet, based on other beliefs I held, I could be said to have had this implicit belief in a dispositional way.[30] But I can actualize this dispositional belief to reflect on it occurrently, and the same is true for other mental states. What this means is that most of my mental states—like my belief that ‘raccoons are pretty neat’—persist only for as

[30] This was pointed out by Dennett (1978, 104), although his example concerned zebras and overcoats.

long as they are occurrently held. Unless I happen to actively think about how neat raccoons are for twenty-four hours, today's belief and yesterday's are numerically distinct actualized *occurrences* of a disposition *to believe that raccoons are pretty neat*. Consequently, this regular succession and exchange of occurrent mental states suggests that, for us to be partly individuated by mental states in SR, the constitutively relevant mental states are *dispositional* states rather than occurrent states.

If we were individuated by occurrent states instead, SR would be scarcely distinct from Hume's well-known claim in Section VI of the *Treatise* that we are bundles of perceptions (where 'perceptions' includes what we would today call mental states). In other words, we would be sets of occurrent states, and hence would persist only as long as the members of the set did not change. Even swapping out another occurrent instance of the same dispositional mental state would bring about a new set and a new and numerically distinct one of us. However, Hume's 'bundle' approach is unacceptable to the present project. This might not be immediately clear. After all, this approach *is* compatible with a moral persistence condition—a change in any moral trait would result in a numerical change in subjects, as the bundle would now be a new collection of perceptions with at least one new member. But this theory has many well-documented problems.

In the first place, consider that on bundle views, the realizer of mental states would not at all be part of what we are. This is bad news for SR, which accepts physical realizers as partly constitutive of us. Suppose I am a set of occurrent mental states. Then I could not persist through any change in the set of occurrent states that constitute me. My mental states constantly change from moment-to-moment, so even the smallest successive change in the membership of the set—say, the set includes a perception of a white door, but now that perception is replaced by a perception of a checkered floor—will bring about a different subject and be the end of *me*. All of this remains true whether or not there is a constant realizer present. Though we might be tempted to claim that a constant realizer would allow me to persist through things like benign perceptual changes in occurrent states, this would entail that I am the realizer and not the set of occurrent states. Such a set has particular persistence conditions. If I do not share those conditions, in what way am I that set? So, if we are sets of occurrent mental states, we are not even partially realizers. In turn, this means that we are not even partially anything physical.

Granting the foregoing, if we are sets of occurrent mental states and not even partially anything physical, it is difficult to ascertain what sort of thing we are. Call this the *Uncertain Existence* problem. On bundle views, if we are something—say, a substance—then it would have to be a purely mental something, since we are not even partially physical. This would seem to presuppose a commitment to mind-body dualism, as we would end up being Cartesian egos or something rather like them. But this is not exactly right, either. The whole point of an ego or soul is that it is the mental substance that persists through mental state changes, *a la* Butler (1736). Yet, for bundle views, there is no underlying mental substance but just a succession of state changes. This is, after all, part of Hume's point: there is no persistent mental subject that unites a constantly changing bundle of perceptions. Hence, on bundle views, we would have to be a mental something, but that mental something would not be a persisting substance.

If this succession of changing states is not itself a substance, and so we are neither mental nor physical substances, there are few options left for what sort of thing we could be. In fact, it seems like we would have to be states of some *other* substance, in this case. To return to the language of sortals I used before, we would be like *phase sortals*—properties that another being, like a human animal, assumes for a temporary period. Of course, this would mean that *we* are not beings at all, then. We would just be properties. I think that it would be rather surprising to find out that we are not beings but are instead temporary properties of things that actually *are* beings, in the way that my shirt temporarily has the property of being colored red until it fades to a mild pink. Bundle views thus commit us to existence as properties, but also imply that we are not even partially whatever substance those properties are instantiated in! Beyond the *prima facie* implausibility of this conclusion, it also is not at all clear what it would actually mean to say that you and I are properties. I can understand the proposition 'being tall is a property of giraffes' without any trouble, but I can find no way whatsoever to make sense of 'Parfit is a property of a certain human animal located roughly here'. So, bundle theorists hardly have a convincing answer for the Uncertain Existence problem.

A further concern is that one's existence on bundle views would be so ephemeral that it might be spurious to say there is ever something that is 'me' that exists. We can call this the *Ephemeral Selves* problem. Again, this is why Hume famously asserts that there is no 'self' available to introspection. Sets of occurrent states change members so frequently that set membership is constant only for a vanishingly small period. As such, it is tempting to follow Hume and call the idea of distinct and persisting selves a fiction. To many, though, this is intuitively unsatisfying. It certainly *seems* to us that we exist over time. It would at least be very unexpected if it turns out that a numerically distinct entity comes into existence (and another goes out of existence) when, say, one's focus shifts from thinking about turtles to thinking about the score in last night's football game. As a result, bundle views offend our felt experience of how we persist and so would make for an intuitively unfulfilling answer to what we are.

Olson (2007), drawing from Reid (1785), raises another major concern with bundle views. How can a bundle of mental states be both the states and the thinking subject that *has* those states? On the face of it, this sounds absurd. Let us call this the *Thinking Bundles* problem. Olson argues that accepting that bundles of mental states can be subjects in this way requires the bundle theorist to accept something even more absurd: that individual mental states are mental subjects. In other words, he says that if the bundle theorist wants to claim that bundles can think, believe, and desire, then she must also accept that thoughts think, beliefs believe, and desires desire.

On the one hand, the suggestion that a bundle of mental states can think but no individual state does seems clearly wrong because a bundle of mental states is nothing other than an aggregate of individual mental states. Putting together several unthinking mental states does not produce a thinking subject any more than putting together several sheets of music produces a sound. So, if no individual mental state thinks, then neither does a bundle of mental states. On the other hand, if we suppose that a bundle of mental states *does* think, then it is difficult to reject the notion that individual mental states can think, too. A single thought is just as much a thought as an aggregate of thoughts is, after all. There seems to be thinking going on in both cases. Adding more thoughts to form a bundle would appear to be a matter of quantity, not quality; that is, it is not evident how merely adding more thoughts could change something from a mere mental state to a bona fide mental subject. So, if a bundle of mental states thinks, then individual mental states do, too. This sounds clearly false, however. Olson thus concludes that it is better to say that a bundle of mental states cannot think. Of course, since we are quite clearly thinking things, then this means that *we* cannot be bundles of mental states.

A final worry is that—contrary to my initial evaluation above—a bundle view may not actually be compatible with a moral persistence condition. That is, it is unclear how on this view moral traits could be importantly distinguished from any other mental state, at least insofar as those states relate to persistence. A change in *any* occurrent mental state results in a change in the set of occurrent mental states—and, therefore, the emergence of a numerically distinct one of us. So, one of us cannot persist through a change in moral traits just because one of us could not persist through a change in *any* mental state. But, if this is the case, there seems to be no way to give moral traits any special status in determining persistence. This is because all changes in occurrent mental states are equally deleterious to persistence.

I do not think that restricting persistence-severing change to occurrent expressions of moral traits would help matters, either. If only changes in moral traits resulted in a numerical change in one of us, then we wouldn't be sets of occurrent mental states. Instead, we would be constituted by a *subset* of some set of occurrent mental states—namely, that subset consisting of all and only occurrent expressions of moral traits. Of course, this subset would be no better insulated against the above criticisms than the greater set of all occurrent mental states; these problems would just be moved down one level. This 'occurrent moral traits' view has the added worry of being a philosophically unprincipled move. Perhaps we could explain this move as motivated by our intuitions about a moral persistence condition, but to do so would also render all non-moral occurrent mental states impotent in determining persistence. To do this would not do justice to the very same intuitions we are here attempting to serve. After all, the prevailing intuition about a moral persistence condition is not that moral traits are the *only* mental states that constitute persistence. So, there is no obvious way to give occurrent expressions of moral traits special status on a bundle view, which means that bundle views are not compatible with the moral persistence condition suggested by our intuitions in Chapter 2.

Instead of calling us bundles, what if we stipulate that we are *successions* of bundles of mental states that have a kind of continuity over time due to some shared/overlapping states between successive bundles? On the surface, this sounds like what Parfit calls the Wide and Widest view of psychological continuity. As we will see in the next section, these versions of Parfit's view come with at least one irreconcilable problem. Independent of Parfit's arguments, though, the notion of successions of bundles does not get us much further than individual bundles would. Successions would still just be aggregates of aggregates, so most of the above problems transfer up a level but do not disappear. This *would* resolve the Ephemeral Selves problem, as our persistence over time is secured through continuity between similar successive bundles. But it is still unclear what we are on such a view—still properties of something else, perhaps—and it is still unclear if sets of sets can think (I doubt they can, for the reasons cited above). Hence, the Uncertain Existence and Thinking Bundles problems continue to threaten the bundle view, and successions of bundles do not fare much better than single bundles. More broadly, given the difficulties presented by the Ephemeral Selves, Uncertain Existence, and Thinking Bundles problems, as well as the fact that the standard bundle view does not actually allow for a meaningful moral persistence condition, I conclude that the mental states involved in determining what we are must be dispositional states and not occurrent states. (Unless otherwise specified, I will hereafter just use 'mental state' in regards to SR to mean 'dispositional mental state'; it will just keep things tidier this way).

3.2.3 Constant Realizers and Parfit's Wide/Widest Views

In addition to specific mental states, SR says that part of what we are is a constant realizer. There are several reasons for this requirement. First of all, the need for a constant realizer clearly allows us to make sense of the fact that there is a something that has the mental states in question—thus avoiding the Uncertain Existence problem. Properly speaking, of course, we are not the realizer, but the entity consisting of the relevant states and the underlying realizer. But by marrying the states and realizer, SR skirts the embarrassing problem of having an entity that is constituted by more or less free-floating states.

Most importantly, the constant realizer requirement avoids some of the troubling implications of Parfit's Wide and Widest views of psychological continuity. Like the Narrow view, both the Wide and Widest views define persistence in terms of psychological continuity, where this continuity consists in chains of overlapping psychological connections over time. Where these positions differ is that the Wide and Widest views accept a *wider* range of what is acceptable as the cause of psychological continuity. On the Narrow view, continuity only results from the ordinary cause of mental state production and maintenance—for us, normal brain functioning. 'Normal brain functioning' can include purely physiological operations, like the brain's physics and chemistry. But this normal functioning can also include processes and causes external to the brain, too. For instance, one 'normal' cause of continuity might be the production and maintenance of mental states through the typical physiological upkeep of the brain, while another equally 'normal' cause might be the modification of select states (and preservation of others) through ordinary learning. 'Abnormal' causes would then be cases of obvious physiological manipulations—like a drug that bypasses typical function to produce certain states—or environmental manipulations—like hypnotic suggestion that simply 'implants' particular states.

However, the Wide and Widest views allow any reliable cause and any cause whatsoever, respectively, to account for psychological continuity. Significantly, Parfit argues on all three views that gradual change in mental states can be persistence preserving because such minor changes do not ultimately disrupt psychological continuity. Despite changes to a few mental states, the vast majority of psychological connections persist, and so there remains an uninterrupted chain of overlapping connections. However, when we consider both his acceptance of this kind of gradual change *and* the Wide and Widest views, a serious problem emerges:

> *Gradual Brain Washing*: I am the victim of a nefarious and uncommonly future-oriented mad doctor. Posing as my regular psychiatrist, this mad doctor is somehow able to slightly modify one of my mental states during my weekly visits, perhaps through a kind of hypnosis or mental conditioning. Over the course of ten years, he gradually changes my moral character traits so that I am uncaring, highly prejudiced, cowardly, and deeply selfish. (Let us assume that I was a virtuous individual before these sessions—caring, open-minded, courageous, and concerned primarily for the wellbeing of others.) Suppose also that the mad doctor has implanted the requisite beliefs so that I consider the above changes to my moral character to be both natural and of my own free will.

Now, many of us intuitively want to say that I am not the same person at the end of *Gradual Brain Washing*; that is, we want to say that I do not persist through these moral-psychological changes. We think it seems that I ought not be held responsible for the vicious actions and character of the individual at the end of the thought experiment. Of course, this is because all of these moral changes were the product of brainwashing—outside manipulation that I certainly would not have consented to—instead of normal brain functioning. Although the individual at the end of the story might consider these changes natural and freely chosen, this 'natural' feeling is as much a result of manipulation as the moral changes. Gary Watson (1996) distinguishes between two senses of responsibility: attribution (who actually performed the action) and accountability (who should bear the consequences of it). But neither of these senses of responsibility seems to appropriately designate *me* as the responsible party at the end of *Gradual Brain Washing*. Rather, some new person has come into existence, and it is this person who can be rightly said to have the vicious character described.

Yet, on the Wide and Widest views, it appears impossible that I am not that individual. The mad doctor's means of altering my mental states *does* seem to be a reliable cause; likewise, psychological continuity is not disrupted, as a single changed mental state hardly severs the necessary overlapping chains of connectedness between the other (at that time) unaltered mental states. On the other hand, if we are partially individuated by a constant realizer, we get the right answer here. I cannot be the individual at the end of the story because my change in character was not brought about in the usual way. My change in mental states was caused by a mad doctor's tinkering, not through normal brain functioning made possible by a persistent physical realizer. So, the requirement for a constant realizer sidesteps the Gradual Brain Washing problem, which makes SR preferable to the ontology recommended by the Wide and Widest views.

The reader might still be unconvinced why there is a particular need for a constant realizer rather than, say, something more specific. For each one of us, this realizer is presumably a human brain. So, why not insist that we are in part *brains* instead of realizers? Mostly, this is because, on SR, we are essentially persons. That is, we are necessarily individuated by our mental states, and—following the traditional and widely accepted Lockean view of persons as rational, intelligent, sentient, and self-conscious—these selfsame states are the characteristic features of personhood. What this means is that another way to state what we are on the SR view is that we are *persons* constantly realized in a particular way—namely, via physical brain states. Since we are essentially persons, a general account of the metaphysics of the person will also double as an account of what we are. Keeping the more general 'realizer' in place of 'brain' permits multiple realizability in the sense that it leaves open the possibility of inorganic persons, or persons whose states are not realized in an organic brain. After all, there might be things other than brains that can realize mental states.

Of course, this is not full-fledged multiple realizability, as my realizer cannot change. Contrary to many science fiction stories, my consciousness could not be ‘uploaded’ onto a synthetic brain surrogate and the resulting entity still be me (see Shoemaker 1984, later recanted in his 2004). To see this, we can consider something as simple as a flash drive. When I transfer information—like this dissertation—from the drive to folders on a PC, it looks like I move one and the same information from one place to another. That is not really what happens, though, if we look at it more closely. In actuality, the dissertation is *copied* to a new destination. Like words written on a page, the information that comprises the dissertation must be realized somewhere and in something. But nothing moves between the drive and the PC except electric current. In other words, no realizer moves between the drive and the PC. So, the dissertation cannot move during the ‘transfer’ as there is nothing in which it could be realized that is transferred.

Consciousness is no different. The mental states that comprise consciousness must be realized somewhere and in something. Just like this dissertation and the flash drive, though, my consciousness cannot move unless its physical realizer does. But the whole point of science fiction ‘upload’ stories is that one and the same consciousness is supposed to move from one realizer to another. This means that ‘transferring’ my consciousness to a synthetic brain surrogate would amount to copying the information of my consciousness to a new destination. In this case, the ‘information’ is a pattern that consists in the totality of my mental states at some time or another. Of course, the copy is not one and the same consciousness. So, on SR, my realizer must remain the same. I cannot be uploaded to a synthetic brain, nor could I persist through a swapping out of substantial chunks of the relevant part of my brain for other brain tissue. Otherwise, SR risks the problem of having a Theseus's ship for a brain![31] Yet, SR does have other significant consequences for the restrictions typically placed on what we are.

For one, if we are essentially persons, it is at least logically possible that mental states could extend beyond the physical brain. SR stipulates that our mental states must be constantly realized in the same brain, but this does not mean that the brain could not be assisted in cognition by external artifacts *a la* extended mind (Clark and Chalmers 1998, Sutton 2010) or distributed cognition theories (Hutchins 1995a, 1995b). In other words, since we are only our realizers because those realizers instantiate our mental states, it would be conceptually coherent for some of our mental states to be instantiated in other realizers appropriately related to the brain (however that might be). This would be contingent on theories of cognitive extension being true, of course.

Yet, even if one refuses to accept extended mind or distributed cognition theories and denies that our mental features can in any way be external to body and brain, SR at least implies that we have what Clark (2003) has referred to as a 'cyborg nature'. On extended mind and distributed cognition theories, persons form coupled systems with external artifacts to perform mental work. But consider that on SR we are effectively already *systems* of a sort—we are entities consisting of mental states and their realizers. So, it is hardly a stretch to think that we might join again with other things and enlarge the system that we already are. If SR is correct,

[31] Perhaps there is a process that involves exchange of brain material but that would also still allow for brain continuity. Suppose that I gradually replace tiny bits of my brain tissue with new tissue, but each time new tissue is inserted, a significant interval is provided to allow the new tissue to be integrated into normal brain functioning. McMahan (2002) discusses such a case. We might also ask if SR—given its functionalist leanings—would accept gradual integration of functionally identical synthetic tissue as 'the same brain'. That is, could the realizer brain be incrementally but completely replaced with an artificial but *ex hypothesi* functionally identical brain without severing persistence? This question is a little far afield for now, but my gut reaction is to say that SR would have to be committed to an affirmative answer. Let this footnote stand as my recognition that what counts as 'the same brain' is a separate and real question, but it is one that I do not engage with further than this paragraph.

being a system-like entity is part of our very nature. Given the above, the preference for ‘realizer’ over ‘brain’ in the SR thus serves an additional purpose: it highlights more precisely that we are first and foremost psychological entities and that whatever physical thing we are is important mostly as a necessary condition for the psychological.

3.3 Why We Ought to Prefer SR to Competing Views of What We Are

With SR established, I now turn to competing views about what we are. I cannot here consider all possible objections or alternatives—fully defending SR would be a book-length project on its own. The best I can hope for here is to present an account that is consistent, avoids problems of similar theories, and permits a theory of persistence that can accommodate a moral persistence condition. To this end, in this section, I examine four theories apparently similar to SR: the Embodied Parts theory, brain theories, metaphysical constitutionalism, and hylomorphism. I argue that, despite appearances, each of these positions is in fact quite distinct from SR.

Further, I contend that there are good reasons to prefer SR to these four alternatives. Competing views are either subject to substantial problems that SR avoids or turn out to be incompatible with a moral persistence condition. The former is obviously undesirable in any theory, but the latter proves especially damning here, as the goal of this chapter is to construct (if possible) a theory of personal ontology that would allow further investigation into a moral persistence condition. There can, of course, be no further investigation if a given theory cannot even support the thing hoped to be investigated.

3.3.1 Embodied Parts/Minds

One competing view of what we are is what Jeff McMahan (2002) calls the Embodied Minds view, though Parfit (2012) later picks up a broader version of this theory under the name Embodied Parts (EP) view which would also include McMahan's theory as a subset of the view. (Since it seems to be more general, I will stick to the Embodied Parts naming convention here.) As seems to be standard practice personal among ontology arguments, McMahan's view begins with comments about persistence, then analyzes what those comments presuppose about what we are. On EP, what matters for persistence is continuity of the *capacity* for consciousness *simpliciter*, not continuity between the mental states that constitute consciousness. McMahan says that we can understand 'same consciousness' in this sense to be "equivalent to the notion of same mind" (2002, 67).

McMahan argues that we ought to look for our persistence at a deeper level than simple continuity of psychological states. Specifically, he says that it is in the conditions that enable consciousness where we find our persistence. In McMahan's words, "a particular mind continues to exist only if enough of the brain in which it is realized continues to exist in a functional or potentially functional state" (67), where the chief function seems to be generating consciousness. There are thus on this view *two continuities* that are necessary and sufficient for persistence: physical continuity of the relevant parts of the same brain and functional continuity of that brain so that it can sustain consciousness. A third continuity—organizational continuity, or continuity of the particular arrangements and structures in the relevant parts of the brain that give rise to certain mental states—is immaterial to EP. Consequently, this means that one of us could survive mental reprogramming but not destruction of the relevant parts of the brain or cerebral death resulting in a persistence vegetative state.

In terms of personal ontology, it thus follows that McMahan's view presupposes that we are 'embodied minds'. There are other versions of this account that might get more specific than this. Parfit (2012), for instance, thinks that we are embodied in a particular part—the cerebrum. Strictly speaking, what we are on EP is some part of a human organism. Namely, we are its conscious, thinking part.

McMahan takes care to distinguish EP from two views that he calls versions of Parfit's Narrow Psychological view, independently advanced by Green and Wikler (1980) and Lockwood (1988, 1994). (Technically, both theories are theories of persistence rather than personal ontology; however, as I maintain in foregoing subsections, theories of persistence often make implicit assumptions about personal ontology.) Given SR's roots in Parfit's theory, I suspect that McMahan would try to pass off Green and Wikler and Lockwood's accounts as breeds of SR. However, I argue that neither of the proposed theories ultimately counts as SR; in fact, I think both actually reduce to EP, despite McMahan's objections to the contrary.

Let us consider Green and Wikler first. As one of the goals of their paper is to draw an equivalence between brain death and death of the person, it might seem like Green and Wikler have a 'brain' theory of what we are (see below in 3.3.2). Yet, they state that

> the ordinary causal processes which link events in a personal history involve more than spatio-temporal continuity of brain tissue. They also require continuity of certain brain *processes*...whatever processes there are which normally underlie that person's psychological continuity and connectedness (Green and Wikler 1980, 126-127).

So, we are not just identical with our brains, as something like what Perry (1976) describes as a 'brain zap'—a complete erasure of *all* mental states—would sever psychological continuity. To Green and Wikler, a brain-zapped Jones, bereft of all continuity with earlier brain processes, is not Jones anymore but just a leftover body. Jones is "gone and dead" (Green and Wikler 1980, 118). Note, though, that this is not ordinary psychological continuity understood as overlapping mental states. In fact, they argue that "[w]hat matters is the substrate, not the psychological states which it produces" (128).

The implications of this latter claim are brought out clearly when they consider a case from Bernard Gert (1971). Supposing that Jones has been both hypnotized to have all of Smith's mental states and that Jones's brain has been transplanted into Smith's former body, Green and Wikler seemingly accept it as obvious that the hypnotized, transplanted individual is still Jones. In this case, Jones's mental states have been completely overwritten by new states matching Smith's. Obviously, this is not ordinary psychological continuity. Somehow, Jones's brain processes—in the same continuous brain—make the transplant patient Jones and not Smith, even though just about the entirety of Jones's mental states have changed.

Now, what I think the above shows is that Green and Wikler do not have an SR theory at all. For one, the realizer or brain certainly adopts a more primary role in determining what we are than on my view. As Green and Wikler say, 'what matters' to them is the substrate, not the individual mental states. Jones *lost* nearly all of his mental states, but his substrate remained the same throughout—and it is the persistence of the substrate that leads Green and Wikler to conclude that Jones persisted through his reprogramming. This means that what *really* matters is that the realizer or brain is physically and functionally continuous. A brain zap—but not a near-total 'rewriting' of the brain's states by a hypnotist—is persistence severing just because the brain only fails to maintain its proper functionality in the former case.

Like McMahan's view—and *unlike* SR—I think that Green and Wikler claim that we could survive reprogramming because reprogramming would *only* affect organizational continuity: functional and physical continuity go on untouched in the Jones/Smith case. A Perrystyle brain zap, on the other hand, might just impact functional continuity, too. When Green and Wikler say that brain processes are constitutive of what we are, they could arguably be taken to mean *functional* processes. If so, then the 'psychological continuity' Green and Wikler discuss is just continuity of functional brain processes, or just plain functional continuity. Indeed, it seems to me that this is exactly what Green and Wikler take them to be. McMahan points to Green and Wikler's claim that the continuities that underlie psychological continuity are what are important to persistence as evidence that their view differs from EP. That is, McMahan argues that Green and Wikler are actually worried about organizational continuity, which is importantly not relevant to persistence on EP. But I find McMahan's reading to be uncharitable.

If Green and Wikler were in fact concerned with organizational continuity, then their distinction between simple reprogramming and brain zaps would not be coherent. After all, both reprogramming and a full brain zap result in massive organizational *discontinuity*, so both ought to be equally persistence severing. However, if Green and Wikler are talking about functional continuity in the case of the brain zap, we can make sense of this distinction. And, if this is right, Green and Wikler's account surely counts as a version of EP. Green and Wikler might even understand 'psychological continuity' to be something like 'continuity of consciousness' rather than continuity of individual states. In discussing anencephalic infants—children born with insufficient cerebral material to have higher cognitive functions—Green and Wikler's focus is on brain processes and the brain itself as "the substrate of consciousness" (1980, 128), and they immediately thereafter mention 'higher level psychological continuity'. This implies a hierarchy of psychological continuities wherein bare consciousness could be seen to occupy (one of) the lowest levels. Furthermore, Green and Wikler entertain the possibility of a coma patient who persists despite discontinuity in psychological states because "enough of the brain remained structurally and functionally intact" (128). If I am right, Green and Wikler would on this reading—like McMahan's account—value consciousness rather than continuity between states and prioritize only physical and functional continuity. Just like EP, this would make what we are out to be consciousness realized in particular parts of the brain; in other words, we would be 'embodied minds'.

We now shift our attention to Lockwood. Like Green and Wikler, Lockwood argues that our persistence cannot

> *consist in* such connections and continuities as are exhibited in consciousness and behaviour. Intuitively, one's identity over time is, I would suggest, conceived of as a *deep* fact: something we think of as lying behind these connections and continuities, something of which the latter are merely a manifestation (1988, 204).

Continuities between mental states are therefore only what Lockwood calls 'superficial' continuities. What really matters for making something one of us are 'deep' continuities whereby "[t]he essence of the person may, in that sense, survive a physical or emotional trauma that results in amnesia and sudden change of personality" (205).

Critically, though, these deep continuities are more than just physical continuity of the brain. Physical continuity is, for Lockwood, necessary but not sufficient for persistence. Lockwood goes on to claim that a subject whose mental states have been 'reprogrammed' to match those of a second subject would fail to survive, as "reprogramming would effect too radical a discontinuity of organisation in the parts of it that subserved mental functioning. What one would have done is make a new person that was a replica of the original person" (206). Despite physical continuity of the brain as an organ, the structures of the brain become thoroughly different in reprogramming; they would have to in order to bring it about that the mental states that depend on those structures and their arrangements are so altered. However, given Lockwood's comments about surviving through amnesia or sudden and complete change in personality, it seems clear that organizational continuity of the brain might—when paired with physical continuity—be sufficient for persistence, but not necessary for it. After all, brain structures would have to rearrange in order to effect changes in mental states like amnesia or personality shifts, too.

There is thus a third sort of deep continuity active in Lockwood's account. His comments on the emergence of consciousness can clue us in to the nature of this third continuity:

> At such time, then, as the brain has matured to the point of being able to sustain any characteristically mental functions, I would expect to find in place the neurophysiological substrate of such psychological continuities and connections as are to be found in consciousness and the behaviour we take to manifest mentality...before the brain has matured to the point of being able to sustain psychological functions, a human life has yet to begin (206-207).

Lockwood here echoes Green and Wikler: there is a causal and functional continuity that undergirds persistence. In particular, Lockwood ties this to consciousness *simpliciter*. Functional continuity, as a necessary feature of persistence, consists in the capacity of the brain to sustain conscious function. Ultimately, there are three 'deep' continuities—physical, organizational, and functional—that Lockwood says "alone are constitutive of personal identity" (205).

Although Lockwood only explicitly mentions organizational continuity, I have argued that the other two continuities are implicit in his characterization of his view of persistence. Translating this back into personal ontology, this means that what we are is brains with a specific functional structure for consciousness. In a similar way to Green and Wikler's position, Lockwood's account seems, on the face of it, to be a kind of SR: he focuses first on continuity of mental states, then argues that we must be at least partially constituted by the realizers of these states in order to metaphysically ground them. Again like Green and Wikler, though, Lockwood appears to favor the realizer over the states. This is evident in his division between 'superficial' state continuities and ‘deep’ realizer continuities. Moreover, Lockwood’s emphasis on the capacity for consciousness instead of self-consciousness is somewhat at odds with my view. SR takes the capacity for self-consciousness as a necessary condition for something to count as one of us. The mere capacity for consciousness *simpliciter* and perhaps a few proto-conscious states—the capacities that are important on Lockwood’s account—would thus seem to fail to satisfy SR’s conditions for something to be one of us.

Lockwood’s view is less obviously a case of EP. For one, Lockwood explicitly denies that one of us could survive reprogramming, which immediately puts it into contention with EP and seemingly implies a commitment to organizational continuity. On the other hand, it is clear that organizational continuity is not necessary for our persistence to Lockwood, as he argues that one of us *could* persist through amnesia and a sudden and complete personality shift. Both of these latter effects would, of course, be substantially organizationally discontinuous. Again, I suggest that the charitable reading is to take Lockwood’s comments about reprogramming to actually refer to functional continuity rather than straightforward organizational continuity. Though Lockwood does say that the underlying organization of the brain would change too much during reprogramming, we could understand this to mean that the *functional* organization would be disrupted and not merely the specific arrangement of brain structures that would support specific mental states in organizational continuity. Like McMahan, Lockwood thinks that consciousness marks when one of us comes into existence. Since the particular states which are the contents of consciousness do not seem to matter for persistence in this case—otherwise, amnesia and total personality shift ought to sever persistence—it must be consciousness as a capacity that is important here. Perhaps Lockwood believes that reprogramming momentarily snuffs out this capacity, only to reinstate a capacity for consciousness via the new arrangement of brain states that follow.

Like Green and Wikler, if Lockwood's account is to be coherent, I think we ought to understand him this way. If I am right, this would mean that much of Lockwood's account lines up with McMahan's EP. Crucially, our persistence would seem to turn on physical and functional continuity, not organizational continuity. Certainly, Lockwood's account diverges from McMahan's in the case of reprogramming, but I argue that this is a difference not about what sort of continuity is at issue, but what reprogramming in fact does or affects. But which continuity is important is what really matters to EP, and Lockwood and McMahan seem to be in agreement there. So, like Green and Wikler, if Lockwood prioritizes physical and functional continuity, his account presupposes that what we are is consciousness realized in a particular part of the brain—again, an 'embodied mind'.

Now that we have a full account of EP and I have defended my claims that Green and Wikler and Lockwood offer versions of EP instead of SR, I will distinguish EP more clearly from SR and argue why SR is preferable for the current project of investigating a moral persistence condition. First, although on both EP and SR we are parts of organisms, we are not the same part on both views. On EP, we are the part of the brain responsible for consciousness. On SR, though, we are the part of the brain functionally responsible for self-consciousness. We are also, in part, a particular pattern of organization of both mental and brain states. EP makes no metaphysical distinction between consciousness and self-consciousness. That is, something conscious with the right realizer is one of us on EP, but SR requires that the entity have the capacity for self-consciousness. Said differently, this is a capacity for very specific kinds of mental state—self-conscious states. But particular states do not matter on EP, so neither does the capacity for self-consciousness.

There are also differences between EP and SR in the case of reprogramming which are indicative of broader differences about what sort of events we can survive. SR takes reprogramming to be tantamount to the death of one of us. Part of what we are on SR is sets of mental states; as a result, the structure of brain states (organizational continuity) matters to what makes something one of us, since mental states depend in some way on those same brain states as realizers. We might say that the EP account of what we are is conscious (or potentially conscious) brains, but the SR account has us as parts of brains *and* certain patterns of organization. Here, the major difference between EP and my view is that EP values only physical and functional continuity, but on SR, all three continuities—physical, functional, and organizational—are individually necessary and jointly sufficient for persistence. This means that I do not persist through loss of the capacity for consciousness (self-consciousness, specifically), through sudden organizational change, or loss of the physical substratum altogether. Either destruction of the relevant parts of the brain or full-on cerebral death would both kill me. EP accepts this result, too. But, on SR, reprogramming *also* kills me, and this is true even if the other continuities remain stable.

Another point of divergence concerns the possibility of successive minds. Both views accept that successive minds could serially occupy a single brain, but they differ in *how* they explain that this might happen. Ignoring the case of dissociative identity disorder, which is still not scientifically well understood,[32] SR accepts successive minds in a wider variety of possible circumstances than EP. McMahan posits just two ways one brain might support successive

[32] Though see Wilkes (1981), Mackie (1985), Radden (1996), and Gunnarsson (2009) for philosophical treatments of dissociative identity or multiple personality disorder.

minds. A first way would be if the relevant regions of the brain with the capacity to generate consciousness were removed, then replaced with new tissue. Assuming the region removed is comparatively small, it seems reasonable to consider this the same brain with a 'new' mind borne out of the new tissue. We can break down exactly how this might work to make things clearer. At removal, the capacity for consciousness is lost, so the first mind goes out of existence. When new tissue is grafted to the brain, the capacity to generate consciousness returns. But because this capacity was not continuously held between removal and regrafting, another mind now comes into existence. The other way one brain might support successive minds depends on whether or not consciousness is or could be linked to some specific brain regions. If it is not, it is hypothetically possible that if the relevant region is destroyed or severely damaged beyond functional continuity, another region might be altered in such a way as to support the capacity for consciousness. This would again be a 'new' consciousness—and thus a new mind—because physical and/or functional continuity was disrupted in the case of the first mind.

SR accepts both of McMahan's possibilities, as loss of either physical or functional continuity is on SR enough to kill one of us. However, SR also regards organizational continuity as necessary for persistence (and, along with physical and functional continuity, jointly sufficient for it). So, SR must also accept the possibility of successive minds in a third way: through any sudden reorganization of the relevant brain states to produce a discontinuous arrangement of mental states. For instance, severe head trauma might bring about substantial reorganization of brain states. Though I will more cleanly distinguish between sudden and gradual change to organizational continuity in the next chapter, it should already be clear that some reorganization is not persistence preserving whether it is sudden *or* gradual. As noted in *Gradual Brain Washing*, reprogramming is the wrong kind of cause to be persistence preserving—even if this is done gradually. Of course, EP would not recognize any amount of organizational discontinuity as persistence severing.

Given these differences, what makes SR a better fit for my project than EP? For one, to say that we are ‘embodied parts’ or minds could be argued to fail to capture what is uniquely special about us as opposed to other conscious entities. Mere consciousness is not enough—that is, we seem to be different from other conscious entities not simply in degree, but in kind of thing altogether. SR attempts to address this by requiring the capacity for self-consciousness at the level of the realizer. So, it might be said that EP focuses on the wrong capacity in highlighting bare consciousness only.

Even if this is wrong, though, EP fails to meet the needs of the current project because it cannot support a moral persistence condition. Since organizational continuity—and thus the particular mental states that depend on this continuity—is on EP immaterial to what we are, moral character traits can actually have no bearing on our persistence on this view. I could on EP survive complete reprogramming (at least, as McMahan understands it). This would, of course, include sudden change in my moral character traits. So, EP cannot sustain a moral persistence condition, and EP is therefore incompatible with the goals of my project. Even if I am wrong and EP *does* capture what is uniquely special about us, we should for present purposes reject EP.

3.3.2 Brains

Another view that might seem similar to SR is the brain view (B): the notion that we are brains in a strict and literal sense. That is, we are the small, fleshy organs we ordinarily think of as being in ‘our’ heads. Olson (2007) suggests that there are few actual proponents of this view, as most (Puccetti 1973, Tye 2003) collapse into other views like constitutionalism or something like McMahan's EP, though Olson does not consider the latter. Nagel (1986) suggests what he calls the 'hypothesis that I am my brain' as

> I am whatever persisting individual in the objective order underlies the subjective continuities of that mental life I call mine. But a type of objective identity can settle questions about the identity of the self only if that thing in question is both the bearer of mental states and the cause of their continuity when there is continuity. If my brain meets these conditions then the core of the self—what is essential to my existence—is my functioning brain (40).

Yet, over the next couple of pages, Nagel doubles back, saying that "I do not know in any detail what is responsible for [what underlies those subjective continuities]" (41) and that "it seems possible...that I may really be any of a variety of things (soul, brain, etc.)..." (42). Hence, Nagel ultimately claims to be agnostic about personal ontology, though he *does* give us a skeletal idea of how B might work.

Whether or not B is adopted by many thinkers, there is much intuitive appeal to the account. After all, the prevailing folk view seems to be that brain death results in the death of one of us, and the American Medical Association in 1980 approved the *Uniform Determination of Death Act*. This act incorporates brain death into the medical definition of death:

> An individual who has sustained either (1) irreversible cessation of circulatory and respiratory functions, or (2) irreversible cessation of all functions of the entire brain, including the brain stem, is dead. A determination of death must be made in accordance with accepted medical standards (UDDA 1980, 5).

So, there is a widely held conviction that continuous brain function is necessary for one of us to survive. On B, this would be easily explained. Destruction of a brain or brain death would be a case of actually killing one of us, not just the loss of an organ integral for our survival. In terms of parsimony, B is fairly straightforward. Other things being equal, this is always a desirable characteristic of our theories. We think, and the brain obviously seems to be the organ of thought. So, it looks to be an equally obvious candidate for what we are. Indeed, it might be said that the onus is on the opponent to explain why *I* think but am not my brain. B is therefore intuitively appealing and refreshingly uncomplicated as far as theories of personal ontology go.

For these virtues, though, B is beset by many problems. Most troublingly, Olson (2007) argues that B inevitably collapses into a form of homuncularism wherein the various systems that see, hear, and cognize must *each* be a distinct entity. This is because Olson insists that the proponent of B is committed to what he calls 'thinking subject minimalism'. Olson attributes the advocate's belief that the brain is the thinking subject to the notion that something must be directly involved in thinking for it to be a proper part of the thinker. Consequently, the advocate might say that nothing outside the brain contributes to thinking in any meaningful way. Setting aside the many other troubling implications this might have, like the vagueness of the idea of 'direct involvement' (is the heart directly involved in thinking, as it keeps the brain alive?), Olson observes that there seems to be little principled reason, if thinking subject minimalism is true, to stop with just general, thinking subjects. For instance, why not also accept a *specific thinking subject minimalism*? That is, what keeps all and only the parts of the visual system from composing a separate subject, or even all and only the parts involved in thinking about philosophy as opposed to, say, the culinary arts? On such a view, it is next to impossible to identify which of the many system-subjects *we* would be.[33]

While I'm not sure that I agree with Olson that the only philosophical motivator—or at least the most likely one—behind B is thinking subject minimalism, I can point out that SR is not susceptible to this problem. As I argued earlier, given SR's focus on mental states and general realizers, it is compatible with views that what is involved in or contributes meaningfully to

[33] Olson *does* propose a solution to homuncularism via the *psychological individuation principle*. This is the notion that we are individuated by unified sets of causally-related psychological states. However, Olson immediately thereafter rejects this view, saying it is altogether incompatible with *any* version of B.

thinking need not be confined to the physical brain. As such, SR can sidestep B's alleged claim that, due to thinking subject minimalism, brains are the boundaries of the thinking subject.

That SR is not vulnerable to homuncularism is certainly one reason to prefer it to B, but it is not the only reason. Like EP, B—in neglecting to emphasize self-consciousness or even mental states at all—fails to capture what is distinctive about us. Beyond this, B is, again like EP, incompatible with a moral persistence condition. Like SR, B implies that we are parts of organisms—specifically, the conscious, thinking part. But to say that we are literally brains is quite different than saying that we are mental states instantiated in a constant realizer that, in our case, happens to be the brain. If we are literally brains, all that is necessary and sufficient for our persistence is the persistence of the brain. No kind of psychological or even functional continuity has any place here—I could survive everything from reprogramming to a brain zap to full-on cerebral death. All that seems to matter in B is physical continuity of the brain. Of course, if no psychological states have bearing on my survival, then no moral character trait could, either. Hence, we ought to reject B in favor of SR for the present project.

3.3.3 Material Constitution

A third competitor in terms of what we are is the constitution view (C). This is the view defended by Shoemaker (see, for just a few examples, his 1999 and 2011) and Lynne Rudder Baker (2000, 2002). Though the view can take many forms, the most common is something like the following: we are entities materially constituted by human animals. We relate to our animals in the way a statue relates to the material—say, gold—it is made of. The statue and the gold seem to be distinct things. I can melt the statue, for instance, without getting rid of the gold. So it is with us and our animals. Taking things a step further, constitutionalists argue that the two things have different properties while they coincide. For one, the statue lacks the property 'able to survive melting'. And *we* have properties that our animals lack—namely, psychological properties. Animals think only derivatively: they think by constituting things that *do* think for themselves and have psychological features (things like us). This means that we share all of the physical properties of our animal, but also have our psychological properties essentially. Like SR, on C we are partly psychological and partly physical.

Yet, in light of the odd nature of the constitution relation, C faces unique challenges to its coherence. Olson (2007) suggests that most constitutionalists will want to deny that animals think (except derivatively, perhaps) in order to avoid the terrible consequence that there would otherwise be two thinking entities that overlap in the same body: the person (one of us) and the animal. But, since both would think in exactly the same way—both would read and contemplate this passage, for instance—there would be no real way of knowing which of the two *you* were. Worse, it seems at least unprincipled and at worst vile to say that, though both person and animal are mentally identical, only *one* of the two entities is properly a person. So, constitutionalists must deny that animals can think.

However, this leads to a medley of further problems. First is the apparent absurdity in accepting that, though you and your animal are physically qualitatively identical, *you* have a suite of mental properties and your animal doesn't. We can thus ask, along with Olson, "How can putting the same parts together in the same way in the same circumstances give you qualitatively different wholes?" (63). This is the indiscernibility problem. Two objects that are apparently indiscernible have wildly different mental properties and resultant persistence conditions.[34]

[34] Advocates like Shoemaker argue that *you* have psychological persistence conditions, but your animal merely has biological ones—presumably having to do with functional organization, metabolizing external material, etc.

Then, there is the related question of when and which properties are constitution-inducing. An organism presumably acquires some physical property that then allows it to constitute an entity that shares this physical property but which also manifests some mental property causally related to the constitution-inducing one acquired by the organism. But which properties *are* constitution-inducing? Olson argues that neither of the answers available to constitutionalists are especially satisfying. On the one hand, constitutionalists could adopt the generous view (per Bennett 2004). On this view, *all* properties are constitution-inducing. This is akin to the universalist answer to the special composition question[35]—the question of when some plurality of things composes another thing—which states broadly that any two things comprise an object (like my left hand and the sunken mast of Sir Francis Drake's flagship). For C, the generous view of when and what is constitution-inducing has the baffling implication that you coincide with a constantly revolving and absurdly large cast of entities. In Olson's (2007) words, this means, for example that

> [w]hen you stand and frown at once, you coincide with a being that essentially stands and contingently frowns, with a second being that contingently stands and essentially frowns, and with a third being that essentially stands and essentially frowns (67-68).

Beyond the apparent ridiculousness of this idea, it also reintroduces the difficulty of fixing just which being *you* are in the case of, say, when you come to think about some subject, like turtles. Are you the being who thinks about turtles essentially and thus has a limited existence only as long as turtle-thought is going on? Or are you the being who was not thinking about turtles a few minutes ago? Psychologically speaking, *both* would think s/he were you. Between this example and Olson's standing/frowning comment, it is clear that the generous view is rife with unpalatable consequences.

[35] See Rea (1998) and Van Cleve (2008) for defenses of this claim.

Opting for a more limited view of when and what is constitution-inducing seems, on the other hand, unprincipled. Whether we say, as many advocates do, that acquiring the capacity for first-person thought demarcates the constitution-inducing, or follow Shoemaker's milder claim that *any thought at all* is constitution-inducing, the answer just seems to be a somewhat arbitrary brute fact. This is made all the worse when we recall that, per the indiscernibility problem, two physically qualitatively identical entities coincide, but one of them constitutes the other because some capacity or another was suddenly acquired. Considering the indiscernibility problem and the problem of when/what is constitution-inducing, we see that C is a difficult view to accept.

Despite this, C is not without positive points. To its advantage, C is well-equipped to handle a moral persistence condition, given its emphasis on the psychological. In fact, there might even be a way to incorporate a moral persistence condition more strongly. That is, it could be possible to construct a version of C which involves a moral persistence condition as in some way specially constitutive among other psychological features. But when we go on to ask whether we should prefer SR or C as a personal ontology that might undergird a moral persistence condition, we see that the constitution relation raises many difficult questions for advocates of C that might just seem insurmountable. Simply put, SR avoids issues of indiscernibility and when/what is constitution-inducing because it steers clear of the conceptually problematic constitution relation. Hence, though C may appear to meet the needs of the current project well, the metaphysically murky notion of its central relation recommends against it and in favor of SR.

3.3.4 Hylomorphism

We turn now to a final competitor to SR in the hylomorphic view (H). Historically, Aristotle and Aquinas both famously advanced hylomorphic theories, and there has been a present resurgence in the field in work done by Kit Fine (1999), Mark Johnston (2006), David Oderberg (2007), Kathrin Koslicki (2008), Michael Rea (2011), Anna Marmodoro (2012), Robert Koons (2014), and William Jaworski (2016). It will be difficult to provide a satisfactory general account of H, since these views offer often competing accounts of just what H amounts to. Nonetheless, this subsection will focus principally on Jaworski's (2016) recent version when I cannot comment about H more generally. This is primarily because Jaworski's theory might appear especially close to SR.

To give a rough picture of H, we can say that H generally accepts that we are particular matter arranged according to a certain form or structure. It is this structure that makes the material what it is. Jaworski puts it like this:

> Structure is a basic ontological principle: it concerns what things are. It is also a basic explanatory principle: it concerns what things can do (9)...organization, order, structure, or arrangement is thus something real, with real explanatory significance (11).

Jaworski calls this 'structural realism'. What this means is that, to the hylomorphist, we are human organisms. But being a human organism on H amounts to matter that is structured in a particular way such that our characteristic capacities emerge.

On the face of it, H is an attractive answer to the question of what we are. It clearly captures what is distinctive about us, as the capacity for self-consciousness can easily be considered a component of the necessary structure something must have if it is to be one of us. H also appears perfectly compatible with a moral persistence condition. That is, if some material must have the right kind of structure to be one of us, it seems coherent to think that a certain moral configuration could be built into my structure—assuming structure is particular to individuals and not universal. Particularity is important here because it would be pure nonsense to claim that there is a universal structure for us that includes a certain moral configuration that we all share in order that something qualifies as one of us. Some hylomorphists (Johnston 2006, Jaworski 2016) even ascribe a dynamic character to structure. This could allow for some degree of change to moral character traits without severing persistence, just as Parfit's Narrow view—upon which SR is based—allows.

H runs parallel to SR in other ways, too. On SR, we are a sort of system composed of mental states and their realizers. In section 3.2, I referred to this relation as that between a pattern and the instantiation of that pattern. These comments seem just as appropriate to H in the way structure relates to matter. And the ontological relation between structure and matter on H is quite similar to the relation between the totality of mental states and their physical realizer on SR—at least insofar as SR takes the realizer to be important as a way of instantiating mental states. For H, what is most important in making something one of us is that it have the right structure. The matter is only valued as a way of instantiating that structure. Given these similarities and the apparent close fit between H and my goals with the moral persistence condition, it is all the more vital that I bring out the relevant differences between SR and H below.

One problem it is incumbent on all hylomorphists to explain is the nature of form or structure. If structure is supposed to be real in any meaningful way, what is its ontological status? Jaworski's complicated characterization of structure treats it as a property. Due to other metaphysical commitments he makes, 'property' is further cashed out as a power. Specifically, structure is a power "to configure (or organize, order, or arrange)" (2016, 94). Moreover, Jaworski states that structure properties are particulars and not universals—my human organism structure is not interchangeable with yours. Hence, at least on Jaworski's version of things, structures-as-properties could be said to sound about as ontologically real as mental states on SR. Given his particularity stipulation, H still looks to be nicely compatible with a moral persistence condition.

However, even if H can avoid the less than clear ontological status of structure, it reveals itself to be vastly different from SR in many ways. In SR, it turns out that we are *parts* of something else—an organism. This is not the case for H. On H, we are matter arranged in a particular way—as a *whole* organism. Of course, this is a significant difference in what we are: a part of a human organism on the one hand and a full-blown human organism on the other. But even apparent similarities between SR and H break down under closer scrutiny. Although what we are on both views could accurately be said to be jointly psychological/physical entities, the relation between the members is not the same on the two views. On H, the structure determines the properties and features its matter will take. In other words, the higher level (structure) determines the features of the lower level (matter). This is a reversal of the relation in SR. Mental states do not determine anything about the realizer's properties; to the contrary, it is the realizer that determines the states that are realized. That is, the lower level determines the features of the higher level in SR.

Continuing with this line of thought, both the state-realizer entity of SR and the structure-matter organism of H can be described as systems. Granting this, the two systems operate in fundamentally different ways. The bottom-up determination means that the SR system is constrained at least partially by its members. So, since SR is compatible with extended mind theories—as I argued in section 3.2—and assuming some version of extended mind is true, the

SR system can grow or shrink as it gains or loses members. That is, the SR gets a unique character from the specific members that constitute it and work together in a complementary way. An SR system that thinks partially through external artifacts is a very different system from one with only biological parts. While there is no guarantee that extended mind is true, SR remains open to the possibility.

On the top-down determination of H, though, the organism-as-system's members are constrained by its structure. That is, what counts as part of the organism-as-system will be limited by the nature of the structure involved; what's more, this is a one-way relation. We would not expect the structure to change to accommodate new membership, as it is the structure that decides what is part of the system in the first place. Again, H and SR are operationally inverted here. Beyond this, there is a greater consequence: H's system is at its core more restrictive due to its top-down determination. What this means is that, unlike SR, H limits what counts as part of one of us to the biological. Jaworski calls this the 'biofunctional account of parthood', saying that any part of one of us consists "either in composing cells or being composed of cells" (115).

If the parts of one of us must be biological on H, this implies another break from SR: H seems incompatible with multiple realizibility.[36] There would be no way on this view for our characteristic mental features to be realized in, say, synthetic material as the relevant structure dictates that these properties manifest only in the biological. Jaworski suggests that his version of H can accept an appropriately qualified multiple realizibility as a kind of analogy applying the same predicates to similar observed behavior between different entities. But this linguistic multiple realizibility surely fails to get at the metaphysical salience of the traditional claim. H,

[36] Indeed, it is incompatible on separate grounds from the biofunctional account of parthood. Jaworski—again due to other metaphysical commitments—denies the existence of higher order properties. Traditionally, multiple realizibility gets its teeth from the fact that higher order properties can be instantiated in different substances.

therefore, still seems incompatible with any kind of meaningful multiple realizibility. SR at least accepts a weakened form of multiple realizibility similar to Clark's sense that we have a 'cyborg' nature that could in principle extend beyond the skin and skull.

Perhaps the reader thinks that the noted differences between SR and H are not as important as I make them out to be. Yet, I think there is still a very good reason to prefer SR over H: SR implies fewer ontological commitments, especially in terms of mereology and composition. Hylomorphic accounts are mired in attempts to ground what we are in a broader metaphysic. This is because H itself is part of a bigger metaphysical enterprise than just personal ontology. To see this, we only need to ask a few questions. Are things other than organisms made what they are by structure? For instance, does a table have ontologically and explanatorily 'real' structure, as Jaworski puts it? What about the object (if there is one) consisting of my right elbow and Bentham's preserved head? If so, then the hylomorphist will need a more robust and principled account of just how many objects there are. That is, the hylomorphist will need to develop an answer to the special composition question. If not, then the hylomorphist must say why not—this also requires an answer to the special composition question. Likewise, just what it means for someone like Jaworski to say that structure is a property presupposes a treatment of properties—whether they are sparse or abundant, for instance. In sum, H is only one aspect of an overall project to carve up reality according to an entire metaphysical system. Alternatively, SR remains reticent about most other problems of metaphysics and thus might be compatible with multiple broader metaphysical systems. For all of H's benefits, I argue we ought to prefer SR for its lack of metaphysical baggage.

3.4 Beginnings and Endings on the State-Realizer View

In the preceding sections, I argued for SR as an account of what we are. On SR, we are sets of mental states instantiated in a constant realizer, and the capacity for self-conscious mental states in particular is a necessary condition for something to be one of us. Then, I suggested reasons why we ought to prefer SR as a theory of personal ontology for the current project. Unlike EP and B, SR is fully compatible with a moral persistence condition. Moreover, and again opposed to EP and B, SR is also able to get at what is uniquely special about us (our capacity for self-consciousness) as a necessary feature of being one of us—something any theory of personal ontology ought to do. SR does not suffer from B's alleged homuncularism because it does not endorse thinking subject minimalism. Similarly, SR avoids the indiscernibility and constitution-inducing problems of C by not invoking the constitution relation. Lastly, accepting SR does not commit one to an entire, complicated metaphysical system as does accepting H; SR is a less restrictive personal ontology because it is *only* a personal ontology. As a result, I concluded that SR is therefore preferable to similar competitors as a theory of what we are.

In the next chapter, I will advance both a theory of persistence and a moral persistence condition embedded in this theory. I feel that I am finally in a position to do this because this persistence theory will be built on the bedrock of the SR personal ontology argued for in the present chapter. To make this clear, I will in this last section explore a few implications of SR in terms of when we come into and go out of existence.

SR admits that, like the title of Parfit's 2012 paper, we are not human beings. Instead, we are the thinking parts of human beings. Even still, we are not just any thinking parts, for we are the parts that allow thought especially characteristic to us—self-conscious thoughts. In terms of beginnings and endings, this means we go into and out of existence when the capacity for self-consciousness emerges or is extinguished. We thus come into existence some time after conception and even some time after the capacity for consciousness is formed. Consequently, we were never fetuses, nor were we even small children, as they likely lack the capacity for self-consciousness. Indeed, SR would seem to imply that the being that is first conscious in a human animal is a different being altogether from either the animal or the later one of us that emerges with self-consciousness. If, as I have argued, we are the thinking part of an animal, then the animal thinks only derivatively. But we are specifically the *self-conscious* thinking part of an animal. If the animal thinks only derivatively, but there is a conscious, thinking entity pre-self-consciousness, then there is *another* thinking part that comes into existence at consciousness and goes out of existence when the capacity for self-consciousness develops.

Maybe it is—and this is probably generous—odd to say that we come into existence only when the capacity for self-consciousness is realized, in the same way a house comes into existence when the bricks and mortar are arranged in a certain way. If we are mental states with constant realizers as SR would claim—or patterns instantiated in a particular realizer—then there is nothing objectionable in saying that we go out of existence when the pattern does. Of course, we are not *just* the pattern. We could not survive Parfit's Teletransporter from Chapter 1, as the pattern but not the realizer is preserved in creating a Replica through Teletransportation.

But even if one is willing to accept the foregoing, it is additionally strange to suggest that there are not one but at least *two* thinking entities that pop into existence between conception and death of the animal. Yet, I do not think this is as *ad hoc* a solution as it sounds. This is perfectly consistent with the ontology of self-conscious entities described by SR. Actually, the conscious entity that exists pre-self-consciousness is a natural consequence of that view: the same rules apply all the way down. On my view, consciousness creates a new entity—one that does the thinking; it is a part of the animal, not the animal itself. It seems to me that, as Locke did by distinguishing between something with the functional organization for life (an animal) and something that is a person, we can also distinguish between something with that functional organization (again, the animal) and the phenomenological entity (plus realizer) that *feels* or thinks. If it is right to say that we are realized patterns, positing a pre-self-conscious 'pattern' that is different from one of us is at least coherent—peculiar though it may sound.

This might even be less onerous than it seems. That something goes out of existence if its 'pattern' does is rather common, on reflection. It even affects organisms. An organism goes out of existence when it dies despite the fact that there is clearly a body still hanging about—that is, the same physical parts are still there. All that has changed is the organization of those parts; in other words, the specific *pattern* has changed from a pattern that supported life to one that does not.

Setting our discomfort with short-lived, pre-self-conscious beings aside, SR at minimum also implies that one of us *goes out* of existence at least when the capacity for self-consciousness is lost. Destruction of the brain certainly qualifies, or even just cerebral death. But, as I argued in 3.3.2, none of this is really out of the ordinary. Cerebral death is intuitively considered to be plain-old death by many. After all, it seems common for so many of those with family members in a persistent vegetative state to say something like "she died in the accident" when deciding to withdraw life support from the vegetative animal in the hospital bed. Such individuals probably do not see themselves as killing one of us when they say this, just an animal with extraordinarily poor quality of life and next to no interests in continuing to live.

Perhaps more troubling is the implication that degenerative brain diseases like dementia or Alzheimer's disease would kill one of us, even though there still seems to be a conscious being present. Like the pre-self-conscious child, this would be a new pattern in the same realizer and thus a new being altogether. Again, this might strike some of us as dubious, but I argue that it remains perfectly consistent with the rules laid out in SR—again, the same rules apply all the way down. What's more, this is not a unique problem for SR. EP accounts that propose that we begin and end with consciousness full-stop still have to deal with what McMahan calls pre-conscious and post-conscious entities. That is, there seem to be beings that have disunified, proto-conscious states but that do not constitute full-on consciousness. This occurs before the emergence of consciousness in a human animal, but also in the most severe brain degeneration cases. There is a point where some sufferers' conscious experience becomes so scattered and disunified that it hardly seems like a genuine stream of consciousness at all. So, EP just moves the problem I describe in SR down a level—but, importantly, it is still there.[37]

The reader might pause at my regular use of 'change in pattern' throughout this section. "Does this mean, for instance, that *we* are strictly a single pattern of mental states and that any change thus kills one of us? This sounds an awful lot like the Ephemeral Selves problem of bundle theories, and I thought we had dispensed with those?" In part, I can dismiss this worry by reminding the reader that SR is grounded in Parfit's Narrow view. This at least limits some of the cases in which change in the pattern of mental states would kills one of us, because the Narrow view requires that our mental states be caused in the right way—namely, by our realizer. So, obvious cases like reprogramming would kill one of us even though our realizer remains physically and functionally continuous because the cause of our changed mental states is not the normal cause. The Narrow view also suggests that, through overlapping chains of continuity, we

[37] It looks like this is a problem for all theories of consciousness except maybe higher-order theories, as those theories seem to fold consciousness and self-consciousness together. However, higher-order theories of consciousness require stiffer commitments to what consciousness consists in than I am willing to settle-on here. So, while higher-order theories might solve this problem, the price to be paid is too high for me right now.

can persist through gradual changes in mental states over time like the changes that evidently happen naturally through age and experience. But my full response is, for now, likely to be unsatisfying, as I cannot fully answer until the next chapter. The goal of Chapter 4 in proposing a theory of persistence out of SR will be to show just how much change (and why) the ‘pattern’ of mental states can tolerate, including how changes in moral character traits affect the general ‘Narrow view’ model of persistence I adopted as a framework here in Chapter 3.

CHAPTER 4

THE NARROW MORAL-PSYCHOLOGICAL VIEW OF PERSISTENCE

This final full chapter moves from investigating and providing the grounds for the possibility of a moral persistence condition to actually postulating an account of persistence that includes a moral persistence condition as a *necessary condition*. Although I argue for only one way moral traits might be a persistence condition, there are most assuredly others. The last chapter laid out the ontological conditions under which a moral persistence condition would be possible. Accepting these conditions—what I called the State-Realizer View (SR)—we now can see how a moral persistence condition fits into a persistence account derived from those ontological conditions.

There are three major sections to this chapter. In the first, I give a detailed account of Parfit's Narrow Psychological View (NSV), as this view will be the starting point for my own persistence theory. Then, in a second section, I present my theory as a modification of Parfit's, what I call the Narrow Moral-Psychological View (NMPV). As a departure from Parfit's view, NMPV states both (1) that a moral persistence condition is a *necessary* condition for persistence and (2) that this condition consists in a relation called moral continuity. In a final section, I consider several problem cases that illustrate my view's explanatory power and help to better distinguish it from Parfit's.

A quick note before continuing: my view takes moral character traits to be psychological features generally. I do not think that the argument for NMPV requires a particular understanding of moral traits beyond this; that is, I do not think it matters to the argument whether moral traits have satisfaction conditions—like beliefs or desires—or if moral traits are instead more like dispositions. Indeed, I believe that it is a virtue of my theory that it is open to and compatible with a range of ways we might characterize moral traits as mental features. If moral traits are *in fact* of a different character than other mental states like beliefs and desires, then this might make a necessary moral persistence condition even more plausible. For, if moral traits are psychologically unique, it would make sense that moral traits play a constitutive role in persistence that is distinct from that played by other mental features. This seems to me to be an empirical question, though. All the same, I think that it is to the credit of the NMPV if it can establish a necessary moral persistence condition without postulating a unique mental character for moral traits, and this is what I attempt to do here.

4.1 A Starting Point: Parfit's Narrow Psychological View of Persistence

Given that SR uses Parfit's NPV as an ontological foundation, it ought to be unsurprising that the persistence account I propose in this chapter would be a modification of Parfit's view. I should stress that this is not circular: the NPV is ultimately a persistence theory, not a theory of personal ontology. So, to adopt it as a foundation for SR—as I did in the previous chapter—means only that I needed there to tease apart what kind of personal ontology an NPV-like view might presuppose. However, it is the business of the present chapter to now turn SR into a bona fide persistence theory. Although I modify Parfit's view as a means of accommodating a moral persistence condition, I do not think the possibility of a moral persistence condition turns specifically on the NPV. It seems entirely plausible to me that there might be other accounts that could accept the ontological demands of SR (or a similar view) but with a persistence theory distinct from Parfit's. In the end, what matters for the coherent possibility of the version of a moral persistence condition I argue for in this chapter is that what we are consists both in something psychological and in whatever physical thing realizes that psychology. Nonetheless, I think that Parfit gives us a ready example of a persistence theory that meets SR's ontological needs. As such, I spend the rest of this section revisiting the NPV and its unique implications for persistence. Much of this was already given in broad strokes in Chapter 1, though I will here be giving it the richer detail called for by my modification of the view.

In many ways, the NPV is a stopgap between bare physiological and psychological persistence 'criteria'—a word Parfit takes to mean "what [persistence] necessarily involves or consists in" (1984, 206). I say this because the NPV grounds persistence in both the psychological and the physiological. Again, though, the physical is important to this view only insofar as it is the constant realizing condition for the existence of the psychological. With that in mind, I will restate only Parfit's Psychological Criterion here, as the full Physiological Criterion does not really come to bear on the NPV.

Broadly, Parfit's Psychological Criterion depends on two relations that hold between mental states. First is psychological connectedness: 'direct' connections between mental states, like the case where a six-year-old forms a memory of first tasting candy corn and a thirty-year-old later remembers that dreadful experience. Similarly, a young philosophy student might share her intention to publish on Kierkegaard with the later professor she becomes. (These relations are not yet part of any persistence theory, so there is little worry in giving illustrative examples that assume persistence—that the young student and the later professor are one and the same person, for instance.)

The second relation, psychological continuity, is actually more important to Parfit's Psychological Criterion and is defined as overlapping chains of 'strong' psychological connectedness. Remember that 'strong' here is somewhat vague by design, as connectedness is not a binary relation but instead admits of degrees (unlike identity, which one either has or does not have). Of 'strong' connectedness, Parfit says:

> there must be over every day *enough* direct connectedness...we can claim that there is enough connectedness if the number of connections over any day is *at least half* the number of direct connections that hold over every day, in the lives of nearly every actual person (206).

So long as there are 'enough' direct connections, there can be psychological continuity through significant psychological disunity. Indeed, such disunity is common during ordinary aging. Many of us are not strongly connected with ourselves at the age of three when we are now eighty. Yet, there is a chain of strong connections at each given stage in the life of that individual that leads from, say, the three-year-old to the eighty-year-old. Once incorporated into the Psychological Criterion, this 'strong' connectedness overlap is what will allow a subject to persist despite often substantial psychological variation. Thus, continuity does not make overly demanding claims that we *never* change psychologically, just that we do so gradually—so that the chain of direct connections is not actually severed.

This leads to Parfit's proper statement of the Psychological Criterion:

> *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) [Persistence] just consists in the holding of facts like (2) to (4) (207).

(4) is a non-branching clause meant to insulate against fission cases like Chisholm's amoeba-man (see Chapter 1). And (5) indicates that Parfit's theory is reductionist. I say more about reductionism below. For now, we will focus on (3), which differentiates the NPV from two other possibilities that Parfit countenances: the Wide and Widest views. On the NPV, the 'right'

kind of cause is normal functioning of the brain. Now SR, my ontological basis for discussing Parfit's view and my own persistence theory, focuses on the realization relation that holds between mental states and their physical substrates, and this is not strictly a causal relation. Despite this, I think that we can still talk of the causal processes behind continuity, as Parfit does. For instance, 'normal functioning of the brain' will include the realization of such-and-such mental states through a corresponding arrangement in the physical substrate.

By contrast, the Wide view goes further and asserts that any *reliable* cause of psychological continuity is the 'right' cause, and the Widest view accepts *any* cause whatsoever as 'right'. The notion of 'any cause' is probably self-explanatory, but the only restriction for the Wide view's reliable cause is that it is regularly reproducible. In this way, the cause of psychological continuity on both the Wide and Widset views can possibly be 'abnormal' in the sense that it can bypass ordinary physiological brain functioning or environmental processes of which brain functioning is a part (like learning) in order to achieve continuity. For instance, last chapter's *Gradual Brainwashing* incorporates a reliable yet abnormal cause. The mad doctor's hypnosis is reproducible week after week, but continuity obtains through the direct manipulation of the subject's mental states—sidestepping the normal physiological and environmental processes of continuity. The reason to prefer the NPV for my project is that the Wide/Widest views have unsavory implications about persistence.

Consider a similar example to Parfit's well-known Teletransporter—but even more implausible, so as to better illustrate the intuitive problems of the Wide/Widest views. Unlike teletransportation, there is not in the case to follow even the pretense of moving one and the same subject from place to place, which could plausibly color the reader's impressions.

> *Murder Clones*: Scientists develop a process of cloning that succeeds without fail. This process is bizarre in that it must genetically 'map' the subject before the

> clone is created. Unfortunately, mapping has the side-effect of utterly destroying the original subject completely before the clone is made out of entirely new material. As the clone is a physically and psychologically identical copy, it has exactly the same physical and psychological features as the original and even thinks itself to *be* the original. In fact, that clone's last memory is of the mapping machine whirring to life.

On both the Wide and Widest views, the original subject persists despite being physically destroyed just because the clone is psychologically continuous with her. The Wide view implies persistence because the cloning process is—*ex hypothesi*—reliable. And, if the Wide view implies persistence, so does the Widest view *a fortiori*. A reliable cause is, of course, also a cause *simpliciter*. That is, since the Widest view accepts psychological continuity by any cause whatsoever, it will obviously accept continuity by way of a reliable cause.

But this is surely the wrong result. Unless we want to say, along with Parfit, that identity (as persistence) is not what matters for our practical concerns like attributing responsibility and blame, it cannot be the case that I survive complete destruction of *all* of my parts just in case a timely clone is created. This would mean that we are only really psychological entities on the Wide/Widest views. To see why, recall that I claimed in the last chapter that SR takes us to be analogous to instantiated patterns. But on the Wide/Widest views we would be *only* the patterns because any particular instantiation is immaterial to persistence. The Wide/Widest views grant that our patterns persist so long as psychological continuity obtains through at least a reliable cause. As *Murder Clones* shows us, it is possible to have a reliable cause that brings about continuity through two *different* instantiations. (Technically, *three* instantiations, as the pattern is briefly instantiated in the cloning machine before it is reproduced again in the Clone; this third instantiation pushes plausibility even further, as it is most probably an instantiation of just ones and zeros.) Since no one thing can have multiple beginnings of existence, we know that the same pattern cannot cease to exist when one instantiation is destroyed and then begin to exist again when the other instantiation is later created. So, if there is continuity just because the pattern is reinstantiated somewhere else—despite complete destruction of the original physical instantiation—what it means to persist as the *same* pattern is to be a sort of abstract idea.

Hence, on the Wide/Widest views we are only really abstract patterns, and only abstract psychological entities. Our particular instantiations do not matter in determining what we are. Indeed, imagine if instead of a digital cloning machine that the pattern of a cloning subject was written down onto a paper notepad. The operators of this incredibly inefficient cloning apparatus would then manually enter the handwritten genetic data into yet another machine that would then create the Clone. Once again, this process is *ex hypothesi* reliable, and causal connections can be traced from the Original to the Clone. That is, it is just because the Original's genetic data is such-and-such that the particular information is written on the page, and it is just because of this information that the Clone has the mental states she does. So, there is psychological continuity here on both the Wide and Widest views. What is more, the subject *persists* through the entire process—briefly as words scribbled on a page! As with the Humean bundle views I considered in Chapter 3, it is unclear what we are if we are this kind of abstract pattern that can apparently exist through even an em-papered instantiation. I called this the Uncertain Existence problem back then, and it rears its head just as strongly on the Wide/Widest views here. *Pace* Parfit, the Wide and Widest views should be rejected in favor of the far more plausible NPV.

As remarked above, (5) implies that the NPV is a reductionist theory of persistence.[38] We will remember that this means that persistence is not some further fact beyond certain physical and psychological conditions holding true. If there is non-branching psychological continuity by the normal cause (the brain), then one of us persists. Unfortunately, reductionism

[38] Note that this is distinct from reductive physicalism, a theory which, while tough to peg down exactly (Jaworski 2016), is about what kinds of stuff make up reality.

is not always clear cut about who persists and when. As Parfit argues, there seem to be conceivable cases on reductionist accounts where it is indeterminate whether I persist or not. This is most telling through the example of the *Combined Spectrum*. Parfit asks us to consider that some mad doctor is gradually converting him—both physically and psychologically—into the late actress Greta Garbo. This is done incrementally. That is, each of Parfit's cells is exchanged for one matching Garbo's one at a time; at the same time, replacing these cells also causes gradual change of Parfit's mental states over to Garbo's. Obviously, there seem to be uncontroversial cases of persistence at each of the far ends of the *Combined Spectrum*. Having a single cell/state replaced is hardly persistence severing. Likewise, it sounds plainly wrong to say that Parfit persists if there is only *one* of his cells/states left and the rest are uncontrovertibly Garbo's.

The difficulty is the cases in the middle. The admittedly imprecise notion of strong connectedness cannot really help us here, either. Strong connectedness obtains when there are at least half as many connections that hold day-to-day. But the question in such cases is 'when does something stop being psychologically continuous with me?' In the Garbo case, we essentially have two persons involved: one, Parfit, is gradually going out of existence; the other, Garbo, is gradually coming into existence. Both have determinate cells/states that individuate them. So, the question in the Garbo case goes further than the one in mere psychological continuity. We here want to know 'is there a clear boundary at which something stops being me and becomes someone else?'

If we try to apply strong connectedness here, it might suggest to us that 51% of cells/states is the demarcation line of persistence. That is, something is me just in case it has 51% of my cells/states. (Clearly, this line could not be 50%, as the subject's composition would, at a certain point in the *Combined Spectrum*, be at the same time 50% Parfit and 50% Garbo. This would make the single subject simultaneously Parfit *and* Garbo. But persistence identity is a 1:1 relation—one thing cannot be two.) Yet, 51% sounds arbitrary and a bit unprincipled. It would mean, among other things, that a single cell/state could make the difference between something being me or someone else. Such a difference seems trivial, yet it is supposed to account for something monumentally important: whether something is me or someone else or, more worryingly, whether I live or die.

If we find this 'sharp borderline' unattractive, Parfit insists that we have only one other option. Instead of a sharp borderline, we should accept that there are cases when, despite having all of the reductionist physical and psychological facts, we just cannot tell if something is me or someone else. In other words, there are times when persistence is indeterminate. That it might sometimes be indeterminate if I persist or not is, as Parfit concedes, hard to believe. To Parfit, this is a much more palatable view than the apparently arbitrary alternative of a sharp borderline. It really does seem like the *Combined Spectrum* presents us with no other viable solutions. The thing that emerges at the end—Garbo—is neither physically nor psychologically continuous with the thing that was first strapped into the surgical table—Parfit. And, though Parfit does not directly address this, it also seems that we cannot wiggle out of the dilemma by deploying vagueness, like in the Sorites. Persistence is not something that comes in degrees; that is, something cannot be more *me* or less *me*. A heap of sand can be a vague concept, but it seems doubtful that persistence can be, too.

In the end, I am inclined to agree with Parfit that reductionism, while metaphysically uncomfortable, is more plausible than a sharp borderline. Returning to the NPV, we can now see that this view has some strange metaphysical implications. Namely, its reductionism implies that our persistence is sometimes indeterminate. However, I think that, unlike the Wide/Widest views—which share this reductionism—the NPV allows us to resist Parfit's controversial claim that persistence is not what matters for our practical concerns. He draws this conclusion in part from the fact that persistence is sometimes indeterminate, but also from the Wide/Widest views' openness about what counts as persisting. If I can persist through destruction of my entire physical structure—which seems, even to Parfit, not to be genuine persistence—then persistence cannot be what matters. If my clone is, in Parfit's words, just about 'as good as' being me, all that truly matters is psychological continuity.

But the NPV rejects this. Physical continuity of the realizer makes a difference, too. Having a clone is not 'about as good' as persisting on this view. Moreover, cases where the NPV would deem persistence to be indeterminate are extremely few and far between. In fact, indeterminacy cases seem limited to fanciful thought experiments like the *Combined Spectrum*. Standard cases of psychological change through ordinary aging, physical or mental trauma, or even profound events like sudden religious conversion can all be explained on the NPV, as none of these cases involve the bizarre kind of incremental physical/psychological replacement of the *Combined Spectrum*. I argue that the upshot is that the NPV allows us to take the orthodox view that persistence *is* what matters, as indeterminacy only affects a handful of highly bespoke cases.

4.2 Adding Moral Weight to the Narrow Psychological View

With a full description of the NPV out of the way, I am finally in a good position to actually argue for a moral persistence condition. The NPV is clearly able to sustain a moral persistence condition, as it is a principally psychological theory of persistence. Since moral character traits are psychological features, this view allows that these traits—caused through the normal functioning of the brain—can be constitutive of persistence by being a subset of our psychologically continuous states. However, the NPV cannot, as is, really account for a moral persistence condition in a way that accords with the moral persistence intuitions highlighted by Strohminger and Nichols in Chapter 2. These intuitions take moral character to be the *most* important feature in determining persistence. But, on the NPV, moral traits are just one set of psychological features among many. This means, problematically, this view is compatible with complete change in all moral character traits. Indeed, so long as 'enough' other direct connections remain, I could persist through full episodic memory amnesia, personality shift, complete change in moral traits, or any other selective change to particular groups of psychological features. In fact, this versatility is evidently one of the NPV's virtues. At the same time, this also means that a moral persistence condition is not necessary for persistence. We can quickly see that a moral persistence condition is also not sufficient for persistence on Parfit's view. If all and only non-moral psychological features were reprogrammed, wiped, or otherwise changed, the presence of just moral features would presumably not constitute enough connections for 'strong' connectedness to obtain. Ultimately, moral features on the NPV have no special place in persistence.

As such, I will in this section propose a modification to the NPV that allows a moral persistence condition to take a more robust role in persistence that better accommodates the moral persistence intuitions discussed in Chapter 2. This modification takes the form of a further requirement that a new relation, moral continuity, must hold in order for persistence to obtain. In this way, I argue that a moral persistence condition is necessary for persistence though not in itself sufficient for it.

4.2.1 Why a Moral Persistence Condition is Necessary

One way to see why a moral persistence condition might be necessary for persistence—and how the NPV misses this—is to consider how the NPV would treat the following cases:

> *Mental State Booth*: in a future society, direct manipulation of one's mental states is a simple process. Subjects can enter 'mental state booths', apparatuses that scan the user's brain and maps his mental states onto a display, whereupon the user can then selectively remove undesirable states and add desirable ones. Perhaps the subject would prefer to replace a desire for sugary foods with one for something healthier, to forget an especially traumatic memory, or to become courageous when he is, in fact, a coward.

This is an unproblematic case for Parfit's view. Only a handful of direct connections are severed in a given visit to the mental state booth. So, even though these changes occur by way of an abnormal cause—the manipulation of the booth's technology rather than ordinary brain function—there are more than enough remaining connections for there to be strong connectedness. Suppose, though, that we alter the case:

> *Malfunctioning Mental State Booth*: Everything from the description of *Mental State Booth* is the same, but a particular booth has started malfunctioning and performs the requested operation for *all* of a given mental state. So, Jones enters the booth to remove a memory, but leaves a full amnesiac (in terms of episodic memory). Next, Smith enters the booth wanting to be kind rather than inconsiderate but leaves with all of his character traits inverted. (As a matter of convenience, we can consider these traits in the manner of virtues and vices, but I do not think my argument turns on this.) For both Jones and Smith, all non-targeted types of mental state remain unaltered. Let us further suppose that at least half of all of both subjects' total mental states are unchanged.

On the NPV, both Jones and Smith persist as one and the same person through their time in the mental state booth. That is, in both cases, the person who enters is psychologically continuous with the person who leaves, and this continuity obtains through normal brain functioning. Of course, manipulation of brain states by way of a malfunctioning machine certainly counts as an unnatural cause. Nonetheless, continuity obtains because the vast majority of non-manipulated connections between mental states *do* hold due to normal brain functioning. Memories or moral character traits do not alone represent the full psychology of a person. Despite sudden discontinuity among those particular mental features, *enough* direct connections remain between the pre-booth and post-booth subjects in both cases for there to be strong connectedness between them. Again, even if the new states are abnormally caused, there is still this strong connectedness between unaltered states. Hence, both Jones and Smith persist on the NPV.

I want to suggest that there seems to be something intuitively wrong with this result. While it is plausible that the amnesiac Jones is still Jones, to say that 'this is still Smith' about the post-booth subject sounds fishy. Why? That Smith is still Smith is highly inconsistent with the empirical results we saw in Chapter 2. Further, in the Smith case, the post-booth subject seems to have an entirely different relationship to the world—one that is not due to the normal cause (gradual change in brain function through experience, learning, age, etc.) but to abrupt manipulation. As a matter of fact, the post-booth subject will respond to the world and act in ways exactly *opposite* the pre-booth subject, Smith. Even though the pre and post-booth subjects are strongly connected and so psychologically continuous, to call the post-booth subject Smith is inconsistent with just about everything about how the post-booth subject engages with others and the world. The post-booth subject will not act in Smith's characteristic way. I think that this indicates that, contra the NPV, altering only moral features does not *ceteris paribus* amount to persistence. In other words, sameness of some amount of moral character traits seems necessary to persistence in a way not captured by Parfit's view. Below, I will argue that this suggests that the NPV needs to be supplemented by another kind of continuity: moral continuity. First, though, I want to consider two potential objections to my claim in this subsection.

4.2.2 Two Objections to a Necessary Moral Persistence Condition

An immediate worry is that *Malfunctioning Mental State Booth* is just too fanciful. Such selective change to mental states, it might be argued, has no practical bearing. And so, cases like *Malfunctioning Mental State Booth* do not show that moral features are necessary to persistence, as moral features and the rest of our psychological profile cannot ordinarily come apart like they do there.

Yet, we should not be so easily convinced by the detractor. There in fact *are* real-world cases where there appear to be great changes in character while other mental features more or less stay the same. In particular, this happens in cases involving traumatic brain injury. Phineas Gage of course comes to mind, but, as I noted in Chapter 2, there is recent skepticism on the degree of moral change Gage in fact underwent. Still, we can draw upon a more timely example that has Gage-like circumstances. Alissa Alfonina is a woman sensationalized by tabloids as turning “from a star student to a lusty dominatrix” (Natalie O’Neill, *New York Post*, Feb. 6th 2015). Setting aside the ridiculous exaggerations of yellow journalism, court documents *do* support a dramatic shift in moral character after brain injury suffered during a car accident. in his summary of *Alfonina v. Jansson*, Presiding Justice Joel Groves reports that, prior to the accident, Alfonina’s Grade 11 teacher

> found Alissa to be a student who was very bright and interested in her pursuits, and in matters generally. He described her as being in the top 2% in terms of engagement in class activities and assignments. She wanted to pursue a media-arts education...He spoke positively of her high motivation, her goals to become a filmmaker or actress, and her abilities to write imaginative works of fiction. He also noted that her interest in the “big questions” led her to taking a philosophy class that he also taught. He testified that he had no sense that she was troubled emotionally in any way. He found her to be noteworthy on the intellectual scale, in the top 5%, someone who met deadlines and always attended regularly (114).

After the accident, however, Alfonina's character changed remarkably:

> [After the accident], [h]e observed a very different Alissa. He said that she showed signs of no impulse control, could not carry through and tasks were not done. Instructions had to be repeated to her. Things had to be read over and over. She became socially isolated and began to have outbursts in class. She made sexual comments during these outbursts that were inappropriate for the class setting. Mr. Byrne was of the view that she did not appear to filter her thoughts and acted as if she was unaware of her social environment...He gave numerous examples of instances where her lack of social appropriateness required counsellors to intervene. Her talk was unfiltered, random, and as he described, not logical for the school social environment, or "out of left field" (116).

Justice Groves was convinced by these claims and corroborating evidence presented by Alfonina's legal team that her accident indeed caused a pronounced change in character in the form of a personality disorder with symptoms that ranged among "emotional liability, cognitive fatigue, reduced insight, reduced judgment, disinhibition, mood swings, apathy and inflexible thinking" (129). As a result of her disorder, Alfonina had tremendous difficulty holding down a job and became increasingly depressed and apathetic. Eventually, she turned to sex work to support herself financially, including—as the tabloids were so quick to point out—work as a dominatrix for hire.

As Alfonina's unfortunate case illustrates, there are legitimate Gage-type cases that mirror the consequences of *Malfunctioning Mental State Booth* but without the latter's sci-fi fancy. Before her accident, Alfonina appears to have been a driven and creative young woman mindful of her education—at least in areas of interest to her—and concerned about how she related to social peers. Following the accident, though, Alfonina became listless, intellectually closed-off, and, most tellingly, was highly inconsiderate of others. Not all of these changes are overtly changes in moral character; Alfonina also suffered an incapacity for retaining new memories and saw changes to personality traits that are decidedly morally neutral.

At the same time, there is no clear indication that she endured anything like widespread loss of or change to other features. For instance, Alfonina did not report amnesia-like loss of episodic memory, and there seemed to be no significant change to her foundational belief structure. It could even be argued that many of her desires—such as aspirations in filmmaking or other creative media—were still present and accounted for, as Alfonina *did* continue pursuing these interests post-accident in a specialized Grade 12 program for skilled students called Connect Film. When she found herself unable to fulfill these desires due to cognitive impairment from the accident, she became frustrated and withdrew from the program. This means that there appears to be overall psychological continuity in the Alfonina case, and that this continuity is the result of normal brain functioning (the parts of the brain responsible for Alfonina's unchanged features presumably continue to function normally, even if there is abnormal function in changed areas).

So, we can see that, as in *Malfunctioning Mental State Booth*, changes to Alfonina's character traits at least in part led to a radical difference in how she participated in and related to her social ecosystem. In particular, Alfonina's failure to meet standards of social propriety seems indicative of a lack of concern for others. It was not that Alfonina no longer recognized what was socially expected of her. Rather, she just did not seem to care anymore whether her words and actions negatively affected those who shared her classroom. There is a whole web of traits involved here: kindness, respect for others, and concern for the welfare of one's peers, probably among others. Again, I suggest that to say that post-accident person is one and the same Alissa Alfonina is dissonant with her behavioral changes and the underlying disruptions to her character that they represent. This is in spite of the fact that the pre and post-accident person are psychologically continuous.

It might be tempting to say instead that the post-accident person is the ‘same’ Alissa Alfonina and just credit her behavioral changes to brain injury. We might think this especially because the post-accident person evidently retained the majority of her pre-accident mental features and has the same, albeit damaged, physical substrate. My point, though, is that the behavioral changes in question are both quite severe and of such a nature that they make the attribution of sameness seem unnatural—substantially more so than if the psychological changes were, say, solely to learning aptitudes. After all, it is not her learning disabilities that made post-accident ‘Alissa’ so unnerving to Mr. Byrne and others, but her extraordinary and sudden character changes. So, in spite of apparent psychological continuity and (more or less) continuity of the realizer, there still seems to be something wrong with affirming that Alfonina persisted through the accident. And this is true even if we qualify the persistence claim as ‘persisted with brain injuries’.

Reading over Alfonina’s case and calling to mind the specificity of *Malfunctioning Mental State Booth*, the reader might now raise a different objection. ‘You say that those cases show that altering moral features is not *ceteris paribus* compatible with persistence. But the *ceteris* is not *paribus* in these cases. Moral traits are interlinked to a network of other mental states: beliefs, desires, etc. While it *is* possible that I can be an amnesiac and keep the majority of direct connections, loss of moral traits requires the simultaneous loss of too many other features with which those traits are interwoven. For instance, I cannot really be a kind person if I believe that helping others is for chumps. Even the Alfonina case involves the loss of other capacities, as you admit. So, there is no selectively changing one’s mental states, and the conclusion in favor of a moral persistence condition does not follow.’

I see two possible responses here. First, I am not sure that it is clear that moral traits *are* as tightly bound up in other states as my imagined interlocutor might think. There is, of course, the phenomenon of akrasia to consider—acting against what one judges to be the 'right' thing to do, or acting against one's beliefs. But we can generalize on this to think of it as plausible that I might act against my *bad* judgments, too. That is, I might apparently hold the belief that I ought to treat others poorly or selfishly but be faultlessly considerate in my actual interactions with people. This leads to the question of whether this would actually constitute kindness—perhaps I am considerate because I see a social advantage in it. Nonetheless, it seems possible that I could be consistently kind without ever forming the belief that kindness is a good thing or that I ought to be kind. Maybe, for instance, I just genuinely never actually think about it but act anyway.

Whether I am right or wrong in the above, I see a second response to the objection that is much more forceful. Suppose I concede that moral features *are* in fact more closely connected to other mental features than, say, memories or desires. Then this is actually a compelling and principled reason to accept a moral persistence condition and modify the NPV. If this is right, then one could not persist through certain moral changes *just because* changes in moral features result in the severing of too many other direct connections. Therefore, this second response to the objection turns the objection itself into good evidence for why moral features are necessary for persistence. Yes, this would mean that the unmodified NPV actually gives us the right answer in the above problem cases, as there could not be moral change without the severing of enough direct connections to preclude persistence. But the NPV would, on this view, be getting the right answer without the right explanation. So, Parfit's view would still need to be modified to at least highlight that a kind of moral continuity is necessary for persistence.

I grant that it might be possible for the objector to hold out and claim that loss of persistence is still best explained by the severing of so many other direct connections, even if those connections are severed precisely when moral features change. However, I do not find this especially convincing. If moral features are closely interlinked with other mental features in the way stipulated for this response, then moral features form a kind of lynchpin holding together a set of mental states. Remove or change the moral feature in this case, and the entire set collapses. To deny that changing the moral feature severs persistence because it is *really* the many other changing mental states that do the work of severing persistence sounds to me like saying 'It isn't the fall that kills you, it's the sudden stop at the end'. That is, this claim seems to mistake the most immediate cause of an effect to have the most explanatory power. While it is certainly true that falling is not particularly deadly in and of itself, it is equally true that sudden impacts do not normally happen spontaneously to people sitting around in loungers. Likewise, if a loss of persistence results from a change in some set of non-moral mental states caused by a change in the moral feature that binds them together, then it seems accurate to say that that loss of persistence resulted from that same change in moral feature. In any event, it is probably an empirical question whether moral traits cannot in fact be detached from a network of other mental features. My point in considering this pair of responses to the objection is that whatever that answer ends up being, a necessary moral persistence condition is not ruled out.

4.2.3 Moral Continuity and What it Consists In

In the previous section, I argued that examples like *Malfunctioning Mental State Booth* and Alfonina-type brain injuries indicate that the straightforward psychological continuity of the NPV is not enough to account for our persistence. This is because this kind of continuity alone does not capture the specially important constitutive role played by moral features—namely, that continuity of moral features are a necessary condition for persistence. But if the straightforward psychological continuity of the NPV could be modified to include a place for moral continuity as necessary, then we can get the right answers for the right reasons in these problem cases. In this section, I propose a way to understand moral continuity in parallel to the basic features of Parfit's NPV. That is, I take it that (a) moral features are subject to the same kind of restrictions as other psychological states on the NPV and that (b) once defined this way, necessary moral continuity can be appended to Parfit's view as an additional clause to form a new account of persistence: the Narrow *Moral* Psychological View.

We will consider (a) first. To say that moral features are subject to the same kind of restrictions as other psychological states is to say that moral continuity is a kind of subset of standard psychological continuity under the NPV. So, this would mean that moral continuity consists in the following parallel premises to the Psychological Criterion:

> *Moral Continuity*: There is *moral continuity* if and only if (1) there are overlapping chains of strong connectedness (in regards to moral features) between subjects, (2) this continuity has the right kind of cause, and (3) there does not exist a different person who is also morally continuous with Y. (4) Moral continuity just consists in the holding of facts like (1) to (3).

Characterizing moral continuity in this way has important consequences. Principally, I use this formulation in order to clearly convey that moral continuity does not preclude *any* moral change at all. (1) allows for gradual changes in moral features: some number of our moral character traits can change without disrupting continuity, so long as there remain 'enough' direct moral connections. As with Parfit's view, what counts as 'enough' will have to be imprecise, and I deal with this indeterminacy in the next section. (2) anticipates and guards against worries like Chapter 3's *Gradual Brainwashing*. Though moral change is incremental in this case, it is not taking place by means of the normal cause. Said differently, the apparent ‘continuity’ of *Gradual Brainwashing* comes about through a mad doctor’s manipulation, not through the process of learning and experiencing in ordinary aging. As in the normal Psychological Criterion, (3) prevents fission problems. And finally, (5) tells us that moral continuity is reductionist—there is no ‘further fact’ in which moral continuity consists. Just like in the NPV, (5) has meaningful implications for our ability to ascertain in some specific cases *when* moral continuity has either obtained or lapsed. I explore these implications in the final section of this chapter. For now, it is enough I think to note moral continuity’s reductionism.

Given this formulation of moral continuity, it is rather simple to approach (b): we just add another clause to the NPV that includes the moral continuity claim. I include moral continuity as an additional clause rather than incorporating it into the definition of psychological continuity because I want to allow for the possibility that psychological continuity can obtain without moral continuity. As stated above, I do not think that I am in a position to argue whether or not moral traits are truly bound up so tightly with other mental features that altering certain moral traits would change enough direct connections to also sever persistence. If moral traits are mostly independent of other mental features, then psychological continuity and moral continuity can be teased apart and an additional clause is required to account for a necessary moral persistence condition. On the other hand, if moral traits are mostly interdependent with other mental features, then psychological continuity always requires moral continuity. In this latter case, updating the definition of psychological continuity would be the preferable solution.

Even so, adding a ‘moral continuity clause’ instead still would not undermine the argument in this case—it would just be redundant, as the requirement for psychological continuity would already effectively contain a moral continuity clause. As a result, since I cannot determine how moral traits in fact connect to other mental features, I will go on assuming that moral traits are at least somewhat independent and bring in an additional clause that leaves the definition of psychological continuity untouched. Should this turn out to be wrong, little needs to be changed to accommodate the interdependence of moral traits with one's total psychology: the additional clause vanishes, and moral continuity is folded into the definition of psychological continuity.

Now, to return to the additional clause of (b). Recall that the NPV formulation is:

> *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) [Persistence] just consists in the holding of facts like (2) to (4) (Parfit 1984, 207).

But I have argued that persistence necessarily requires moral continuity, so we should now rearrange the premises to include both what moral continuity is and that X must be morally continuous with Y in order for X and Y to be one and the same person at different times. Hence, the revised formulation is:

> *The Narrow Moral Psychological View of Persistence*: (1)' There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. (2)' There is *moral continuity* if and only if there are overlapping chains of strong connectedness among the relevant moral features. X today is one and the same person as Y at some past time if and only if (3)' X is psychologically continuous with Y, (3)' X is morally continuous with Y, (4)' these continuities have the right kind of cause, (5)' there does not exist a different person who is also psychologically or morally continuous with Y, (6)' Persistence just consists in the holding of facts like (2)' to (5)'.

We can see that only minimal refiguring is needed to get the NPV to the more specific NMPV. All the same, I think that this small change is enough to comfortably integrate the necessity of moral continuity to persistence in a way that satisfies the examples from the previous section *and* helpfully meets the moral persistence intuitions described in Chapter 2.

4.3 What the Narrow Moral Psychological View Implies About Persistence

If we accept that moral continuity is a necessary condition for persistence, and if we further accept that NMPV adequately captures this necessity, then we now have a successful account of persistence that includes a moral persistence condition. In this last section, I examine what kind of implications this account has for our persistence.

First, we can see that, although moral continuity is necessary for persistence, the NMPV does not hold that it is sufficient. General psychological continuity is also necessary for persistence, per (1)'. My view would not be a theory of psychological continuity at all if this were not true. Consider the following case:

> *The Moral Accident Victim*: a man, Quaid, is involved in a catastrophic accident (the details are unimportant, other than it is a head injury) that results in a most peculiar impairment. Quaid loses access to all previously held memories, beliefs, desires, intentions, and personality traits. In fact, the only previously held mental features that remain are his moral character traits. Quaid still interacts with others in his social environment with the same warmth, kindness, and generosity that he did before, even though he does not remember anything about himself or others, lacks a beliefs structure that verifies that these acts are acts that he ought to do, and has no real desire to do these things.

Quaid is not psychologically continuous with the post-accident man. There are exceedingly few direct connections between them. Despite the fact that his behavior has not changed, we can and should say that the post-accident man does not relate to others in the world in the way that the pre-accident man did: indeed, just about all of his relations with others in the world were severed by the accident. So, it is clear that we cannot accept that Quaid persisted through the accident if we believe any sort of psychological continuity is necessary for persistence. And *The Moral*

Accident Victim implies that, absent psychological continuity, moral continuity is not sufficient for persistence. There just is not enough of 'Quaid' left over, and this is clear by the deficit in the post-accident man's relations to the world.

As promised, I also need to here return to what reductionism about moral continuity means for persistence. Remember that the implied reductionism in the NPV resulted in a sort of indeterminacy in persistence. That is, there were situations along Parfit's *Combined Spectrum* where it is quite unclear if something is me or someone else. In such cases, the best that can be done is to give an account of the physical and psychological facts. But we cannot say more than this—it's me! It's not!—without drawing an implausible borderline. Since moral continuity adopts this same framework, it will also suffer from indeterminacy in select cases. Indeed, since there are two continuities involved, and since we are assuming that psychological continuity and moral continuity can come apart, there may be cases where one continuity but not the other is indeterminate.

We can imagine a more selective *Combined Spectrum*—let's call it the *Moral Combined Spectrum*, out of convenience. On this spectrum, the only brain states Parfit exchanges with Greta Garbo are those which specifically give rise to moral traits (assume, of course, that we have a sophisticated enough neuroscience to make this kind of fine-grained identification possible). As before, these states are replaced only incrementally. Then, also as before, there will be cases in the middle of the spectrum where either it will be true that a single brain cell makes the difference between something being me and it being someone else, or there will be no answer beyond just giving the physical and moral-psychological facts.

Following Parfit's earlier argument for reductionism, it seems highly implausible for there to be a sharp borderline when the difference between persistence and not is an apparently trivial cell (or single moral trait). Indeed, as awkward as the reductionist result is—that whether I persist is sometimes indeterminate—it seems much more plausible than the sharp borderline. And anyway, the number of cases where we would get this indeterminate result is extremely limited to only bizarre *Moral Combined Spectrum* type scenarios. As with Parfit's view, I do not think these limited cases actually threaten the coherence of the NMPV. In the overwhelming majority of cases—what is more, in all *actually* possible cases—moral continuity and so persistence will always be determinable.

What is more, the fact that two continuities are involved does not actually complicate things in the way it might suggest. Since psychological continuity is also required for persistence, there will be no persistence cases of any concern where psychological continuity is indeterminate but moral continuity is clear. Moral continuity alone is insufficient for persistence, so indeterminacy at the level of psychological continuity is enough to give us pause and say that we can only give the reductionist description of facts. Any complexity moral continuity contributes would come from cases where psychological continuity is clear but moral continuity is indeterminate. Again, though, these cases are limited to *Combined Spectrum* and *Combined Moral Spectrum* type cases, and the number of these cases is both extremely small and restricted to only possible (but not actual) cases. So, most cases by far (and all actual cases) are still determinable—even with two continuities.

Setting concerns of indeterminacy aside, it would be good to point out where determinations of persistence differ between Parfit's view and mine. Most of the time, both views will agree on when I persist. For instance, both the NPV and the NMPV are clear that we do persist through brain transplants but not cerebral death. Similarly, both straightforwardly accept that degenerative brain diseases like Alzheimer's will kill us through psychological disunity long before the human animal of which we are a part dies. The difference comes in cases where moral continuity but not continuity of my total psychology is severed. Arguably, victims of traumatic brain injury like Alissa Alfonina represent a real-world example: genuine Phineas Gage-types who suffer moral discontinuity while retaining most of their other psychological features (if not all of their psychological capacities). Of course, I could on Parfit's view persist through a more limited version of *Gradual Brainwashing*, too. Supposing that only a small number of features were altered, it ultimately does not matter that these changes were not brought about via the normal cause. On the NMPV, though, I certainly cannot survive gradual and complete change of my moral features, for reasons given above.

Towards the beginning of this dissertation, I mentioned a few cases that might intuitively prompt us to seriously consider a role for the moral in constituting persistence. At the time, I meant those to be merely suggestive. We *do* in fact often say things like 'after the accident, she's not the same person' or 'he came out of prison a new man.' But we might only be speaking qualitatively: she—one and the same person—has qualitatively different features/behavior after her accident. When we look at these and other paradigmatic cases involving apparent moral change, the NMPV reveals that we do in fact persist through most such cases and, in turn, through most moral change. The prisoner's new moral traits were probably taken on as part of a gradual process of learning and experience and are thus part of ordinary mental continuity. So, he persists through his time in prison and indeed emerges as a 'better' man.

Most religious converts also persist through what is nonetheless a powerful transformation of character, as there is good reason to believe that some conversion is also a gradual process of development rather than momentous and instantaneous. That is, this type of conversion is a process of *becoming*. Admittedly, at least some cases of conversion appear to be authentically instantaneous. However, even Augustine’s sudden turn to God in Book Eight of the *Confessions*—the standard in terms of ‘instantaneous’ conversion—looks like the result of many *years* of steady change in character and ‘soul searching’. As Steven Affeldt (2013) argues, Augustine’s conversion probably also did not *end* when he picked up a Bible in a Milan garden. Rather, the conversion experience was for Augustine an ongoing practice of sustaining and inhabiting the new values of his faith. If there are cases of genuine instantaneous conversion, then those cases would present situations where on the NMPV the subject does not persist, as moral continuity is disrupted by the sudden severing of so many direct moral connections. But in fact, instantaneous conversion might disrupt plain old psychological continuity on Parfit’s view, too, depending on how wide-ranging the psychological changes are. So, instantaneous conversion cases notwithstanding, the NMPV holds that we can persist through religious conversion.

Both the prisoner and religious convert cases involve gradual moral change over time. Stepping back a bit to adopt a big-picture stance, the fact that we can persist through this kind of gradual change holds real consequences. Perhaps unfortunately for those among us who lament jarring moral changes we observe in old friends, we can survive complete inversion of character that may occur naturally and slowly through aging, as moral continuity is not there interrupted. Recall the kind of surprise found in the Strohminger and Nichols example from Chapter 2, where someone has dinner with a friend not seen in many years. Though respondents reported that moral changes in this friend would be most indicative that the original person no longer persists (and is not, in fact, sitting in front of them sharing dinner), we see now that on my view this is wrong. It would, of course, be a very nice comfort to be able to say that the bitter, mean, and intolerant person sitting before us really is not one and the same person as the warm, considerate, and open-minded friend we knew years ago. But this cuts both ways. If the NMPV *did not* accommodate the gradual moral change of aging—in other words, if moral continuity did not allow *some* direct connections to be severed—then the NMPV would have the unhappy and highly implausible result that we could never morally improve ourselves.

Despite the fact that my view runs counter to some of our moral persistence intuitions in this case, I argue that it still gives us the right answer. First, this is because the alternative of a sense of moral continuity that does not permit personal moral improvement (or decrement) is likely far too difficult to accept. It is better to admit that gradual moral change resulting from aging and both institutional and environmental factors is perfectly compatible with moral continuity. Second, the NMPV still captures something important about our intuitions in even this divergent Strohminger and Nichols case. Namely, it fully accommodates the intuitive sense reflected in this case that moral character is not just important to persistence, but actually more important than any other psychological feature. The dispute is not over whether or not moral continuity is necessary for persistence. Instead, the Strohminger and Nichols case pushes back against a moral continuity that accepts gradual moral change. So, even this challenging case can accept the necessity of moral continuity, even if it intuitively questions whether that continuity goes far enough.

Excluding instantaneous conversion, the only paradigmatic practical cases that the NMPV says we *would not* survive are incidents involving severe brain injury, like Alissa Alfonina's. Such accidents disrupt moral continuity through sudden change that is abnormally caused. We can consider, though, a variant of this case type that is less obviously harmful to moral continuity:

> *The Gradual Accident Victim*: a woman, Rachel, is involved in an accident which causes traumatic brain injury. However, there are few immediate moral changes. Rather, over time there are physical complications from her injury that slowly begin to manifest. Rachel's moral character traits change little by little until she is, years later, morally the inverse of pre-accident Rachel.

This case seemingly preserves moral continuity by making moral change a gradual process; few connections are severed all at once. However, I think we should resist the urge to say that Rachel persisted through her full moral transformation. Like *Gradual Brainwashing*, Rachel's moral changes are brought about by an abnormal cause: a malfunctioning and degenerative cognitive system. Yet, things are not quite as straightforward as *Gradual Brainwashing*, either. It seems as if Rachel *does* in fact persist through the initial accident, and it is only later that continuity is severed at whatever point when the accident victim is no longer strongly connected with pre-accident Rachel. What this means is that the *Gradual Accident Victim* gives us a real-world example of the *Moral Combined Spectrum*—what we might call *Rachel's Spectrum*. On the NMPV, there would come a point in Rachel's moral changes at which the accident victim was no longer morally continuous with Rachel. But this forces us to revisit the problem of the sharp borderline from before. There may be such a borderline in *Rachel's Spectrum* between something being Rachel and someone else, but we could never know where that line is, and defining it would appear quite arbitrary. In the interest of plausibility, the NMPV's reductionism would instead recommend that we can only describe the physical and psychological facts. That is, we can at some points in *Rachel's Spectrum* do no more than tell what is happening in the brain and which moral character traits have changed. What we cannot do is feasibly say whether the present person is Rachel or someone else.

Obviously, this is an unwelcome result. Part of my dismissal of the indeterminacy of persistence in certain cases was the fact that such cases were limited to extremely fanciful sci-fi thought experiments. But *The Gradual Accident Victim* seems to at least be grounded in real possibility. Does this not threaten the legitimacy of my view? I do not think so. For one, *The Gradual Accident Victim* is still a highly unlikely case—it requires that moral changes be incremental and very nearly trivial. It is therefore very implausible that any *actual* accidents could fit the bill.

Even if some incredibly small number of actual cases *did* fit, though, this does not mean that we need to throw up our arms and accept the Parfitian conclusion that persistence is not what matters. For the same reasons that we separate babies and bathwater, a small number of cases where persistence turns out to be indeterminate does not recommend in favor of tossing out persistence altogether in guiding our worries about things like blame, attribution, and accountability. After all—as we have observed since at least Aristotle—the practical is concerned with the actual and everyday. In terms of the everyday, the NMPV reliably gives a determinable answer about persistence nearly all of the time. So, any threat posed by *The Gradual Accident Victim* to my view is a false one. Occasional rulings of indeterminacy do not necessitate that we abandon or even diminish the importance that persistence holds in shaping our practical judgments.

CHAPTER 5

Conclusions

In this dissertation, I have argued for a moral persistence condition. More specifically, I argued that moral continuity—as part of overall psychological continuity—is a necessary condition for one of us to persist. Getting to this argument was a complicated process. After an in-depth look at the literature in Chapter 1, I concluded that moral character is a neglected feature of persistence. Despite being regularly involved in persistence discussions both as a motivator and an important consequence, a survey of the persistence literature clearly showed that moral character is rarely if ever taken to figure into what constitutes persistence. At this point, I also made the case that only a psychological view of persistence could support a moral persistence condition, as moral character traits are psychological features.

The neglect of moral traits in philosophical discourse about persistence is all the more surprising when we compare philosophical talk of persistence with folk or commonsense notions of it. As I argued in Chapter 2, there is good evidence that ordinary people tend to regard moral character as the most important condition in determining whether I persist or not. Of course, it is still just as true at the end of this dissertation as it was in Chapter 2 that the mere fact that people have these moral persistence intuitions says nothing about whether or not they are right. But, I argued that as long as our intuitions are not unstable or biased—and further that our moral persistence intuitions are not either—we are justified in thinking these intuitions might be on to something. In other words, that we have these intuitions is not proof of a moral persistence condition, but they do suggest that a moral persistence condition is plausible (and, moreover, they give us a good reason to investigate whether there is, in fact, a moral persistence condition).

In the second half of the dissertation, I developed an account of how a moral persistence condition might work. Chapter 3 was devoted to constructing an underlying personal ontology that could sustain a persistence theory that contained a moral persistence condition. That personal ontology was the State-Realizer view of what we are. On this view, we are psychological entities with a constant physical realizer; that is, we are composed both of sets of mental states and the physical realizers in which these states are instantiated. While I do not think my account of what we are is the only account of personal ontology that is compatible with a moral persistence condition, I argued that the State-Realizer view is both internally consistent—despite some unwelcome consequences—and preferable to competing views.

Finally, in Chapter 4 I built upon the State-Realizer view to posit the Narrow Moral Psychological View (NMPV) of persistence. This is a modified version of Parfit's Narrow Psychological View (NPV) which adds a requirement for moral continuity as a necessary condition for persistence. Furthermore, I defined moral continuity broadly as overlapping chains of direct connectedness between moral features; these chains are subject to the same non-branching, normal cause, and reductionist requirements of standard psychological continuity on Parfit's NPV. One thing that is important to note is that the NMPV does not ultimately change very much in terms of the NPV's practical effects. As I argued in the last section of Chapter 4, only certain cases seem to separate NMPV from NPV—namely, Gage-type accidents involving traumatic brain injury, *The Gradual Accident Victim*, and some versions of *Gradual Brainwashing*. It might turn out, as I considered in an objection to *Malfunctioning Mental State Booth* in the last chapter, that moral traits are more deeply bound up in other mental states than some of my examples would intimate. If so, then general psychological continuity and moral continuity can *never* come apart.

That NMPV practically changes little about NPV is the desirable outcome, to me. My goal was not to pose a radical reevaluation of persistence. In general, I think that NPV probably gives the right answers to most persistence questions. Yet, NPV seemed to miss something significant about how we persist in certain cases. It could not accommodate our moral intuitions from Chapter 2, and it seemed to give the wrong answer in cases like *Malfunctioning Mental State Booth*. Compared to other mental features, moral character traits might be said to have a special link to actions. I can arguably act against my beliefs, I can certainly suppress my desires, and my intentions can be frustrated and set aside. But my actions—in an Aristotelian sense—seem to stem from and reflect my character. Among psychological features, character traits have a special relationship with the world. And if they *are* uniquely interlinked with other mental states, then moral traits even more clearly have a distinctive status among mental features. NMPV tries to give a metaphysical underpinning to this special relationship by allowing our moral features to be a necessary part of what we are and how we persist.

This is a novel idea: again, the literature downplays or outright ignores any constitutive role moral features could play in persistence. What I hope to have accomplished here is then twofold. First, I think that my treatment at least shows that we ought to be thinking more seriously about how moral features *could* take on such a constitutive role in persistence. Perhaps others can use this account as a platform out of which to incorporate moral persistence conditions into other psychological persistence theories. Second, I believe my account gives us a way to bridge the gap Schechtman observes between our practical concerns and persistence (see Chapter 1). Against Schechtman, I have repeatedly argued that we can on NMPV or even NPV resist Parfit's argument that persistence is not what matters. This would, in turn, save us from the Extreme Claim—the view that none of our practical concerns are grounded in anything at all.[39]

I think that we can still ground these concerns in persistence. In fact, I have made the stronger claim that our moral features are deeply constitutive of who we are. In a certain sense, the practical is tightly tied to persistence in just this way. Though obviously the specific practical concepts Schechtman talks about—survival, moral responsibility, self-interested concern, and compensation—are not constitutive elements of persistence, my account necessarily builds the moral features that generate practical worries like these *into* persistence. The result is that not only is persistence still what matters, but what matters—the moral—is part of persistence.

[39] Unfortunately, I cannot really argue for this more fully here and anticipate ways that Schechtman might respond to my resolution to her problem; this would be the topic of another project entirely.

REFERENCES

Affeldt, Steven G. 2013. "Being Lost and Finding Home: Philosophy, Confession, Recollection, and Conversion in Augustine's *Confessions* and Wittgenstein's *Philosophical Investigations*." In *Wittgenstein Reading*, edited by Sascha Bru, Wolfgang Huemer, and Daniel Steuer. Berlin: Walter de Gruyter and Co.

Armstrong, David. 1984. "Consciousness and Causality." In *Consciousness and Causality*, edited by David Armstrong and Norman Malcolm, 103-92. Oxford: Blackwell.

Aquinas, St. Thomas. 1956. *Summa contra Gentiles II*. Translated under the title *On the Truth of the Catholic Faith, Book II: Creation*, translated by James F. Anderson. New York: Doubleday.

Augustine. 2009. *Confessions*, translated by Henry Chadwick. Oxford: Oxford University Press.

Ayers, Michael. 1991. *Locke: Volume II, Ontology*. London: Routledge.

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

Baker, Lynne Rudder. 2002. "The Ontological Status of Persons." *Philosophy and Phenomenological Research* 65: 370-88.

Beck, Simon. 2016. "Technological Fictions and Personal Identity: On Ricoeur, Schechtman, and Analytic Thought Experiments." *Journal of the British Society for Phenomenology* 47: 117-32.

Bennett, Karen. 2004. "Spatio-temporal coincidence and the grounding problem." *Philosophical Studies* 118: 339-71.

Butler, Joseph. 1736. "Of Personal Identity." In *The Analogy of Religion*, reprinted in Perry 1975, 99-105. Berkeley: University of California Press.

Carroll, Lewis. 1998. *Alice's Adventures in Wonderland*. Chicago: Volume One Publishing.

Carruthers, Paul. 1996. *Language, Thought, and Consciousness*. Cambridge: Cambridge University Press.

Chisholm, Roderick. 1969. “The Loose and Popular and the Strict and Philosophical Senses of Personal Identity.” In *Perception and Personal Identity*, edited by Norman S. Care and Robert H. Grimm, 82-106. Cleveland: Press of Case Western Reserve University.

Chisholm, Roderick. 1991. “On the Simplicity of the Soul.” *Philosophical Perspectives* 5: 167-81.

Clark, Andy. 2003. *Natural Born Cyborgs*. Oxford: Oxford University Press.

Clark, Andy and David Chalmers. 1998. "The Extended Mind." *Analysis* 58, 7-19.

Cline, B.W. 2012. “You’re Not the Same Kind of Human Being: the Evolution of Pity to Horror in Daniel Keyes’s *Flowers for Algernon*.” *Disability Studies Quarterly* 32.

Collins, Steven. 1982. *Selfless Persons*. Cambridge: Cambridge University Press.

Dainton, Barry and Tim Bayne. 2005. *Australasian Journal of Philosophy* 83: 549-71.

Davis, Lawrence. 1998. “Functionalism and Personal Identity.” *Philosophy and Phenomenological Research* 58: 781-804.

Davis, Lawrence. 2001. “Functionalism, the Brain, and Personal Identity.” *Philosophical Studies* 102: 259-79.

Dennett, Daniel. 1978. “A Cure for the Common Code.” In *Brainstorms: Philosophical Essays on Mind and Psychology,* 90-108. Cambridge: Bradford Books.

Descartes, Rene. 1911. “Meditations on First Philosophy.” In *The Philosophical Works of Descartes: Volume I,* translated by Elizabeth S. Haldane and G.R.T. Ross, 131-200. Cambridge: Cambridge University Press.

Descartes, Rene. 1934. “Reply to Fifth Objections.” In *The Philosophical Works of Descartes: Volume II,* translated by Elizabeth S. Haldane and G.R.T. Ross, 204-33. Cambridge: Cambridge University Press.

Fine, Kit. 1999. “Things and Their Parts.” *Midwest Studies in Philosophy* 23: 61-74.

Foster, John. 1991. *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. London: Routledge.

Fuller, Gary. 1992. “Functionalism and Personal Identity.” *The Personalist Forum* 8: 133-43.

Gert, Bernard. 1971. “Personal Identity and the Body.” *Dialogue* 10: 458-78.

Giles, James. 1993. "The No-Self Theory: Hume, Buddhism, and Personal Identity." *Philosophy East and West* 43: 175-200.

Green, Michael B. and Daniel Wikler. 1980. "Brain Death and Personal Identity." *Philosophy and Public Affairs* 9: 105-33.

Grice, H.P. 1941. "Personal Identity." *Mind* 50: 330-50.

Gunnarsson, Logi. 2009. *Philosophy of Personal Identity and Multiple Personality*. London: Routledge.

Hume, David. 1896. *Treatise of Human Nature,* edited by L.A. Selby-Bigge. Oxford: Clarendon Press.

Hutchins, Edwin. 1995a. *Cognition in the Wild*. Cambridge: Bradford Books.

Hutchins, Edwin. 1995b. "How a Cockpit Remembers its Speeds." *Cognitive Science* 19: 265-88.

Jaworski, William. 2016. *Structure and the Metaphysics of Mind*. Oxford: Oxford University Press.

Johnston, Mark. 1987. "Human Beings." *The Journal of Philosophy* 84: 59-83.

Johnston, Mark. 2006. "Hylomorphism." *The Journal of Philosophy* 103: 652-98.

Kant, Immanuel. 1964. *Critique of Pure Reason*, translated by N. Kemp Smith. London: Macmillan.

Koons, Robert. 2014. "Staunch vs. Faint-Hearted Hylomorphism: Toward an Aristotelian Account of Composition." *Res Philosophica* 91: 151-78.

Koslicki, Kathrin. 2008. *The Structure of Objects*. Oxford: Oxford University Press.

Lewis, David. 1983. "Survival and Identity." In *Philosophical Papers Volume I,* 55-72. Oxford: Oxford University Press.

Lichtenberg, G.C. 1971. "Schriften und Briefe." In *Sudelbucher II*. Munich: Carl Hanser Verlag.

Locke, John. 1975. *An Essay Concerning Human Understanding*, edited by Peter Nidditch. Oxford: Oxford University Press.

Lockwood, Michael. 1988. "Warnock versus Powell (and Harradine): When Does Potentiality Count?" *Bioethics* 2: 187-213.

Lockwood, Michael. 1994. “Identity Matters.” In *Medicine and Moral Reasoning,* edited by K.W.M. Fulford, Grant Gillett, and Janet Martin Soskice, 60-74. Cambridge: Cambridge University Press.

Lycan, William. 1996. *Consciousness and Experience*. Cambridge: MIT Press.

Mackie, J.L. 1985. "Multiple Personality." In *Persons and Values: Selected Papers Vol II,* edited by Joan Mackie and Penelope Mackie. Oxford: Clarendon Press.

Macmillan, Malcolm and Matthew L. Lena. 2010. “Rehabilitating Phineas Gage.” *Neuropsychological Rehabilitation* 20: 641-58.

Marmadoro, Anna. 2013. “Aristotelian Hylomorphism without Reconditioning.” *Philosophical Inquiry* 36: 5-22.

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

Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press.

Nichols, Shaun and Michael Bruno. 2010. “Intuitions about Personal Identity: An Empirical Study.” *Philosophical Psychology*, 23, 293–312.

Noonan, Harold. 1991. *Personal Identity*. London: Routledge.

Oderberg, David. 2007. *Real Essentialism*. New York: Routledge.

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

Olson, Eric. 2007. *What Are We?* Oxford: Oxford University Press.

Olson, Eric. 2009. “Personal Identity.” In *Science Fiction and Philosophy*, edited by Susan Schneider 69-90. Oxford: Blackwell.

Parfit, Derek. 1984. *Reasons and Persons*. New York: Oxford University Press.

Parfit, Derek. 1995. “The Unimportance of Identity.” In *Identity,* edited by H. Harris, 13-45. Oxford: Oxford University Press.

Parfit, Derek. 2012. “We Are Not Human Beings.” *Philosophy* 87: 5-28.

Peirce, C.S. 1935. *The Collected Papers of Charles Sanders Peirce, Vol. VI: Scientific Metaphysics,* edited by Charles Hartshorne and Paul Weiss. Cambridge: Harvard University Press.

Perry, John. 1975. "Introduction." In *Personal Identity*, 3-30. Berkeley: University of California Press.

Perry, John. 1976. "Review: Problems of the Self: Philosophical Papers, 1956-1972 by Bernard Williams." *The Journal of Philosophy* 73: 416-28.

Perry, John. 2002a. "Personal Identity and the Concept of a Person." In *Identity, Personal Identity, and the Self*, 119-44. Indianapolis: Hackett Publishing Company.

Perry, John. 2002b. "Personal Identity, Memory, and the Problem of Circularity." In *Identity, Personal Identity, and the Self*, 84-104. Indianapolis: Hackett Publishing Company.

Puccetti, R. 1973. "Brain Bisection and Personal Identity." *The British Journal for the Philosophy of Science* 24: 339-55.

Radden, Jennifer. 1996. *Divided Minds and Successive Selves*. Cambridge, MA: MIT Press.

Rea, Michael. 1998. "In Defense of Mereological Universalism." *Philosophy and Phenomenological Research* 58: 347-60.

Rea, Michael. 2011. "Hylomorphism Reconditioned." *Philosophical Perspectives* 25: 341-58.

Reid, Thomas. 1785. "Of Memory." In *Essays on the Intellectual Powers of Man*, reprinted in Perry 1975 as "Of Identity," 107-12. Berkeley: University of California Press.

Ricoeur, Paul. 1992. *Oneself as Another*, translated by Kathleen Blamey. Chicago: University of Chicago Press.

Rosenthal, David. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.

Rovane, Carol. 1998. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.

Shapiro, Michael. 2005. "The Identity of Identity: Moral and Legal Aspects of Technological Transformation." In *Personal identity: Vol 22 Part 2*, edited by E.F. Paul, F.D. Miller, and J. Paul, 308-74. Cambridge: Cambridge University Press.

Schechtman, Marya. 1996. *The Constitution of Selves*. New York: Cornell University Press.

Schechtman, Marya. 2014. *Staying Alive*. Oxford: Oxford University Press.

Shoemaker, Sydney. 1963. *Self-Knowledge and Self-Identity*. New York: Cornell University Press.

Shoemaker, Sydney. 1970. “Persons and Their Pasts.” *American Philosophical Quarterly* 7: 269-85.

Shoemaker, Sydney. 1984. “Personal Identity: a Materialist’s Account.” In *Personal Identity,* edited by Sydney Shoemaker and Richard Swinburne, 67-132. Oxford: Blackwell.

Shoemaker, Sydney. 1997. “Self and Substance.” *Nous* 31: 283-304.

Shoemaker, Sydney. 1999. “Self and Body.” *The Aristotelian Society* 73: 287-306.

Shoemaker, Sydney. 2003. “Realization, Micro-Realization, and Coincidence.” *Philosophy and Phenomenological Research* 67: 1-23.

Shoemaker, Sydney. 2004. “Functionalism and Personal Identity—A Reply.” *Nous* 38: 525-33.

Shoemaker, Sydney. 2008. “Persons, Animals, and Identity.” *Synthese* 162: 313-24.

Shoemaker, Sydney. 2011. “On What We Are.” In *The Oxford Handbook of the Self,* edited by Shaun Gallagher, 352-71. Oxford: Oxford University Press.

Sider, Theodore. 1997. “Four Dimensionalism.” *Philosophical Review* 106: 197-231.

Sider, Theodore. 2001. “Criteria of Personal Identity and the Limits of Conceptual Analysis.” *Philosophical Perspectives* 15: 189-209.

Snowdon, Paul F. 1990. “Persons, Animals, and Ourselves.” In *The Person and the Human Mind,* edited by Christopher Gill, 83-108. Oxford: Clarendon Press.

Stcherbatsky, Theodore. 1919a. “The Soul Theory of the Buddhists I.” *Bulletin de l’Academie des Sciences de Russie* 13: 823-54.

Stcherbatsky, Theodore. 1919b. “The Soul Theory of the Buddhists II.” *Bulletin de l’Academie des Sciences de Russie* 13: 937-58.

Strawson, Galen. 2011. *Locke on Personal Identity: Consciousness and Concernment*. Princeton: Princeton University Press.

Strawson, P.F. 1966. *The Bounds of Sense: An Essay on Kant’s Critique of Pure Reason*. New York: Harper and Row.

Strohminger, Nina and Shaun Nichols. 2014. "The Essential Moral Self." *Cognition* 131, 159-71.

Sutton, John. 2010. "Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process." In *The Extended Mind,* edited by Richard Menary. Cambridge: MIT Press.

Swinburne, Richard. 1984. "Personal Identity: the Dualist Theory." In *Personal Identity*, edited by Sydney Shoemaker and Richard Swinburne, 1-66. Oxford: Blackwell.

Thomson, Judith J. 1997. "People and Their Bodies." In *Reading Parfit*, edited by J. Dancy, 202-29. Oxford: Blackwell.

Tobia, Kevin. 2015. "Pesonal Identity and the Phineas Gage Effect." *Analysis* 75: 396-405.

Tye, Michael. 2003. *Consciousness and Persons: Unity and Identity*. Cambridge: MIT Press.

Unger, Peter. 1979a. "I Do Not Exist." In *Perception and Identity*, edited by G.F. MacDonald, 235-51. London: The Macmillan Press.

Unger, Peter. 1979b. "There Are No Ordinary Things." *Synthese* 41: 117-54.

Unger, Peter. 1979c. "Why There Are No People." *Midwest Studies in Philosophy* 4: 177-222.

Van Inwagen, Peter. 1990. *Material Beings*. New York: Cornell University Press.

Van Cleve, James. 2008. "The Moon and a Sixpence: A Defense of Mereological Universalism." In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean Zimmerman, 321-40. Oxford: Blackwell.

Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24: 227-48.

Wiggins, David. 1967. *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.

Wiggins, David. 1980. *Sameness and Substance*. Cambridge: Harvard University Press.

Wiggins, David. 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Wilkes, Kathleen. 1981. "Multiple Personality and Personal Identity." *The British Journal for the Philosophy of Science* 32: 331-48.

Wilkes, Kathleen. 1988. *Real People: Personal Identity Without Thought Experiments*. Oxford: Oxford University Press.

Williams, Bernard. 1957. "Personal Identity and Individuation." *Proceedings of the Aristotelian Society* 67: 229-52.

Williams, Bernard. 1970. "The Self and the Future." *The Philosophical Review* 79: 161-80.

Zimmerman, Dean. 2009. "Properties, Minds, and Bodies: An Examination of Sydney Shoemaker's Metaphysics." *Philosophy and Phenomenological Research* 78: 673-738.