Persons and Their Parts

Persons and Their Parts

CHARLES MARINUS JANSEN

A thesis submitted for the degree of
Doctor of Philosophy

UCL, Department of Philosophy

UCL

Declaration

I, Charles Marinus Jansen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

<div align="right">

—————————————————————

Charles Jansen
</div>

Abstract

This thesis investigates the metaphysics of human persons. In it, I consider two questions.

The first question concerns what types of change we can survive. Here, I argue against the popular 'wide psychological view'. According to this view, our survival requires the presence of particular kinds of psychological connection between our present and future selves, but does not set constraints on the way in which these connections are realised; it is all the same whether our bodies and brains continue to exist, or whether we are beamed from one place to another by teleportation. I argue that we do not have firm intuitions in favour of the wide psychological view, that the view must be carefully formulated in order to give a coherent account of our identity, and that – when so formulated – it is less attractive than an alternative 'narrow view', on which our survival requires the continued existence of the realisers of our mental states.

The second question concerns the parts that we have. Here, I defend the thesis that we have humanoid form against the view that we are only the size of our brains. I do so by arguing that many of our mental states are primarily attributable to humanoid entities, and (at best) only derivatively attributable to anything that is presently brain-sized.

The picture that emerges from the thesis as a whole is one on which our 'normal' or 'characteristic' form is humanoid, but we have a central individuating nucleus, whose continued functioning is sufficient for our continued existence. I find this an attractive view, and conclude the thesis by considering its implications for Kit Fine's account of the nature of material things.

Impact Statement

As of April 2018, UCL requires an impact statement to be included in all PhD theses, which should describe how the expertise, knowledge, analysis, discovery or insight presented in the thesis could be put to beneficial use.

This thesis articulates a series of ideas about our nature. The conception ultimately developed resembles, and builds upon, David Wiggins' one-time belief that we have an 'individuating nucleus', which serves as the 'seat' of our mental capacities (Wiggins 1967, p. 51), and Mark Johnston's claim that we could survive as lone brains, though this would be, for us, an uncharacteristic 'mutilated' form of existence. (Johnston 1987a, p. 79). Neither thinker argues for these views in any detail – they simply highlight their feasibility. The novel contribution of this thesis is to indicate how these views naturally emerge from attempts to spell out what is required for us to continue to exist. In this, and much of the methodology pursued, there are clear resemblances to Peter Unger's defence of his 'Physical View' of personal identity (1990).

One hopes that by building upon these thinker's arguments, and articulating the reasons to believe their views, this thesis highlights clear avenues for future research into personal identity, and – further afield – into discussions of the nature of material objets. In connection with the latter, one chapter – chapter 6 – examines Kit Fine's theory of parthood, and develops a novel objection; I argue that there are objects that Fine's theory cannot accommodate. A paper based on this chapter has recently been accepted for publication at *dialectica*.

Acknowledgements

Mike Martin suggested that I begin my thesis with the joke that I do not know whether what I have written is the same thesis as the one that I set out to write. Fortunately, this is not true; I know that what I have written is not the unwieldy sequence of impressions that I set out to describe. I offer the following list in order to attribute responsibility

First, and foremost, I would like to thank my supervisor, Rory Madden. Rory was my first undergraduate metaphysics lecturer, and introduced me to the problems of personal identity around ten years ago. In the years since, he has been impossibly and unfailingly generous with his time and support – ranging from supervisions that have lasted a couple of hours, to weekly skype check-ins towards the end of the writing process.

I have also spent terms with three other supervisors – Mike Martin, Daniel Rothschild, and Douglas Lavin. Each term has been well spent, and has allowed me to develop and think through ideas related to the topics discussed in this thesis.

My time at UCL has been blessed by a large number of colleagues who are as friendly as they are sharp and engaging. Of these, I would particularly like to thank Showkat Ali, Julian Bacharach, Vanessa Carr, Tony Cheng, Niels Christensen, Nicholas Currie, Catherine Dale, Pete Faulconbridge, Jessica Fischer, Alex Geddes, Androula Kafa, Andrew Knox, Kirstine La Cour, James Laing, Michael Markunas, Rowan Mellor, Polly Mitchell, Oda Ottosen, David Olbrich, Rebecca Rowson, and Anaïs White. Without such kind friends, I think that the process of studying for a PhD would have been lonely and unfulfilling. It has not been so.

I would like to thank Laura Silva, Edgar Phillips, and Karoline Phillips for being my closest friends throughout my time as a graduate student. I have also been tremendously lucky to regularly play board games with some friends from before university – Josh Higgins, and Asher Lewis.

Acknowledgements

I would like to thank my dad, Robert Jansen, my mum, Debra Jansen, and my sister, Caroline Jansen, for their support throughout my time at university. My Oma read and commented on almost all of my undergraduate essays, and has been a source of wisdom, inspiration, and typographic advice throughout my university career. Finally, I would like to thank my partner – Elle Wilkins – for her support, humour, and good grace, in putting up with me through the writing of this thesis.

There are also some to thank who lack the capacity to read these acknowledgements. Stevie, in particular, has kept me sane (despite her best efforts) over the past months. Thomas and Max acted as writing companions for me through my secondary education, up until earlier this year. I also acknowledge the presence of Hamsner and the sticks.

Contents

Contents

Contents

Introduction

This thesis investigates the metaphysics of human persons. As a whole, it develops and tentatively recommends a picture on which we are humanoid entities, but have a privileged proper part – the brain – whose continued functioning is necessary for our continued existence. This picture is similar to those developed by Peter Unger (1990), Mark Johnston (1987; 2007; 2016), and David Wiggins (1967, part 4).[1] It has two components. The first component is the claim that we have such an 'individuating nucleus'. The second is the claim that we have parts other than this nucleus. I investigate these components in order.

The first part of the thesis deals with what I call the persistence question – the question of what is necessary and sufficient for our continued existence. I consider, and reject, one answer to this question – the wide psychological view, on which my continued existence just requires the presence of certain types of psychological connection between my past and future selves, but it does not matter how these connections are realised. I suggest that those drawn to this view should accept that our continued existence requires continuous physical realisation of our mental states. In the case of human persons, this is to say that our persistence requires that of our brains. I divide this part of the thesis into four chapters.

Chapter 1 introduces and contextualises the wide psychological view. I contrast this view with an alternative – the narrow psychological view. Ultimately, I will argue that the narrow view is preferable to the wide one. In chapter 2, I evaluate the dominant methodology for evaluating questions about personal identity, and, in particular, for garnering support for the wide psychological view – the appeal to hypothetical scenarios, or 'thought experiments'. Two questions

---

1. It is worth noting that, though Wiggins initially holds a view similar to mine, he later retracts this view – largely as a result of the discussion in Snowdon 1996 (see also Wiggins 1996). As of Wiggins 2001, ch. 7, he now holds that psychological continuity is neither necessary nor sufficient for personal identity, even if underwritten by the continued existence of the brain.

arise here. First, do we have reason to think that this methodology is, in general, sound? Second, if the methodology is sound, does it support any view about our identity? My contribution to the first question is to argue that we should take thought experiments about personal identity to reveal what kind of entity we believe ourselves to be. These beliefs, in turn, help to determine the referents of our terms for persons (though they do not do so alone). Thus, we ought to take them into account in philosophising about personal identity. Unfortunately, however, our responses to thought experiments do not reveal a settled set of beliefs about our nature; we are inclined to respond to different thought experiments in conflicting ways. Further, our intuitions about many thought experiments change depending upon the way in which they are presented. Since our intuitions do not reflect a stable set of beliefs about our nature, appeals to singular intuitions do little to support the wide psychological view. I consider, and reject, a way to break the impasse by appeal to a hypothetical community who do have stable beliefs in favour of the wide psychological view; there is no reason to believe that there could be such a community, or that we should have the same concept of personal identity as they possess.

Chapter 3 considers an alternative way to support the wide psychological view – an argument presented by Sydney Shoemaker in his 'Personal Identity: A Materialist's Account' (1984). The argument claims that the wide psychological view is entailed by a functionalist theory of mental states. I begin by showing that this is false; functionalism does not commit us to any particular theory of personal identity. Nevertheless, this is not enough to dispel the appeal of Shoemaker's argument; his proposal promises to neatly characterise personal identity using the materials of a functionalist view. The second half of the chapter aims to show that things go awry even here. Shoemaker proposes that we define a notion of 'psychological continuity' by reference to correlations between mental states and behaviours exhibited by subjects at different moments in time. Thus defined, however, psychological continuity can take a branching form; it can hold between a single subject at one time, and multiple subjects at a later moment in time. Thus, psychological continuity can hold between subjects who are provably not identical to one another. What is needed is the introduction of further constraints, detailing exactly when psychological continuity suffices for identity. It is not clear how we should derive these additional constraints. Further, the most plausible suggestion rejects the wide psychological view, holding that our persistence requires the continuous realisation of our psychological

states. I conclude that functionalism gives little support to the wide view.

I turn, in chapter 4, to give further reason to reject the wide psychological view, paying special attention to the version of this view developed by Barry Dainton. The argument again relies upon scenarios in which a person's psychological life branches. Here, however, the branching takes an unfamiliar form, involving frequent separation and recombination of capacities for experience. I do not think it is possible for subjects to have rich psychological lives that are so messily intertwined. This brings into focus the fact that it must be explained, on any view of personal identity, how there could be psychological subjects at all. I propose that the best explanation makes note of the unity imposed upon a subject's mental states (and its difference from other subjects) by a system which realises her existence. In our case, these systems are made from the parts of enduring material entities. Thus, the metaphysically basic case of personal persistence involves the continued existence, and continued functioning, of such a system. Since we have failed to motivate the wide psychological view, there is no reason to think that our persistence can be extended beyond this metaphysically basic case by 'translating' us between material systems.

This concludes my argument against the wide psychological view; we have failed to find reason to motivate this view, and have argued that, if it cannot be motivated, we have reason to reject it. Subsequently, I turn to the second question that this thesis investigates – the question of our composition.

It is appealing to think that we have material parts. Chapter 5 begins by defending this claim against an objection. The objection, which is developed in Lowe 2006, is based on the claim that no material entity is necessary for the existence of each of our mental states. Since we are necessary for the existence of each of our mental states, it follows that we are not material. This argument depends on controversial claims about what conditions are sufficient for the existence of my mental states, and about what is necessary for the existence of various material objects. It can therefore be rejected. I then proceed to elaborate, and defend from criticism, a view that I will reject in chapter 6. According to this view – which I follow Olson in calling 'thinking subject minimalism' (2007, pp. 87–99, 2015b, pp. 49–51) – we are identical to the smallest entities to contain everything that realises our mental states. I note that thinking subject minimalists need not be committed to thinking that there are entities composed of just the realisers of our mental states, and show how this allows them to avoid criticisms raised by Olson.

Introduction

Chapter 6 continues to assess which parts we have. The chapter begins with a challenge. Many entities contain the realisers of my mental states. What reason do I have for identifying myself with one of these entities, rather than any other? I begin by considering possible motivations for thinking subject minimalism. I argue that these motivations fail. I then consider one objection to this view – the claim that nothing brain-sized could sensibly be said to think – and argue that the objection does not succeed. Finally, I present my reasons for thinking that we nevertheless should conclude that we have humanoid form. The argument is based on the fact that many of our mental states can only be ascribed to our brains in virtue of their co-ordination with the perceptual systems of a larger humanoid, on whose behalf the brain thinks. My claim is that we should therefore take this thing to be the 'strict' thinker of our thoughts, and our brains to think (at most) derivatively. Since we do not think derivatively, it follows that we are humanoid. Thus, contra thinking subject minimalism, we are much larger than the smallest entities which contain the realisers of our mental states.

According to the previous chapters, it is plausible to think that it is sufficient for us to continue to exist that our brains continue to exist, and to sustain our mental lives. Yet, I have also argued that we have parts other than our brains. This suggests a desideratum for any account of the metaphysics of material objects; the account should make room for objects whose existence depends on central, unchangeable parts, but which possess other parts aside. I call these 'nucleated wholes'. Chapter 7 brings this to bear upon an extant account of the metaphysics of material objects. I argue that Kit Fine's theory of embodiments is unable to accommodate nucleated wholes. As such, it should either be rejected or revised.

The thesis develops a view of our nature on which our persistence is to be understood in psychological terms, but yet requires the continued existence of some material entity. I have argued that those who believe that our persistence is to be understood in psychological terms should accept this thesis. I do not, however, argue that this thesis is correct, tout court. I finish with some concluding remarks. These remarks summarise the findings of the thesis, and indicate what remains to be proven before it would be reasonable to conclude that the view developed in this thesis is the correct representation of the metaphysics of human persons.

The Two Psychological Views

In the first four chapters of this thesis, I will present and argue against a view about what is required for our continued existence. The view in question is the wide psychological view, according to which our continued existence can be analysed in terms of psychological connections between our past and future selves, without regard to how those connections are realised. I will argue that this view is overly liberal, counting us as surviving scenarios which we should not survive. Before entering into the main body of my argument, I would like to explain and contextualise this view. I begin with the question to which it is a response – the persistence question. This chapter explains how I understand this question, and introduces the two answers I will consider.

1.1   The Persistence Question

To say that something persists is to say that it exists at one time and then continues to exist until a later time.[2] We persist; I existed yesterday, and I exist today. The persistence question asks what is required for a person who exists at some time to have existed at an earlier time, or to continue to exist until a later time.

---

2. A minor quibble here: Some say that for something to exist is just for it to be in our domain of discourse. However, our best logic has a temporally (and modally) invariant domain of discourse. It would seem to follow that 'exists' does not admit of temporal qualification; we cannot sensibly speak of things existing in time, or at times (indeed, Williamson 2013 expands this into an argument that everything that exists exists necessarily). This worry does not affect the point being made; we can substitute the notion of temporal existence with that of being temporally located (which can, in turn, be analysed in terms of present-tensedly bearing intrinsic, non-relational properties). An alternative analysis along similar lines is given by Wiggins 2001, 68–69 (see particularly footnote 9).

Chapter 1: The Two Psychological Views

This question is often expressed using the language of identity.[3] Following suit, I shall take the following to be the canonical form of this question:

PQ Under what conditions is some person x, existing at time t_1, identical to some entity y, existing at time t_2?[4]

The persistence question is almost certainly the most discussed, and the most contentious, question in the personal identity debate. To see what it demands from us, we should begin with the extremely plausible idea that there are constraints upon what an object can survive. Thus, for example, one cannot say that a wooden chair which was burnt on a bonfire last November later existed as a metal lamp constructed from entirely new material. We are similarly constrained in what we can plausibly count as the end of an object's career; given that it has not undergone (intrinsic) change, it would be absurd to say that the writing desk my partner purchased three years ago ceased to exist before today (and is hence distinct from the desk beneath me now).

These constraints provide us with a fairly minimal outline of what is involved in an object's persistence. Some of these pronounce against identifying objects existing at different times. By examining them, we can formulate necessary conditions upon the identification of objects existing at different times (and hence, upon an object's persistence). Other constraints speak against counting

---

3. David Lewis famously argues that questions of persistence are not really questions of identity, but rather questions about something else (in his view, about the principles which determine how an object is composed from its temporal parts). He does so on the basis that there are no real questions about identity:

> Identity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything else except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never any problem about what makes two things identical; two things never can be identical. (D. Lewis 1986, pp. 192–193)

We need not disagree with Lewis in order for the point that I make here to stand. It may be that what I have labelled a question of identity has a more fundamental analysis in which the notion of identity is not used at all; I simply prescind from giving such an analysis. For discussion of the analysis of identity questions, see Fine 2016; Noonan 1989, pp. 105–116; Perry 2002; Wiggins 2001, ch. 3.

4. This corresponds to Olson's 'broad' question: "What are the conditions under which something that is a person at one time is identical with *anything at all* that exists at another time?" (Olson 1997, p. 25). Note that it might be doubted whether all persons are alike in the conditions required for their continued existence. If so, I invite the reader to substitute for the word 'person' the phrase 'human person in possession of a 'normal' human body and 'normal' human intellectual capacities'. I assume that all such entities are alike in their persistence conditions.

certain kinds of event as the end of an object's existence. These can provide us with knowledge of sufficient conditions for the object's continued existence. It is hoped that further reflection will fill in additional details, eventually producing what has been labelled a 'criterion of diachronic identity' – a necessary and sufficient condition for an object's persistence. This tells us what the object's persistence necessarily involves, or, to use a popular turn of phrase, what it 'consists in'.[5] To give a concrete, though completely implausible, example, we might suppose that it is necessary and sufficient for my desk to continue to exist that 90% of its original parts remain attached to one another.[6] This tells us what is involved in any situation in which my desk continues to exist through time.

A criterion of identity needs not only to be necessary and sufficient for an entity's persistence, but it should also be informative.[7] Anti-criterialists, such as Trenton Merricks (1998), deny that there are any such criteria.[8] According to them, we cannot derive informative necessary and sufficient conditions for an object's continued existence. It is difficult to understand quite what is being claimed here. One possibility is that two scenarios may be qualitatively alike whilst differing as to facts about the persistence of the objects involved. This is difficult to countenance; if the view is supposed to have this consequence, I think we should reject it. In elaborating and defending his views, Merricks notes that we do not ordinarily require criteria of identity in order to make judgments about identity over time (p. 108). This strikes me as correct, but what is denied is not an essential element of the view that there are criteria of diachronic identity; the epistemology of identity can, to some extent, be separated from its metaphysics.[9] Though more must be said to properly evaluate anti-criterialism,

5. See, for example, Noonan 1989, pp. 104–116; Perry 2002; Wiggins 2001, pp. 56–57; 59–61 for more on the relationship between criteria of identity and persistence through time.

6. This stripped down example is artificially precise. Real examples of criteria of identity need not be so neat. I find it very plausible that there may be irresolvable vagueness in our formulations of criteria of diachronic identity, though it is controversial just what this vagueness consists in. For discussion of vagueness in general, see especially Williamson 1994; Barnes 2010; Fine 1975. For discussion of vagueness applied to identity, see Evans 1978; Wiggins 2001; Parsons 2000, pp. 161–171.

7. One might think that for a criterion of identity to be informative, it must not be circular. However, Shoemaker argues that accounts of persistence conditions often make use of an unobjectionable form of circularity (c.f. Shoemaker 1984; Shoemaker 1979, pp. 98–101; 2012). I examine Shoemaker's views in chapter 3.

8. The articles in Gasser and Stefan 2012 provide a good range of perspectives on the closely related 'simple' view of personal identity.

9. See Johnston 2007, pp. 67–68; Wiggins 2001, pp. 78; 82–84 for two different ways to make this point.

Chapter 1: The Two Psychological Views

I shall assume that there are criteria of identity, though they may be impossible to state without considerable vagueness.

To answer PQ, we must provide a criterion of identity for persons. Various proposals exist. For example, the body view has it that it is necessary and sufficient for a person to remain in existence that her body remains in existence.[10] The biological approach, in contrast, holds that one's existence depends upon continued lower-level biological activity.[11] Though these responses are, I think, in good standing, I shall not discuss their merits in this thesis, or compare them with the view I suggest that we adopt. Instead, I will focus upon different versions of a widely accepted third alternative – the psychological view. The first part of the thesis considers the best way to formulate this view.

1.1.1   The Psychological View

From Locke's contribution to the debate up until at least the critical remarks of Bernard Williams, it was usual to assume that our persistence conditions could be glossed in psychological terms. Call this the 'psychological view' of our persistence. Despite lengthy philosophical discussion, Bourget and Chalmers indicate that the psychological view remains the dominant belief among professional philosophers (2014, p. 476).[12] For this reason, the view merits close investigation. I begin by setting out what it claims.

At a first pass, we can say that proponents of the psychological view claim that one's persistence 'consists in' the holding of certain types of psychological connection; it is necessary and sufficient for one's continued existence that some psychological relation hold between one's mental states at different times.[13]

---

10. C.f. D. Mackie 1999b; Williams 1956–1957; Williams 1973; Thomson 1997. An alternative version of the body view might be to hold that it is necessary not only that one's body remain in existence, but also that it remain the body of a person (i.e. that it be associated with a conscious subject) for so long as one exists.

11. See, for example, Johnston 1984, pp. 137–139; Olson 1997 for statements of this view. See Snowdon 2014a for an objection to this view of animal persistence. Also relevant is the exchange between Hershenov and LaPorte over D. Hershenov 2005; D. Hershenov 2009; LaPorte 2009.

12. It is worth noting, however, that an even greater number of respondents (37.3% vs. 33.6%) rejected all responses listed, selecting the 'other' category.

13. This is only intended to be a first pass – there are tweaks discussed later, such as Nozick's closest continuer view (1981, ch. 1), on which psychological connections between different phases of one's existence are not necessary for one's persistence. See also Langford 2014 for a view which denies that psychological connections are necessary for one's persistence. These 'hybrid' views are complex, and I do not think that it is a problem that they do not immediately fit into a characterisation of the psychological view.

§1.1: The Persistence Question

There are, however, various ways to develop an account of the relevant psychological connections. Let us begin with Locke's account, from which all versions of the psychological view take inspiration. Locke claims that "as far as [a person's] consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person" (1689/1975, Book II, Ch. 27, §9/p. 335). One might thereby interpret him as espousing the following criterion of identity:

Lockean A person x, existing at time t_1 is identical to some entity y, existing before time t_1 if and only if x can (at t_1) remember y's actions and experiences.[14]

Needless to say, this criterion can be reversed to give a condition of identity with a future person: a person x existing at t_1 is identical to some entity existing at a later time if and only if that entity can (at that later time) remember x's actions.

Locke's version of the psychological view fleshes out the notion of a 'psychological relation' between mental states existing at different times by reference to 'direct' memory connections which obtain between a person at one time and that person's earlier actions and experiences. It is well known that Locke's criterion is untenable in this form.[15] However, this has not stopped others from pursuing

---

14. This involves taking consciousness of a past action to require memory. This interpretation seems to me strongly supported by the following passage:

> [A]s far as any intelligent being *can* repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self. (1689/1975, Book II, Ch. 27, §11)

It seems fairly plain that memory is at least a necessary condition for Lockean 'repetition'. Marya Schechtman stresses that Lockean consciousness also contains an affective component (2001) which is lacking from most contemporary discussions of the topic (for a noteworthy exception, see Wollheim 1979, pp. 211–224; 1984, ch. 4). This affective requirement is also picked up on by Galen Strawson (G. Strawson 2011, ch. 9).

15. The decisive objection is ably communicated by Thomas Reid:

> Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.
>
> These things being supposed, it follows, from Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who

Chapter 1: The Two Psychological Views

alternative forms of the psychological approach. A few examples are as follows:

Updated Memory A person x, existing at time t_1, is identical to some entity y, existing before time t_1 if and only if either (a) x can (at t_1) remember y's actions and experiences; or (b) there is a w, existing before t_1, such that x can (at t_1) remember w's actions and experiences, and w can remember y's actions and experiences; or (c) there is a w and a v, both existing before t_1, such that x can (at t_1) remember w's actions and experiences, and w can remember v's actions and experiences, and v can remember x's actions and experiences; or ... [16]

Psychological Continuity A person x, existing at time t_1, is identical to some entity y, existing before time t_1 if and only if x's psychological characteristics (including her beliefs, personality traits, intentions, and actions) are appropriately causally linked to y's psychological characteristics.[17]

Experiential Continuity A person x, existing at time t_1 is identical to some entity y, existing before time t_1 if and only if x's experiences and y's experiences fall (or would fall, were y conscious) within the same 'stream of experience'.[18]

I have intentionally kept these characterisations somewhat loose and nontechnical; more detail will be added later where necessary. For now, however, I want to distinguish two versions of the psychological view. The distinction I have in mind can be raised without endorsing any particular psychological relation as relevant to our continued existence.

---

took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging therefore, according to Mr. Locke's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school. (Reid 1785/1941, Essay III, Ch. 6/pp. 213–214)

16. C.f. Grice 1941; Shoemaker 1970; Wiggins 1974.

17. C.f. Parfit 1984; Shoemaker 1984; Quinton 1962; D. Lewis 1976. Due primarily to Parfit's influence, this is now the dominant form of the psychological approach.

18. C.f. Foster 1979; Dainton and Baine 2005; Dainton 2008. Unger 1990 holds that what is required for personal persistence is preservation of one's 'core psychology', by which he means "my capacity for conscious experience, my capacity to reason at least in a rudimentary way, and my capacity to form some simple intentions" (p. 68). Of the views I have described, Unger's is closest to the experiential continuity view.

1.1.2 The Wide and Narrow Views

The distinction I have in mind was originally raised by Parfit:

> There are three versions of the Psychological Criterion. These differ
> over the question of what is the *right* kind of cause. On the *Narrow*
> version, this must be the *normal* cause. On the *Wide* version, this
> could be *any reliable* cause. On the *Widest* version, the cause could be
> *any* cause. (1984, p. 207)[19]

Parfit holds that personal identity requires the presence of appropriately
caused psychological connections between a person's mental states at subse-
quent stages of her existence.[20] I would like to abstract from this specific version
of the psychological view; what is significant is that the above quotation high-
lights three different views about the 'appropriate' means of realising the psy-
chological connections relevant to our persistence. One more element of Parfit's
characterisation should also be set aside. If, like Parfit, we are to distinguish the
wide and the widest versions of the psychological approach, we should need to
identify different kinds of cause, such that it can then be established whether or
not these causes are reliable means of ensuring psychological continuity.[21] It is

---

19. It is not entirely obvious how to interpret the distinction Parfit draws here. For excellent dis-
cussions of this passage, see S. Campbell 2005, pp. 376–379 and Sidelle 2011. Campbell continues
to argue that we should consider an even wider view, on which no cause whatsoever is necessary
for what matters in survival; see Elliot 1991 and Kolak and R. Martin 1987 for similar arguments
to the same conclusion.

20. More precisely, Parfit defines personal identity via the following two notions:

Strong Connectedness$_P$ There is strong connectedness between x at t_1 and y at t_2 if and only
if the number of psychological connections between them is "*at least half* the number of
direct connections that hold, over every day, in the lives of nearly every actual person"
(Parfit 1984, p. 206).

Psychological Continuity$_P$ x at t_1 and y at t_2 are psychologically continuous if and only if they
can be connected to one another by the ancestral of the strong connectedness relation (i.e.
they participate in an overlapping chain of strong connectedness) (ibid.).

With this defined, he can give the following criterion of identity for persons:

Identity$_P$ "x today is one and the same person as y at some past time if and only if (1) x is
psychologically continuous with y, (2) this continuity has the right kind of cause, and
(3) there does not exist a different person who is also psychologically continuous with y"
(ibid., p. 207)

21. To see what is meant here, imagine an operation aimed at 'bringing historical figures back to
life' by creating present-day replicas in clones of their original bodies. Imagine that the amount of
information about different figures differs; we know enough to create a fairly close likeness of Guy
Fawkes (an example familiar from Williams 1956–1957), but much more guesswork is involved
in replicating Genghis Khan. According to one typing, the operation utilises a single process

not entirely obvious how one might do this. Consequently, it has become normal to speak only of the wide and narrow views, where the wide view encompasses any account on which identity does not require the normal cause, whatever that should be. I shall follow suit. The aim of the first part of the thesis is to argue that we should prefer the narrow view to the wide view.

To differentiate the wide and the narrow view, we need to agree on what counts as the 'normal' cause of psychological connections between our mental states. The usual way to do so is by reference to the fact that personal identity usually involves the preservation of one's brain:

> The normal causes of memory involve the continued existence of the brain. And some or all of our psychological features depend upon states or events in our brain. The continued existence of a person's brain is at least part of the normal cause of psychological continuity. (Parfit 1984, p. 208)

If we were to pursue this thought, the narrow view would say that we can survive through changes only if our brain continues to exist, and continues to be responsible for our psychological lives. I would like to abstract away from the requirement that my brain persist in order to avoid discussion of at least two cases about which I would like to remain neutral:

Inorganic Replacement Suppose that, very gradually and over much time, thousandths of my brain are sequentially replaced by suitable bionic structures. The parts replaced are killed. Originally there will be a brain that is entirely organic; finally, there will be only a brain that is entirely bionic. During the brief periods when replacements are made, there will be 99.9% of a brain. Between replacements, of course, there will be 100% of a brain. Except at the first and the last steps, it will be an integrated brain. When there is only 99.9% present, the least there ever is, there will be a slight loss as regards my distinctive psychology. But, even at those times, there will be no loss of core psychology. (Unger 1990, p. 122)

---

which is unreliable because it frequently requires guesswork. According to another. According to another typing, the recreation of someone based on extensive (and reliable) historical information involves a different kind of cause than their recreation in accordance with guesswork. The former typing has it that the operation never succeeds in bringing the original persons back to life, whilst the latter allows that it is sometimes successful.

§1.1: The Persistence Question

Zippering and Separation Suppose that, while my brain is stimulated to support conscious experience throughout, the left and the right half of my brain are each fitted at their interface, slowly and gradually, with radio communicators, or transceivers . . . These transceivers will at first be only a fancier, more indirect means of having my neural impulses proceed just as they normally would, without the presence of the devices. We may suppose that, though the total physical explanation is more complex, my mental processes will proceed just as they would normally . . . We next separate each of the people into three spatially distant parts. During my separation, my three parts continue communicating in ways that are, in relevant respects, quite normal. Because of this, I continue to have the same conscious experience. At the end of the separation, I am in three rooms. (ibid., pp. 177–178)

I do not know how to resolve whether or not my brain would survive the changes described, rather than being replaced by an inorganic replica, or destroyed and scattered. I also do not have particularly strong feelings about whether or not I could survive the changes described, and I do not know how one could settle the question. However, these thought experiments have no bearing on the arguments that I present against the wide psychological view. As such, it will be useful to adopt a slightly wider characterisation of the narrow view than is usual, on which the narrow view is consistent with one's survival of Inorganic Replacement or Zippering and Separation.

As I characterise it, the narrow view should be formulated as follows, adapted from Unger's *Identity, Consciousness and Value* (1990, p. 109):

Narrow View A person x, existing at time t_1, is identical to some entity y, existing at t_2, only if (a) x and y stand in some particular psychological relationship R; and (b) there is sufficiently continuous physical realisation of x's psychological states, or of her capacities to have those states, at times between t_1 and t_2, culminating in the realisers of y's mental states, or of her capacities to have those states.[22]

---

22. C.f. also the constraint Agar imposes on psychological accounts of personal identity:

> A necessary condition for two states to belong to the same person that it is possible to trace a chain of actual and potential causes between them that includes no non-

Chapter 1: The Two Psychological Views

What is meant by this is simple. Just as ordinary material objects must be composed by much the same material from moment to moment, so too must we; if there is to be a complete change in the material which realises our mental states, this must happen gradually, rather than all at once. Proponents of the wide psychological view deny this constraint. According to them, facts about the material realisation of our mental states are irrelevant to our judgments about personal identity. We can represent their view as follows:

Wide View A person x, existing at time t_1, is identical to some entity y, existing at t_2, only if (a) x and y stand in some particular psychological relationship R; where (b) this does not require the continuous physical realisation of x's psychological states, or of her capacities to realise those states.

If the wide psychological view is correct, it is at least logically possible for some entity to realise one's mental states at one time, and for another entity, made from completely new matter, to realise one's mental states shortly afterwards. I will argue that we have reason to reject this view.

I want to make four further comments about these two views. First, note that, as formulated, both elaborate necessary conditions upon one's survival, but are silent about whether these conditions are also sufficient for one's continued existence. This is because certain kinds of logical difficulty, covered in chapter 3, may necessitate adding further necessary conditions to one's analysis of personal identity. Many options are available, and I do not want to complicate discussion by considering them here.

Second, it should be noted that the narrow view is rather vaguely characterised; it says that our survival requires 'sufficiently continuous' realisation of our mental states, but does not tell us precisely what this involves.[23] I assume that a more detailed account would need to be developed on a case by case basis. I also assume that there are limits to the amount of precision to which one can

---

psychological intermediate. (Agar 2003, p. 62)

As I point out in chapter 3, it is reasonable to believe that this requires the continuous realisation of our psychological states.

23. I leave open, for example, that a subject's mental states could be sufficiently continuously realised in a 'gappy' way, allowing for her brain to be separated into largish chunks, and then later put together again (c.f. the discussion of systemic energy on pp. 129–130 of Unger 1990). This is another case about which I wish to remain neutral.

aspire in developing the view. However, for our purposes, this need not matter; when discussing the wide view, we will focus on cases in which our survival is possible despite the complete absence of physical continuity in the realisers of our mental states.

Third, the Narrow Constraint, as written, is committed to the idea that our mental states are physically realised. This excludes the view that our mental states are realised in immaterial entities (or collections of immaterial entities). It will be a basic assumption of this thesis that our mental states are wholly physically realised. However, were this false, and our mental states to be realised in immaterial entities, I presume that we could still define an analogous type of continuity in the realisers of our mental states.[24] Indeed, as we shall see, the earliest proponent of the wide view – John Locke – developed his account precisely as an alternative to the claim that we must be constituted by the same immaterial thinking substances whenever we exist.

Fourth, I speak about the continuous realisation of the capacities for our mental states. I do so in order to accommodate cases such as the following:

Super Freezing A super freezing machine will stop a person's molecules in

24. To generalise the wide and the narrow views, we need only introduce the following two technical notions:

Mereological Overlap An entity x mereologically overlaps at t_1 with an entity y at t_2, if and only if something is a proper part of x at t_1 and y at t_2.

Mereological Continuity An entity x is mereologically continuous at t_1 with an entity y at t_2 if and only if either (a) x mereologically overlaps at t_1 with y at t_2; or (b) there is a temporally ordered series σ, consisting of objects picked out at one or more times intermediary between t_1 and t_2 such that:

(i) The earliest member of σ mereologically overlaps with x at t_1 if t_1 is earlier than t_2, or with y at t_2 if t_2 is earlier than t_1.

(ii) The latest member of σ mereologically overlaps with x at t_1 if t_1 is later than t_2, or with y at t_2 if t_2 is later than t_1.

(iii) Each member of σ mereologically overlaps with any member of σ which is adjacent to it (i.e. either closer to it in the past than any other member of σ, or closer to it in the future than any other member of σ).

Armed with these notions, we can define the wide and narrow views as follows:

Wide View A person x, existing at time t_1, is identical to some entity y, existing at time t_2, only if (a) x and y stand in some particular psychological relationship R; where (b) this does not require mereological continuity.

Narrow View A person x, existing at time t_1, is identical to some entity y, existing at time t_2, only if (a) x and y stand in some particular psychological relationship R; and (b) the realisers of x's mental states at t_1 are sufficiently mereologically continuous with the realisers of y's mental states at t_2.

> their tracks, so to say, almost perfectly preserving their relative arrange-
> ment for minutes, or for years. Almost instantaneously, it reduces the
> temperature of a person's body to within a very minute fraction of one
> degree Kelvin of absolute zero, and keeps it there until it operates in re-
> verse. Later, the device may operate as a super thawing machine, raising
> the temperature of that body, virtually instantaneously, to normal body
> temperature . . . After the thawing, experience, thought and behavior will
> take up where, right before the freezing, they left off. (Unger 1990, pp. 3–4)

My reaction to Super Freezing, which I take to be the dominant intuition, is to think that were the technology medically possible, it would be possible for one to survive being frozen and subsequently thawed. In the interim, I do not think that one's super-frozen brain would realise mental states. Thus, continuous physical realisation of one's psychological states is not required for one's continuous existence. However, I think that one's super-frozen brain would retain the capacity to realise conscious states. I have formulated the narrow view so as to be compatible with our survival of Super Freezing.

The wide and the narrow psychological views agree that our persistence is to be understood in psychological terms; for us to continue to exist, there must exist psychological relations of the right kinds between our future psychological states, and the states that we now possess. Where these views disagree is about how these psychological connections must be realised. The narrow view holds that our continued existence requires the physically continuous realisation of our capacities to enjoy psychological states. The wide view maintains that this is not a requirement on our continued existence. The following three chapters will argue for the conditional claim that if one wishes to uphold a version of the psychological view, one should prefer the narrow view to the wide view. I begin, in the next chapter, by considering a methodology often employed to support the wide view – the use of thought experiments. I argue that this methodology fails to show that we are inclined to believe the wide psychological view.

Persisting with Thought Experiments

According to the wide psychological view, our persistence consists in certain types of psychological connection between our mental states, no matter how these connections are realised. In particular, proponents of this view claim that our continued existence does not require physical continuity. In this chapter, I consider the usual method of generating support for this view – appeals to our judgments about hypothetical cases. Section 2.1 gives three examples of such cases. In 2.2, I consider two objections to this methodology. The first objection fails, but the second presses us to explain why our intuitions should count as evidence for the wide psychological view. Section 2.3 takes up this challenge. I dismiss one response – that these judgments show that the wide psychological view has an epistemic advantage over its rivals – and argue instead that our beliefs about personal identity, as revealed by thought experiments, are likely to play a role in determining the referents of our terms for persons. Unfortunately, however, 2.4 argues that it is unlikely that we have settled beliefs in favour of the wide psychological view. I present two obstacles. First, intuitions in favour of this view seem to conflict with the results of other thought experiments. Second, one and the same hypothetical scenario may generate different responses depending on how it is framed. Section 2.5 ends the chapter by considering a thought experiment of Shoemaker's which attempts to break the impasse, and to show that our concept of personal identity lines up with the wide view. The argument fails because it assumes too readily that it is acceptable to individuate persons as the wide psychological view requires. Thus, I conclude that appeals to thought experiments do little to motivate the wide psychological view.

Chapter 2: Persisting with Thought Experiments

2.1   Intuition and the Wide Psychological View

One might accept the wide psychological view for a variety of reasons. Historically, two such reasons have been attempts to make sense of the possibility of life after death,[25] and the desire to build an account of personal identity using only the supposedly more straightforward 'given' materials of individual experiences, and the relationships between them.[26] These motivations hold little sway today (though I will connect the discussion of chapter 4 to the second of them). Within the contemporary literature, one is more likely to find the wide psychological view supported by appeal to thought experiments – hypothetical, as yet unrealised, scenarios. It is claimed that our intuitive judgments about these scenarios line up with the wide psychological view, and that this provides us with reason to believe that this view is correct. I will give three examples of scenarios which have been thought to support the wide psychological view.

We are prima facie inclined to judge that someone could survive the following scenarios:

Teletransportation The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. (Parfit 1984, p. 199)[27]

Upload [Y]our stream of consciousness is, in effect, lifted from your brain and seamlessly joined to a stream of consciousness that is machine-produced. Your brain, meanwhile, lapses into a profound coma-like condition. (Dainton 2008, pp. 18–19)[28]

---

25. This motivation was certainly in the background of Locke's discussion of personal identity, asthough he talks explicitly about life after death, and judgment day, on only a few occasions (Locke 1689/1975, §§13; 22; 26). For more recent discussion of personal identity and life after death, see Penelhum 1970 and Quinton 1962, pp. 400–401; 407–409.

26. Wollheim suggests, quite plausibly, that psychological analyses of the self received such easy assent through the first part of the 20th century because they promise 'metaphysical economies' of a kind attractive to those writing in the wake of British empiricism, allowing the existence of persons to be understood in terms of (at the time supposed) simpler facts about mental states and the relations between them (Wollheim 1979, pp. 189–191).

27. This is a common trope in science fiction – `https://tvtropes.org/pmwiki/pmwiki.php/Main/Teleportation` gives a list of examples.

28. Again, there are many depictions of this phenomenon in popular fiction. See, for example,

Change of Soul [I]f the same consciousness (which, as has been shewn, is quite
a different thing from the same numerical figure or motion in body) can be
transferr'd from one thinking substance to another, it will be possible, that
two thinking substances may make but one person. (Locke 1689/1975,
Book II, ch. 27, §13/p. 338)[29]

Each of these scenarios faces a person with a complete change in the objects
which constitute her; in Teletransportation and Upload, she comes to be consti-
tuted by a new material entity, in Change of Soul, she is constituted by different
immaterial substances at the beginning and the end of the hypothesized sce-
nario. People do not ordinarily undergo such changes. Hence, were personal
identity to require the normal cause of psychological continuity, the imagined
scenarios would be unsurvivable. These cases therefore directly test whether
we are (sometimes) willing to make judgments in accordance with the wide
psychological view. They confirm this hypothesis. The next section explains
two challenges to taking these responses to count in favour of the view.

2.2 Two Objections to the use of Thought Experiments

Although thought experiments are often employed in philosophy, their use in
discussions of personal identity is sometimes thought especially problematic.[30]
At least two factors have influenced this judgment. First, discussions of per-

---

https://tvtropes.org/pmwiki/pmwiki.php/Main/BrainUploading. O'Connell 2017,
pp. 42–69 gives a fascinating account of time spent with some prominent transhumanists whose
ambition it is to produce technology capable of uploading persons onto computers.

29. Kant presents an evocative take on much the same thought experiment:

An elastic ball that strikes another one in a straight line communicates to the lat-
ter its whole motion, hence its whole state (if one looks only at their positions in
space). Now assuming substances, on the analogy with such bodies, in which repre-
sentations, together with consciousness of them, flow from one to another, a whole
series of these substances may be thought, of which the first would communicate its
state, together with its consciousness, to the second, which would communicate its
own state, together with that of the previous substance, to a third substance, and this
in turn would share the states of all previous ones, together with their conscious-
ness and its own. The last substance would thus be conscious of all the states of all
the previously altered substances as its own states, because these states would have
been carried over to it, together with the consciousness of them; and in spite of this
it would not have been the very same person in all these states. (Kant 1781/1998,
A363–364/p. 423)

30. See, for example, Johnston 1987a; Johnston 2007, pp, 92–99; Wilkes 1988, Ch. 1. Wilkes' and
Johnston's claims are critically assessed in Snowdon 1991.

sonal identity frequently make use of a striking range of fantastical thought experiments. It is far from clear that our judgments about such cases will be reliable. Second, it has been thought – I suspect unfairly – that discussions of personal identity rely upon thought experiments to a greater extent than philosophical discussion elsewhere.[31] These two factors might be thought to leave us with very few materials by which to evaluate whether our intuitive responses to these cases should be trusted. I want to begin by considering this objection, arguing that it fails.

The lynchpin of the objection is the fear that we are likely to be confused or befuddled by unfamiliar thought experiments. Wilkes gives one justification for this concern when she writes:

> The experimenter – any experimenter, in thought or in actuality – needs to give us the background conditions against which he sets his experiment. If he does not, the results of his experiment will be inconclusive. The reason for that is simple and obvious: experiments, typically, set out to show what difference some factor makes; in order to test this, other relevant conditions must be held constant, and the problematic factor juggled against that constant background. If several factors were all fluctuating, then we would not know which of them (or which combination of them) to hold responsible for the outcome. (Wilkes 1988, p. 7)

She continues to argue that thought experiments in the personal identity debate suffer from "ambiguous uncertainty concerning the[ir] relevant background conditions" (ibid., p. 8), and – as such – are poor tools for philosophical investigation. Since we are in the dark as to what would change and what would remain the same were these scenarios actual, our judgments about what has happened do not have an adequate basis.[32] This worry is exacerbated by the thought that discussions of personal identity rely almost solely upon thought experiments; if we rely on thought experiments alone, we cannot check our intuitions except by relying upon other thought experiments. Thus, one has no external measure of the validity of our intuitive judgments, and no reason to think that they are trustworthy.

---

31. This criticism is more often voiced than found in print. One printed example is found on p. 6 of Wilkes 1988.

32. In connection with this, see Wilson 2016.

§2.2: Two Objections to the use of Thought Experiments

My response to broad-scale worries of this kind is rather simple. Parfit puts it like this:

> [The criticism that we have little to learn from thought experiments] might be justified if, when considering such imagined cases, we had no reactions. But these cases arouse in most of us strong beliefs. And these are beliefs, not about our words, but about ourselves. By considering these cases, we discover what we believe to be involved in our own continued existence, or what it is that makes us now and ourselves next year the same people. We discover our beliefs about the nature of personal identity over time. (Parfit 1984, p. 200)

Not all thought experiments generate firm reactions. For example, I am inclined to be unsure what to say about the case of Gyges ring – a case which Wilkes cites as involving ambiguous background circumstances (Wilkes 1988, p. 11).[33] However, when a thought experiment generates a strong intuitive reaction, rather than sheer uncertainty, this illustrates that we take ourselves to know enough about the imagined scenario to make a judgment. This is to say that we take ourselves to grasp what is envisaged, and that we thereby presume ourselves to have a fix on all relevant variables. If this is correct, then we can bypass ambiguity concerning a hypothetical scenario by describing more fully how we envisage it – paying special attention to what features of the scenario drew our attention – and why we took this to license a particular conclusion. There is no reason to think that we will be unable to say what leads us to make the firm judgments that we do. Hence, there seems no reason to concede that we will be utterly stymied by ambiguities in our description of a given thought experiment.

This leads to a further point. If it is possible to unpack our reasons for making a given judgment, it should also be possible to scrutinise prospective justifications for our intuitions, and to consider what bearing they have on an account of our persistence. Of course, our first impressions may not always survive closer examination, and they may be distorted by a variety of factors.[34] How-

33. Indeed, I am inclined to be unsure about this case in part because I have little idea quite how to envisage what is going on.

34. Here, we should particularly heed Bernard Williams' warning against artificially neat constructions of thought experiments intended "to produce a situation which would naturally elicit, with minimum hesitation, [a desired intuitive response]" (Williams 1970, p. 179). C.f. also the discussion of thought experiments in Unger 1990, pp. 84–87.

ever, this just cements the point at hand; it is often possible to establish our reasons for making an intuitive judgment (and, indeed, what assumptions we make about unstated 'background' elements of the thought experiment), and, having done so, to subject them to assessment.

I think that there are two further mistakes involved in the complaint that thought experiments are not subject to external checks and balances. First, its proponents are insufficiently imaginative; in providing an account of personal identity, we can fall back upon a great number of beliefs about the persistence of subjects through real-life circumstances (and about our ability to gain knowledge of their persistence).[35] Unless sufficiently strong reason is given to revise these beliefs, it is a requirement on any account of personal identity that it validate them. Thus, there are at least some external constraints upon the conclusions that we can draw from thought experiments. Second, even when these external checks fall silent, we may compare intuitions drawn from a variety of thought experiments against one another. I conclude that the objection under examination overlooks key ways in which we may interrogate our responses to thought experiments. As such, it should be set aside.

Though this objection fails, there is room for a more subtle 'sceptical' objection to the use of thought experiments in the personal identity debate. Note that Parfit endorses thought experiments because of their role in uncovering our *beliefs* about personal identity. To this, the following objection seems apt:

> Our intuitions about cases of personal survival, be those cases real or imaginary, are just beliefs we have, and they may well be false. (Johnston 2016, p. 95)

Put otherwise, if we are to draw conclusions from our intuitive responses, we need reason to believe that they are reliable. But it is far from clear how we could establish this, at least without being in a position to check them against the correct answer to the persistence question. The following section considers two ways to justify the claim that our intuitions in favour of the wide psychological view are reliable.

---

35. I am, of course, drawing on Johnston's proposed methodology for theorizing about personal identity – reflection on our ability to "reliably and unproblematically identify ourselves and others" (Johnston 1987a, p. 63). I am suggesting that this need not be an alternative to the use of thought experiments, but can instead be used to complement them.

2.3   From Thought Experiments to Reference

This section considers two ways to argue that our intuitions show the wide psychological view to be correct. The first route argues that our intuitions reflect what would be the only reasonable ways to make judgments in hypothetical scenarios, given the evidence that would then be available to us. Thus, accepting the wide psychological view explains how we might make knowledgeable judgments about our persistence, whereas rejecting it calls into question our ability to do so. I will argue that this response is difficult to believe, and that it rests on a controversial epistemological view. That said, I consider another response, on which our judgments about personal identity play a role in determining which entities we refer to when we talk about persons. This response is more promising, though it has some controversial underlying structure.

2.3.1   The Epistemic Response

The first way to buttress our intuitions in favour of the wide psychological view begins with the role that 'consciousness' (or memory) ordinarily plays in our knowledge that we have persisted through some period of time – justifying our beliefs about the past without requiring corroborating evidence.[36] Thus, for example, Locke writes:

> Had I the same consciousness that I saw the ark and Noah's flood, as that I saw an overflowing of the Thames last winter, or as that I write now, I could no more doubt that I who write this now, that saw the Thames overflowed last winter, and that view'd the flood at the general deluge, was the same *self*, place that self in what substance you please than that I who write this am the same myself now whilst I write (whether I consist of all the same substance material or immaterial, or no) that I was yesterday. (Locke 1689/1975, Book II, ch. 27, §16/pp. 340–341)

This passage evaluates the Change of Soul thought experiment (presented on p. 35). As I read it, the starting point of this thought experiment is the claim that memory provides a secure basis for our judgments about identity over time; I can know that I carried out some action just by remembering doing so. In con-

---

36. In this context, see Shoemaker 1970. For criticism of this kind of strategy, see McDowell 1997, who argues that it reflects a 'Cartesian mistake' of holding that the materials for an account of personal identity (and for our knowledge of it) are provided from within consciousness itself.

trast, memory does not provide a secure basis for judgments about the identity of immaterial thinking substances:

> [C]onsciousness being interrupted always by forgetfulness, there being no moment of our lives wherein we have the whole train of all our past actions before our eyes in one view, . . . and we losing the sight of our past selves, doubts are raised whether we are the same thinking thing, i.e. the same *substance* or no.   (Locke 1689/1975, §10/pp. 335–336)

This passage argues that we cannot know that one and the same soul was present for every event that we can remember. Hence, were personal identity to require that we always possess the same soul, we would be unable to know about our persistence on the basis of memory. This contradicts the claim that memory provides a secure basis for judgments about our persistence. We should therefore conclude that sameness of soul is not required for personal persistence, in line with the wide psychological view.[37]

Let's now evaluate the argument. The first thing to note is that it is incomplete if presented as an argument for the wide psychological view. If Locke's argument succeeds, it shows that sameness of soul is not required for personal identity. However, it does not rule out the existence of further requirements upon our persistence, such as preservation of the same body or brain. Thankfully, it is not difficult to find other statements to fill this gap. For example, though he rejects the wide psychological view, Johnston provides an argument which would also be available to its proponent when he says:

> Once it is admitted that among the necessary conditions on personal identity are certain bodily conditions having to do with the survival of an organism or crucial parts of it, the question arises how experiential memory, i.e., something whose internal phenomenology makes it seem like a faculty suited to picking up only mental connections between earlier and later mental states, could deliver any more than conclusions of the form: the same *mind* that had the remembered experience is the mind that is now remembering. And if it could not do this, by what right are its deliverances properly taken to involve claims about personal identity as opposed to claims about mere mental identity? (Johnston 1987a, p. 77)[38]

---

37. Martin and Barresi attribute this argument to Locke on p. 143 of their 2006.

38. Anscombe's sensory deprivation thought experiment is of particular relevance here:

§2.3: From Thought Experiments to Reference

Again, the suggestion is that memory is unable to discern whether or not some condition holds which is considered necessary for personal identity. Since it is to be supposed that memory is a reliable faculty for making judgments about personal identity, we should reject the claim that this condition is necessary.

We can formulate the argument for the wide psychological view as follows:

1. We can know about our persistence (on the basis of memory).

2. If personal identity requires psychological continuity with its usual cause, then we can know about our persistence (on the basis of memory) only if we can establish the presence of this usual cause throughout what we take to be our careers.

3. We cannot establish the presence of a usual cause throughout what we take to be our careers.

C. Hence, personal identity does not require psychological continuity with its usual cause.

This argument is valid. However, the second premise is extremely questionable.[39] It seems to rely on the principle that one can know a fact only if one is in a position to show that any necessary conditions for this fact obtain – including ruling out any possible defeaters of one's claim to know this fact. This is a demanding epistemological premise, and entails the following conditionals:

2a. I can only know that the world has existed for more than five minutes if I can establish that it was not created five minutes ago, leaving us with apparent memories that it existed beforehand.

And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other. Now I tell myself "I won't let this happen again!" If the object meant by "I" is this body, this human being, then in these circumstances it won't be present to my senses; and how else can it be 'present to' me But have I lost what I mean by "I"? ...I have not lost my 'self- consciousness'; nor can what I mean by "I" be an object no longer present to me. (Anscombe 1975, p. 58)

Anscombe concludes that 'I' is not a referring device. A less radical conclusion, though, would be that information about one's body is not required for self-reference because sameness of body is not required for personal identity.

39. I have already cited two thinkers – John McDowell and Gareth Evans – who are extremely sceptical of this kind of argument. In this context, one might also see Snowdon's criticisms of Johnston's argument on pp. 214–215 of his 2014b.

2b. I can only know that I have hands if I know that I am not a handless brain in a vat.

2c. I can only know that this sample of yellowish material is gold if I can establish that it contains atoms with atomic number 79.

2d. I can only know that I bought the table before me last year if I can establish that the table I bought has not been, unbeknownst to me, removed and replaced with a duplicate.

In each of these cases, I assume that we cannot establish that the supposed necessary condition holds. Thus, the most likely motivation for the second premise of the above argument leads to an unattractive, and quite general, scepticism. Because of these unpalatable consequences, the underlying principle has been widely rejected by contemporary epistemologists, who have given a variety of less demanding accounts of what is required for knowledge.[40] Consequently, we may reject this principle, and, with it, the second premise of the above argument.

Though the second premise is the main bone of contention in the above argument, it is worth noting that, once we accept a more relaxed epistemological view, the third premise also seems open to doubt. Take first the view that our persistence requires the continued existence of our bodies, or of our brain. We standardly have very good evidence that our bodies and brains continue to exist; our bodies are before us from the moment we are awake to the moment that we are asleep, and we can usually observe that they have not been tampered with (and so, infer that our brains have not been removed from them and replaced with duplicates). We certainly have at least as good evidence of the continued existence of our bodies and brains as we do of that of other material objects around us. Given that we take ourselves to be able to make knowledgeable judgments about the identity of these items, and to do so on the basis of memory, it is hard to sustain the objection that our memory does not establish that our brains have not been replaced by duplicates. The argument does not establish that we must accept the wide psychological view in order to account for the fact that we may know about our own persistence on the basis of memory.

---

40. Of particular relevance here is Pryor 2000. Pryor argues that one can know something without being able to offer a non-question begging argument that defeaters for that knowledge do not obtain.

§2.3: From Thought Experiments to Reference

It is instructive to note that our relaxed epistemological requirements may prevent the above argument from succeeding even against Locke's intended opponent – the proponent of the view that we are souls ('thinking substances' in his terminology). Locke's discussion of forgetfulness is preceded by the following observation:

> [That a single thinking substance constitutes one's self across time] few would think they had reason to doubt of, if these perceptions, with their consciousness, always remained present in the mind, whereby the same thinking thing would be always consciously present, and, as would be thought, evidently the same to itself. (Locke 1689/1975, Book II, ch. 27, §10/pp. 335–336)

This shows that Locke believes that we can sometimes know that numerically the same substance thinks in us now as did some short time ago.[41] Further, it seems reasonable to take this passage to tell us that whenever we remember a continuous period of conscious awareness, we can tell whether a single thinking substance was involved throughout. There are, then, two hypotheses to contend with. The first is that the substance thinking within us was changed in the midst of a period of conscious awareness, but we have forgotten that this occurred. The second is that it was changed during a period of unconsciousness, such as whilst we were asleep. I will briefly suggest that we have some reason to discount both scenarios.

Consider first the supposition that one's soul was changed whilst one was conscious. There are a few ways to rebut this. First, one might argue that the possibility is logically incoherent. John Foster presents an argument to this effect (Foster 1991, pp. 252–255). This argument relies upon two claims. First, successive experiences within any stream of conscious awareness 'overlap', sharing some of their temporal parts. Second, if a thinking substance is responsible for any part of an experience, then that entity is (solely) responsible for the whole of that experience. This entails that any experience is associated with a single soul for its entirety, including for the portions which it shares with experiences immediately before or after it. Because these overlapping portions are shared, it follows from the second claim that successive experiences are associated with

---

41. This is a significant difference from the thought experiment imagined by Kant in footnote 29 (p. 35), which imagines that it is never possible to tell whether or not one has undergone a change of soul. Kant's version of the argument will not be subject to all of the criticisms raised in this subsection.

the same soul throughout.[42] It follows that there cannot be a change in the thinking substance associated with any continuous stream of conscious awareness. If correct, this argument would refute the suggestion that we could undergo a change of soul whilst awake.

Even if it is not incoherent to think that a changeover of soul could occur within a continuous period of conscious awareness, we may nevertheless have good reason to believe that it has not happened to us. We can make the case in two different ways by reflection on the experiences that we can remember. First, none of those experiences feature a changeover of thinking substance. If our memories are sufficiently rich, then this might justify us in concluding that we have never undergone a changeover of soul whilst conscious. The argument here is inductive; we remember a great number of experiences, none of which have a certain feature, and thereby conclude that we have never experienced this feature. Second, this response could be strengthened further were there grounds to believe that had I undergone a change of soul, it would be very unlikely that I would have forgotten it. Since I cannot remember undergoing any such changes, and since I would be unlikely to have forgotten experiencing them, the argument runs, we have good reason to believe that I never have done so. Thus, we have no evidence that people ever change thinking substances whilst awake, and some reasons to think that they do not.

Given the above, it seems sensible to conclude that I have never undergone a change of soul whilst conscious. How about during a period of unconsciousness – for example, during sleep? Here, it must be agreed, memory has nothing to contribute. Nevertheless, we may still have independent reason to believe that such events do not occur. For example, Snowdon cites the ease of rousing people from sleep in response to a similar argument by Mark Johnston:[43]

---

42. Put more formally, let there be two successive experiences, X and Y. Let X have temporal parts *a* and *b*, and Y have temporal parts *b* and *c*. Because any experience is associated with the same soul throughout, it follows that *a* and *b* belong to the same soul as one another, and it follows that *b* and *c* belong to the same soul as one another. It follows from this, and the fact that 'belongs to the same Z as' labels a transitive relation, that *a*, *b*, and *c* all belong to one and the same soul. Hence, there can be no changeover of soul within a continuous period of conscious awareness.

43. Johnston writes:

> [If we are to know that a single person is present throughout] typical cases of unconsciousness or sleep, there cannot be a relevant alternative to the effect that people understood as bare loci have been replacing each other behind the stage of one human body's life. For that is an alternative that we are not in a position to rule out when faced with a sleeping or unconscious human body. If such an alternative were relevant we would have little chance of knowing who was in bed with us at night. (Johnston 1987a, p. 72)

§2.3: From Thought Experiments to Reference

> [T]here is very good evidence that people do not quit their normal
> body links during sleep, namely, that, however random the timing
> of reawakening they are always there when reawakened (Snowdon
> 1991, p. 124)

The Change of Soul scenario is not the only thought experiment in which there are changes in the thinking substance associated with a single body; we might also imagine cases in which a thinking substance inexplicably leaves a body altogether, or it is replaced by a fresh soul, with different psychological features. Unlike the example of a change of soul, these scenarios, were they to occur, would not be undetectable. We have no evidence that these scenarios actually occur. Hence, one must either explain why the change of soul thought experiment is possible, whilst these other scenarios are not, or accept that it is (at best) very rare that there should be alterations to the thinking substance associated with a single body. If one takes the latter route, then we can probably say that we are justified in assuming that we have not undergone a change of soul. I suspect that, on balance, we should extrapolate further, and deny that there ever is a change in the soul associated with a single person.

Thus, to conclude, the epistemic argument rests on too demanding an epistemology. Once this is set aside, the argument has little appeal. The next section considers an alternative way to read Locke's argument, and thereby motivate the wide psychological view.

2.3.2   The Response from Reference Determination

I have argued that the epistemic argument is unsuccessful. However, it is not the only argument that can be developed based upon Locke's example of a change of soul. One may read Locke as pointing out that the supposition that one person's soul has been exchanged for another is inert; were we to find out that this had actually occurred to someone, this would have no bearing on our judgments about her persistence. It might be argued that this fact discredits the claim that sameness of soul is required for our persistence.

The argument I have in mind rests upon the following principle:

Reference The reference of a singular term x is determined in such a way as to validate our treatment of different kinds of information as relevant or irrelevant to x's persistence.

Chapter 2: Persisting with Thought Experiments

Our responses to the cases given above suggest that we are inclined to take information about psychological connections between persons as relevant to their identity, and that we are frequently unconcerned about how those connections are realised. If this is correct, then Reference supports the wide psychological view. I will shortly argue that our intuitions are not as robust as would be required for this to be the case. First, however, I would like to briefly suggest that we have reason to attribute this argument to Locke, and to explain two ways in which one might motivate Reference.

There is some evidence that Locke supports his account of personal identity by reference to the ways in which we would respond to information about the identities (and non-identities) of thinking substances. Thus, for example, he says:

> Let any one reflect upon himself, and conclude that he has in himself an immaterial spirit, which is that which thinks in him, and, in the constant change of his body keeps him the same: and is that which he calls *himself*: let his also suppose it to be the same soul that was in Nestor or Thersites, at the siege of Troy ...[H]e now having no consciousness of any of the actions either of Nestor or Thersites, does or can he conceive himself the same person with either of them? ...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 of the soul of immaterial spirit that now informs him had been created, and began to exist, when it began to inform his present body. (Locke 1689/1975, §14/p. 339)

Here we are asked to suppose it known that our present soul is identical to that of Nestor and Thersites. Locke claims that this knowledge would not prompt us to take ourselves to be identical to either, or to take concern in their actions as though they were our own. He takes this to prove that sameness of thinking substance is not sufficient for personal identity. It is fair to interpret this as evidence that Locke takes our individuative practices, and the information we take to bear on them, to be relevant to personal identity.

Our interest is not primarily in Locke exegesis.[44] Thus, we should ask what reason we have to believe that Reference is true. I will provide a brief sketch of two possibilities. The first is specific to singular reference, whilst the second

---

44. Though Tabb 2018 argues convincingly that Locke's thought experiments are intended to reveal our patterns of linguistic usage, rather than to pump intuitions in favour of his views.

may also apply to other kinds of terms (such as words for properties).

In order to secure reference to an object, we need to display sensitivity to what it is to which we wish to refer. It is reasonable to think that this is manifested, in part, by our dispositions to reidentify the object – by what we take as evidence for and against its persistence. Dummett seems to have something like this in mind when he says:

> [T]o grasp the sense of [a proper name] is to have a criterion of identification of an object as the referent of the name . . . a grasp of the sense of the name consisted in a capacity to say, of any given object, whether or not it was the referent or bearer of the name (1973, p. 488)[45]

If we successfully refer to persons using their names, and to successfully refer using a name, one must grasp the sense of that name, then it follows from Dummett's view that we have criteria of identification for the persons to whom we refer. Now, it is reasonable to think that these criteria of identification are revealed through our responses to hypothetical scenarios. Hence, if Locke is correct that we are inclined to treat information about sameness of soul as irrelevant to claims about personal identity, this indicates that our criterion of identification for persons makes no reference to sameness of souls. It is therefore reasonable to conclude that sameness of soul is not required for personal identity, as the wide psychological view claims.

The second possibility relies on a different claim about how reference is determined (though not necessarily one which is incompatible with the first). Sider presents the claim as follows:

> The meaning of a term, T, is the candidate meaning for T that achieves the best balance of eligibility and fit with use. (2001, p. 191)

If Sider is correct, a number of criteria are involved in determining the meaning of the referring terms that we use. One of these is fit with use. To see how this is motivated, consider the (extremely plausible) principle that reference is only possible against a background of successful belief-formation. This principle entails that if a term refers at all, then we will ordinarily use that term to express correct beliefs. In turn, this suggests that when we set out to determine

---

45. This is a familiar idea, developed also in Evans 1982, pp. 106–107; 111; Quine 1950; P. F. Strawson 1959, pp. 202–209; Wiggins 2001, ch. 5.

the meaning of a disputed term, it is useful to begin by asking which candidate meanings make true the greatest number of sentences using that term that we accept, and false the greatest number of claims that we reject. If the wide psychological view fits better with use than its alternatives, then we have some reason to accept this view.

This is another compelling argument. It is worth noting, however, that the argument is not entirely conclusive, even if we agree that the wide psychological view fits best with our ways of re-identifying people. Sider claims that a term's meaning is determined by balancing our usage with 'eligibility'. As a result, we may sometimes accept an interpretation of a term which departs slightly from usage if it is rated as more eligible than other available interpretations which fit usage more closely. Sider gives the example of the natural kind term 'gold':

> Compare our community, which is enlightened as to the difference between gold and fool's gold, with another community that is not. It is plausible that even though the use of the term 'gold' in the unenlightened community fits fool's gold, nevertheless their term 'gold' does not apply to fool's gold. For there is a highly eligible meaning – namely, gold – that fits most of their use very well, and which does not apply to fool's gold. (That their term 'gold' means gold rather than gold-or-fool's-gold, or yellow metal, is particularly plausible if they encounter fool's gold only very rarely.) (Sider 2001, p. 191)

Sider thinks of eligibility as a matter of how 'natural' an interpretation is – how well it carves nature at the joints.[46] Thus, in order for the wide psychological view to succeed, it either needs to be at least as natural as its alternatives, or have a significant enough advantage in fit with use to outweigh any more natural alternatives. Without inquiring closely into alternatives, or into the exact mechanisms by which eligibility is balanced against fit with use, Sider's account only offers provisional support to the wide psychological view.

2.3.3 Controversies

This is not the place to enter into a full defence of Reference. However, in passing, I want to briefly engage with two possible sources of concern – one very

---

46. 'Naturalness' is introduced into philosophical discussion by D. Lewis 1983. Dorr and Hawthorne 2013 provides a nice overview and interpretation of debate around this notion. Sider 2011 provides an interesting extension, according to which it is not just terms for properties which can be evaluated for naturalness.

general, and one more keyed to the topic of personal identity. Though I will not say all that needs to be said, I want to suggest that it is at least not obvious that there are insuperable problems with Reference.

The General Worry

The general worry that I have in mind can be elicited by noting that both interpretations of Reference presuppose that our individuative practices are (at least ordinarily) valid; there (usually) exist objects whose persistence conditions match up to our dispositions to make judgments about their persistence. Some ontological views will bear this out. For example, one might think that there exist a plenitude of objects, one corresponding to each possible criterion of identity. Or, alternatively, one might think that the existence of objects is somehow to be explained in virtue of our adoption of terms that 'divide their reference', differentiating repeat encounters of a single object from encounters of qualitatively identical but numerically distinct objects.[47] However, other views may not be so amenable to Reference. I want to consider one such view, and to argue that it is not entirely obvious that the arguments in its favour justify us in rejecting Reference.

The view I have in mind is developed by Michael Ayers.[48] Ayers objects to views on which reference is determined by our beliefs on the basis that such views overlook the existence of 'psychologically basic' objects whose persistence is defined in terms of simple 'material coherence'.[49] We may recognise (and refer to) such entities without developing a conception of what they are, or of their criteria of identification; we need only to attend to our perceptual experience in order to recognise them. If we presume that we are such entities, then we must reject the wide psychological view, no matter how firm our intuitions in

47. See Quine 1950; *Word and Object* 1960, pp. 82–87 for an explanation of what is meant by the notion of 'dividing reference'. See also the discussion of 'feature placing' on pp. 202–209 of P. F. Strawson 1959).

48. See Ayers 1974; Ayers 1991b; Ayers 2005. Particularly important are chapters 20 and 21 of Ayers 1991a.

49. The notion of an object used here is roughly that of what is known as a 'Spelke object' (c.f. Spelke 1990)–a bounded, rigid, object whose parts move continuously through unobstructed space. For philosophical discussion of the impact of such objects on our notions of reference, see Xu 1997, and the replies in Wiggins 1997, and Ayers 1997. For ideas in the vicinity of Ayers, see also Snowdon 2009a; Brewer 2015; Brewer 2019. For a very high quality discussion of the idea that there is a connection between reference and the fact that objects come under various specific kinds, see Dickie 2014.

its favour; it will simply follow from our being entities of such kinds that we cannot survive breaks in our material continuity.[50]

I do not want to engage in a full discussion of Ayers' views. Instead, I shall content myself with making four small points which aim to diminish the effect of these views upon our discussion.

First, we may push back in favour of Reference. If we share firm convictions about what is required for the persistence of some kind of entity, and these convictions prove difficult to unseat, then this is at least prima facie evidence in favour of the existence of entities with the required persistence conditions. Indeed, if we are not mistaken about what we believe, or about some relatively basic facts by reference to which we justify our views about the persistence of the entities in question, it is not easy to sustain the view that we are making a mistake (rather than adopting different individuative practices, referring to different kinds of entity). That this is so is easily explained if Reference is true.

Second, Ayers' ontology is based on the idea that some kinds of objects are psychologically basic; our perceptual systems are particularly attuned to the motion and the persistence of coherent, bounded, entities.[51] It is consistent with this position that other, non-basic entities could exist; Ayers concedes this point when he claims that events, processes, etc. are individuated by appeal to our concepts. If our convictions in favour of the wide psychological view are firm, one might take this to indicate that we are non-basic entities, constituted by entities of a more basic kind.

Third, I am sceptical of the claim that psychologically basic entities cannot survive breaks in material continuity. To say that some kind of entity is psychologically basic is to say that under ordinary conditions, our attention is drawn to such an entity, and we are capable of tracking it without employing concepts, such as those of its persistence conditions. Ayers is certainly correct that there are such entities. I also agree that they are easily identified because of the boundedness of their parts, and because they ordinarily move as continuous, materially unified wholes. However, it is a further step from this to the conclusion that the persistence of these entities requires material continuity. So far, we have only been given reason to believe that these entities *typically* exhibit material continu-

---

50. See also Ayers 1991a, chs. 23–25, and van Inwagen 1997 for versions of this criticism. Cassam 1997 argues for the related thesis that we must conceive of ourselves as such entities in order to think of ourselves as continuants observing an objective spatial world.

51. See, in particular, Ayers 2005 for development of this thought.

ity, not that they must always do so.[52] Thus, what is most plausible about Ayers' views – the idea of a psychologically basic, non-conceptual, way to keep track of objects – does not strictly entail that the wide psychological view could not be true of such entities.

Fourth, and finally, it has been suggested that young children intuitively discriminate psychologically endowed objects from non-psychologically endowed ones, and that they are disposed to track the former through breakdowns in material continuity.[53] One might take this to suggest that there are two different types of perceptually basic object, one of whose persistence does not require material continuity. If so, then Ayers' basic framework, and his focus on psychologically basic entities need not be inconsistent with the wide psychological view.

Given these points, and given that any opposing ontological view is liable to be controversial, I want to suggest that we should not, at this juncture, reject Reference's commitment to the validity of our individuative practices.

The Specific Worry

Let's move on to the second worry. If Reference is correct, then the reference of our terms for persons depends, in part, on our individuative practices. This has the consequence that what I am referring to when I use the term 'I' depends on the individuative practices to which I subscribe. This might be thought objectionable – particularly if it is taken to entail that I could have referred first-personally to some other object than I actually do (and so, entails that which of two thinking beings I am depends on my linguistic choices). Surely 'I' just refers to me – the entity speaking.[54] It seems that my beliefs about myself should be

---

52. A different example may help here. It is plausible to think of artefacts, such as ships and watches, as perceptually basic when their parts are tightly connected to one another, and remain so over time. However, this fact does not prevent such objects from being disassembled, and later reassembled. Further, it does not stop them from being disassembled, losing parts, and then being reassembled with additional replacement parts – as might happen when the parts are strewn around a disorganised workshop, and left there for many decades. Under these circumstances, we would undeniably have to use some conceptual apparatus to judge whether or not the original artefact has survived. This gives an example of an object which is typically easily traced, but whose individuation sometimes needs additional conceptual resources.

53. See, for example, Kuhlmeier, Bloom, and Wynn 2004.

54. As Madden writes in discussing a related proposal by Noonan (Noonan 1998, pp. 312–317):

> It is much more plausible that 'I' and its cognates, and synonymous words in other languages, are governed by the convention of self-reference, or reflexive reference.

irrelevant to the reference of the first person. I have three comments to make in response.

First, we need not take Reference to entail that the reference of the terms that I use depends upon my, possibly idiosyncratic beliefs. The claim states that the reference of a singular term is determined by our individuative practices. We could therefore insist that the reference of the term 'I' is determined by the individuative practices adopted by one's linguistic community. On this view, I could not change the reference of the first-person by adopting different beliefs to my peers about my own persistence (indeed, if I were to do so, I would simply be wrong about what it takes for me to continue to exist). This brings me to my second and third points.

My second point consists in a very brief defence of the suggestion that our shared practices for reidentifying persons help to determine the reference of the first-person pronoun. The case consists in the observation that the first-person pronoun is correlative with a range of other referential devices, such as the second-person, our names, and indeed the word 'person' itself. These devices can be used interchangeably when we indirectly report speech. When Holly says something about Edward, I can say to Edward 'Holly was talking about you', and he can ask 'What did she say about me?'. In this exchange, three referential devices refer to the same object; when I say 'you', Holly says 'Edward', and Edward says 'me', we are all referring to the same person – Edward. A philosophical analysis of the reference of these devices should explain how each refers, and why they are each guaranteed to refer to the same objects on occasions such as this. Reference provides a simple answer; these different devices co-refer because we treat them as co-referential (and so, bring the same individuative practices to bear upon their use). I do not find it disconcerting that the referents of the second-person pronoun, and of our names, should be determined by reference to our individuative practices. Given this, and given that the first person is correlative with these other devices, I am disinclined to think that there is something objectionable about the claim that the reference of the term 'I' depends on the judgments that people in my linguistic community

---

They are words by means of which one may intentionally perform the linguistic act
any x performs when x refers to x. (Madden 2016b, p. 189)

This is compatible with the claim that we need a conception of our persistence conditions in order to perform the linguistic act in question. However, I suspect that it will be felt that it is implausible that we need such a conception in order to refer to ourselves.

are inclined to make about my persistence.

Third, I want to comment on the suggestion that we could change the reference of the first person pronoun by changing our practices for individuating persons.[55] I agree that this seems objectionable. However, whether this is a possibility depends on how malleable our individuative practices are, or could become; one cannot claim, from the armchair, that it should be possible for a linguistic community to think of persons, and about their persistence, in a different way than we do.[56] If our practices are not very malleable, then this objectionable scenario could never arise. It would follow that the objection to Reference is based upon a false presupposition about the number of ways in which we could form a workable conception of personal persistence.

I conclude that if Locke is correct that we treat information about sameness of brain or soul as irrelevant to personal identity, then there is good reason to believe that the wide psychological view is true. I have shown that some of our responses to thought experiments suggest that this is the case. Unfortunately, this evidence is far from conclusive; the following section raises two challenges for these intuitions. I conclude that appeals to intuition alone do not provide sufficient reason to believe the wide psychological view.

2.4 Problems with Thought Experiments

To summarise: Our intuitions reflect our beliefs about personal identity. If they cohere with the wide psychological view, this provides some evidence that we believe in that view (or, at least, that we believe in some of its consequences – that we could survive teletransportation and so on). Our beliefs, in turn, at

---

55. Johnston presents an interesting alternative proposal in his 1989. According to this proposal, we can change our own persistence conditions by changing our individuative practices. This interesting proposal avoids the worry that, by changing our individuative practices, we can change the referents of our first-personal thought. However, I find it difficult to believe the proposal itself, and so will set it aside without further discussion.

56. Here, it is instructive to consider Eli Hirsch's example of the contacti – a strange community of creatures who adopt a practice on which when two people come into contact, they swap bodies until they later cease to be in contact (Hirsch 1982, pp. 287–292). Though this is a possible concept to have, I do not think that it is a possible way to think about ourselves, or about one another. Indeed, Johnston (I think successfully) argues that the contacti do not think of themselves as such beings on the basis that, according to Hirsch, they show the same patterns of self-concern as we do (if they know that two people will be put into contact and then tortured, they show concern for the person who will come to occupy their body, and not for the body that they will, on their conceptual schema, come to occupy) (Johnston 1989, pp. 446–448).

least partly determine what we are talking about when we refer to persons. If we really do believe in the wide psychological view, this constitutes evidence that this view is true. This evidence can only be overcome if it can be shown either that the view is incoherent, or that it is outweighed by a significantly more eligible alternative. Unfortunately for the wide psychological view, we need not consider either of these alternatives; the argument falters at an earlier step.

This section argues that it is unlikely that our intuitions reflect belief in the wide psychological view. This is revealed by the fact that our tendency to respond to thought experiments as this view predicts is far from robust. First, we are sometimes inclined to make judgments about other scenarios which are at odds with the wide psychological view. Second, our intuitions about particular cases are easily affected by their framing. We should conclude that we do not have deeply held convictions in favour of the wide psychological view.

2.4.1   Contrary Evidence

I have shown that we have some intuitions in favour of the wide psychological view. However, some of our intuitive responses to other thought experiments (and, indeed, to circumstances that arise in real life) do not line up with this view. First, consider the following thought experiment, introduced by Parfit:

Branch-line Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: 'It's not working. What did I do wrong?'

'It's working', he replies, handing me a printed card. This reads: 'The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.' (Parfit 1984, p. 200)

This case evokes the intuition that virtually the same procedure as was used in Teletransportation (p. 34) – scanning and replicating one's body and brain – creates a new individual who is a mere copy of the original. This intuition, at least initially, looks to conflict with the thought that one survives the original teletransportation procedure. Further, I think that it is quite normal to find this

judgment more compelling than the judgment that one could survive Teletransportation.

There are various ways to respond to this conflict, including ways to preserve our intuitive verdicts about both cases.[57] Thus, my claim is not that this completely stymies debate about personal identity. However, even if our differing intuitions could be reconciled, it would be a further step to say that our intuitions reveal that we believe the resulting theory. For this to be the case, we would need to ascribe to people what is likely to be a highly complex and philosophically rich belief. This is in itself unlikely. Instead, I suspect that we should simply concede that we don't have fixed beliefs about what our persistence requires.

We can find two further reasons to doubt that our seemingly conflicting intuitions reflect a unified theory of our identity. The first is the fact, discussed in the next subsection, that our responses to a single scenario are very easily affected by its framing. This suggests that we are undecided about what would happen to a person placed in these cicumstances – and hence that we don't have a settled view of our persistence conditions. The second is the observation that we are, at least initially, puzzled to find mismatches between our intuitions about different cases; we take this to need resolution.[58] This further discredits the claim that we already believe a background theory which unifies these different reactions. Given this, I think that we should deny that the wide psychological view gathers any strong support from our responses to thought experiments.

Both Teletransportation and Branch-line are hypothetical – depending on circumstances which are, at present, far beyond our abilities to create. It is therefore worth contrasting them with the following cases, both of which occur:

Amnesia: As I age, my capacity to form new memories declines. Eventually, I find myself unable on any given day to recall anything about the previous day, no matter how active I was on that day.

---

57. See in particular Nozick 1981, Ch. 1. I will briefly discuss Nozick's views in the next chapter.

58. Nichols and Bruno 2010 provides evidence that people are generally inclined to resolve disputes in favour of the view that psychological continuity is sufficient for identity, and bodily continuity unnecessary. We should note, though, that the hypothetical scenario they use to support the psychological view – a version of Shoemaker's brain transplant thought experiment (Shoemaker 1963, pp. 23–24) – is consistent with the narrow psychological view. Thus, even were they correct, this would provide little by way of evidence for the wide psychological view. Nichols and Bruno's argument is discussed in Berniūnas and Dranseika 2016.

Chapter 2: Persisting with Thought Experiments

Persistent Vegetative State: On my commute home, I am involved in a serious
car accident. This accident causes damage to much of my upper brain,
removing from me the capacity for conscious awareness, and erasing my
personality traits and memories. It leaves enough of the lower-brain intact
that my body can be kept alive if supplied with external life support.

The ordinary reaction to both cases is to think that the subject has survived;
in the first case, I am suffering from a debilitating disease, in the second, I re-
main in a coma.[59] These reactions suggest that psychological connections are
not necessary for our continued existence. As such, they seem to conflict with
both the wide and the narrow psychological views (although, as above, reconcil-
ing moves are possible). Yet again, we find that our intuitions do not give strong
evidence in favour of the wide psychological view.

As I have already stated, the fact that these conflicting intuitions exist is not
decisive evidence against the wide psychological view; there is much that might
be said in response.[60] However, it does show that we cannot trust entirely to
intuition to resolve debates about personal identity; we seem to have conflicting
intuitions, and without examining the conflict, or unpacking the results, we risk
a distorted picture of our nature.

2.4.2 Framing Effects

I have argued that different thought experiments can throw up competing in-
tuitions. This section argues that things are worse still; a single thought ex-
periment may, depending on the context within which it is presented, provoke
differing responses. Thus, for example, after meeting the branch-line case, one

59. Snowdon vividly illustrates the plausibility of these intuitions, as against the verdicts sup-
ported by the wide psychological view:

> The relatives and friends would unquestionably accept responsibility for [an amne-
> siac or comatose mother]. They would treat it with respect and tenderness, look af-
> ter its possessions, and express and feel the strongest emotions directed at it. There
> is no sense, it seems to me, in which they do not really identify what is there *as* the
> mother. (Snowdon 2014b, p. 135)

60. For example, Shoemaker distinguishes two types of amnesia (Shoemaker 1984, pp. 86–87). In
normal amnesia, whilst the capacity to form new episodic memories is lost, the subject retains a
great number of learned psychological capacities (e.g. she may remember how to play the piano).
Here, it is implausible that the subject has ceased to exist, but an appropriately liberalised version
of Locke's theory can accommodate her survival. In a 'brain zap', by contrast, every particular
feature and personality trait associated with a person is lost. Here, Shoemaker claims that it is not
implausible that the subject has ceased to exist.

is unlikely to feel a residual pull to judge that one might survive teletransporta-
tion.[61] Here, our intuition about one case is altered by presenting it alongside
another, about which we have firm intuitions.

I want to give one more example of the impact of framing effects on our
intuitions. Bernard Williams very famously exhibits this phenomenon in the
following example:

> Someone in whose power I am tells me that I am going to be tortured
> tomorrow. I am frightened, and look forward to tomorrow in great
> apprehension. He adds that when the time comes, I shall not remem-
> ber being told that this was going to happen to me, since shortly be-
> fore the torture something else will be done to me which will make
> me forget the announcement ... [T]he person in charge [adds] lastly
> that the impressions of my past with which I shall be equipped on the
> eve of torture will exactly fit the past of another person now living,
> and that indeed I shall acquire these impressions by (for instance) in-
> formation now in his brain being copied into mine. (Williams 1970,
> pp. 167–168)[62]

This case is, as Williams puts it, only one side of the familiar idea of a
'body swap', in which the memories and psychological traits associated with
one person are transferred to another's body. It is usual to take such examples
to motivate the wide psychological view; two people literally swap bodies in

---

61. Of course, this intuition may later return, particularly if the case is described vividly, as
within the context of a fiction. Thus, I suspect that someone who watches an episode of Star
Trek after philosophizing is unlikely to have any trouble suspending disbelief about the efficacy
of teletransportation as a method of travel. This, though, underlines the point that context and
framing are liable to affect the intuitions that we have. My claim is that this may be to negative
effect.

62. Another example is Unger's extremely clever variant on Teletransportation:

> [At a] later date, there might be ways of trying to preserve a [diseased[person until
> a cure is found for her particular affliction. Let us imagine some of the technologies
> that might then be available ... The cheapest might be a purely informational taping
> process. In this taping process, a person's brain and body are quickly decomposed as
> detailed information concerning their molecular arrangement is recorded and stored
> on tape for as long as needs be ... Perhaps ten years later a cure is found. Then the
> device uses the stored information to make, out of a new batch of appropriate matter,
> a peson who is, at the least, almost precisely like the original person last was ... The
> cure is applied to this constructed person, who then goes on healthily to live for at
> least another twenty years. (Unger 1990, pp. 21–22)

I think he is right to observe that few of us think that the original person survives to see herself
cured.

virtue of an unusual method of generating psychological connections between them.[63] Here, however, a different description of the very same thought experiment seems to speak against the wide view, and in favour of the opposed claim that memory and personality traits are irrelevant to personal identity. If some thought experiments present conflicting intuitions depending on how they are presented, and depending on which thought experiments are presented antecedently to them, we cannot say that they establish our beliefs about personal identity. Given that the intuitions which favour the wide psychological view are so fragile, it seems that we should be hesitant to endorse it.

Williams' thought experiment has given rise to a range of responses. The pessimistic conclusion drawn by Johnston is that our concept of personal identity is too unspecific to ground any but the most permissive account of our persistence conditions – something which, he thinks, conflicts with the claim that we are able to make reliable judgments about personal identity (Johnston 1987a). Dainton and Bayne, on the other hand, suggest that such framing effects are evidence of 'hidden variables' which, once identified, point the way towards a unified account of personal identity (Dainton and Baine 2005). We need not pause to consider these, or other, responses; Williams case suffices at least to show that a single presentation of a lone thought experiment is unlikely to reveal how we are inclined to make judgments about personal identity; much more than this is needed to support the wide psychological view.

2.5   The Argument from Possible Societies

I have argued that it is far from clear that we have settled intuitions in favour of the wide psychological view; whilst we sometimes seem inclined to accept this view, at other times, we seem far less certain. Shoemaker suggests that there is a way to see past this deadlock of intuitions. I will briefly argue that his suggestion is unsuccessful.

Shoemaker's argument begins with the following scenario:

Periodic Body-change Imagine a society living in an environment in which an

---

63. The most famous use of this thought experiment is Locke's 'prince and the cobbler' case:

> [S]hould the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, 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. (Locke 1689/1975, §15/p. 340)

increase in some sort of radiation has made it impossible for a human body to remain healthy for more than a few years. Being highly advanced technologically, the society has developed the following procedure for dealing with this. For each person, there is a stock. of duplicate bodies, cloned from cells taken from that person and grown by an accelerated process in a radiation-proof vault, where they are stored. Periodically a person goes into the hospital for a 'body-change'. This consists in his total brain-state being transferred to the brain of one of his duplicate bodies. At the end of the procedure, the original body is incinerated. (Shoemaker 1984, pp. 108–109)

We are to imagine that this occurs routinely, and that "all of the social practices of the society presuppose that the procedure is person-preserving" (ibid., p. 108); when one undergoes a body-change, everything proceeds as though you had survived. Further, we are to suppose that the members of that society are not under any illusion as to what happens; they don't believe, for example, that you have an immaterial soul which is transferred along with a body-change. According to Shoemaker:

> If we confronted such a society, there would, I think, be a strong case for saying that what *they* mean by 'person' is such that [a body-change] is person-preserving (using 'person' in *their* sense) ... But there would also be a strong reason for saying that what they mean by 'person' is what we mean by it; they call the same things persons, offer the same sorts of characterizations of what sorts of things persons are, and attach the same kinds of social consequences to judgements of personal identity ... But if they are right in thinking that [a body-change] is person-preserving, and if they mean the same by 'person' as we do, then it seems that *we* ought to regard the BST-procedure as person-preserving. (ibid., pp. 109–110)[64]

Shoemaker presents the case as though the people involved have an entirely settled conception of personal identity. It is easy to put pressure on this by introducing scenarios which do not fit well with the wide psychological view. Imagine, for example, a variant on the Branch-line case, in which a person's original body is locked away immediately after the BST-procedure, but remains alive

---

64. Johnston responds to this case in Johnston 1987b, and Unger on pp. 89–92 of his 1990. They argue that the imagined community mean something slightly different by 'person' than we do. I suggest an alternative response.

and operative. After a number of days, the body is finally incinerated. Here, it is difficult to dispense with the feeling that the original person has died and been replaced. Were the members of the imagined society to feel no conflict here, we should wonder whether they really do mean the same thing by 'person' as we do. This would undermine Shoemaker's argument.[65] Indeed, it strikes me that we should want to see how the members of the imagined society respond to examples such as this before concluding that they are right to think of body-changes as person-preserving, even using the conception of personal identity that they adopt. Faced with unsuitable examples, they need to convince us that their conception of personal identity is sustainable.

If this is correct, then Shoemaker's imagined society must face up to the same conflicts of intuition as we do. They must engage in similar philosophical debates in order to decide upon a conception of personal identity. In this, their situation is parallel to ours; they need to trust to more than individual intuitions (or social practices) in order to determine what they mean by 'person'. Shoemaker's example therefore fails to resolve the conflict of intuitions, or to show that we already accept the wide psychological view.

2.6 Conclusion

Although it is right to allow our intuitions to play a role in theorising about personal identity, our intuitions in favour of the wide psychological view are insufficiently robust to motivate it. However, our discussion has not hit a complete dead end. This chapter has explained how our intuitions are relevant to theorising about personal identity. They are relevant because they reflect our existing beliefs, and these beliefs help to determine reference. I have also explained what is wrong with (a rather quick) way to break the impasse. Given what I have argued, proponents of the wide psychological view would do well either to develop a relatively natural account of our persistence through time, which will thereby provide a highly eligible interpretation of our talk about persons, or to present an independent argument for their position. The next chapter evaluates

---

65. Of course, it is conceivable that the members of this community should feel conflicted about such cases, or understand why we do, but nevertheless be able to explain and defend their view. Here, however, we have a neat test as to whether their conception differs from our own. We should look through their defence, and see how natural it seems to us. Is it the type of defence that we might be prepared to take up and offer? If not, this again suggests that we do not share their conception of personal identity.

§2.6: Conclusion

an attempt by Sydney Shoemaker to do both by reference to functionalism in the philosophy of mind. If Shoemaker is correct, then when we ascribe mental states to someone, we are committed to thinking of them as a psychological continuer. I will argue that this is not the case. Further, I will suggest that we should prefer an alternative to Shoemaker's account of personal identity which places material constraints upon our continued existence. Shoemaker's argument supports the narrow psychological view, rather than the wide view.

Functional Subjects

According to the wide psychological view, our persistence requires neither gradual material turnover nor the persistence of a fundamental substance, such as a soul. The previous chapter considered whether our intuitive responses to thought experiments support this view. I argued that they do not. This chapter considers an alternative argument, once developed and endorsed by Sydney Shoemaker.[66] The argument purports to show that anyone committed to a functionalist theory of the mind should accept a psychological account of persistence. I argue that the argument fails.

Section 3.1 of the chapter explains how Shoemaker understands functionalism, and how this can be used to define the notion of psychological continuity. Subsequently, 3.2 introduces Shoemaker's argument. I argue that one can be a functionalist without accepting the wide psychological view.[67] Shoemaker's argument is therefore unsuccessful.

Though Shoemaker fails to show that functionalism necessitates a psychological account of persistence, he may still succeed in showing that it motivates, or provides the resources by which to develop, such an account. 3.3 considers this claim. I face Shoemaker's account of psychological continuity with two puzzle cases in which it appears that psychological continuity takes a branching form. These cases demonstrate that psychological continuity, as it is defined by

---

66. Shoemaker no longer endorses this argument; in Shoemaker 2004b, he explicitly endorses the kind of view that I shall support here, as against the wide psychological view. Nevertheless, the argument is worth investigating in its own right.

67. This claim is touched on in Ch. 7 of Wiggins 2001, and in the 2004 exchange between Shoemaker and Wiggins in *the Monist*. Shoemaker's position is also thoroughly discussed in Agar 2003; Davis 1998; Fuller 1992 and Olson 2002. Davis 2001 presents a very interesting argument to the effect that Shoemaker's position does not even entail that one would survive having one's brain transferred into what was another person's body.

the functionalist, does not suffice for personal identity. Shoemaker can respond by modifying his analysis, telling us that psychological continuity suffices for personal identity only provided some further constraint is not violated. I analyse three possible modifications. Two of these – the inclusion of a no-branching requirement, and adoption of the closest continuer theory of identity – can be faulted. The third response seems promising, but fails to uphold the wide psychological view, leading instead the the narrow view. I conclude that functionalism does not support the wide psychological view.

3.1   Shoemaker's Account

Though Shoemaker's published work on personal identity begins in 1959, the theory that interests me emerges later, in his 'Personal Identity: A Materialist's Account' (1984). In the years since, Shoemaker's account has been developed in conversation with a variety of other topics in metaphysics and the philosophy of mind.[68] As a result, it has many ingenious applications. Notably, Shoemaker derives from it the surprising claim that animals cannot think.[69] This claim has been discussed elsewhere at length.[70] Accordingly, I set it aside to focus on the more basic motivations of Shoemaker's views. This section provides a brief overview. Following sections discuss the claim that functionalism entails Shoemaker's account and argue that, in light of some technical obstacles, the account is unlikely to provide much motivation for the wide psychological view.

3.1.1   Psychological Continuity

We can get Shoemaker's account into focus by considering how it departs from traditional 'memory-based' accounts of personal identity, such as Locke's. According to such theories, one is identical to the subject of some past experience only if one can remember undergoing that experience. In recent years, there has been a move away such accounts of personal identity, and towards alternatives

---

68. See, for example, Shoemaker 2007; Shoemaker 1979; Shoemaker 1997; Shoemaker 1985, Ch. 5; Shoemaker 2012.

69. See Shoemaker 1999; Shoemaker 2008. In Shoemaker 2016, the narrative changes slightly. Wherever each of us is, there is both an animal which can think (whose persistence conditions are not biological), and one which cannot (whose persistence conditions are biological).

70. See Árnadóttir 2010; D. B. Hershenov 2006; Olson 2002; Zimmerman 2009, pp. 723–737 for responses to Shoemaker on this point. Shoemaker responds to Olson in his 2004b.

according to which psychological connections other than memory play a role in determining personal persistence. Derek Parfit – one of the figures responsible for this trend – puts it as follows:

> Locke suggested that experience-memory provides the criterion of personal identity ... We should also revise the view so that it appeals to other facts. Besides direct memories, there are several other kinds of direct psychological connection. One such connection is that which holds between an intention and the later act in which this intention is carried out. Other such direct connections are those which hold when a belief, or a desire, or any other psychological feature, continues to be had. (Parfit 1984, pp. 205–206)

According to Parfit and Shoemaker, personal identity requires the preservation of a 'sufficient number' of direct psychological connections between adjacent stages of a person's career.[71] This can occur in the absence of memory connections; an amnesiac might retain enough of her former skills and personality traits to persist despite the loss of her memories. A proponent of this more liberal view therefore does not have to claim that memory is necessary for personal identity. At the same time, though, they owe us a characterisation of the types of psychological connections which are relevant to our persistence.

Shoemaker turns to functionalism to characterise the psychological connections relevant to an account of personal identity. The following subsection explains this aspect of his views.

3.1.2 Functionalism

Functionalists hold that mental states are individuated by reference to their causal role.[72] To explain what this means, it is useful to introduce a few technical notions from Shoemaker's writings. We shall begin with the notion of a 'successor state':

> While the content of a person's mental state (belief, intention, etc.) will to a certain extent depend on the nature of all of his states at

---

71. I summarise Parfit's views in footnote 20 of chapter 1 (p. 27).

72. Various forms of functionalist account exist. Shoemaker introduces functionalism as a theory of the nature of mental states. Though he does not make the contrast here, functionalist accounts which detail the nature of our mental states are often contrasted with accounts which aim only to explain the meaning of our psychological terms (and which may be neutral about the nature of their referents). D. Lewis 1972; Putnam 1967; Putnam 1965 provide influential early statements of functionalism. Shoemaker 1981; Block 1978 provide some helpful distinctions between different forms of the view.

earlier times, there will often be a particular state [or collection of states] at an earlier time on whose content its content especially depends – and it is of that state that we will call it the 'successor state'. (Shoemaker 1984, p. 95)

Though Shoemaker's characterisation only applies to states with cognitive content, I think that the notion of a successor state can usefully be extended. We shall say that a mental state's successors include any states or behaviours whose features particularly depend upon it, whether or not those states are aptly described as contentful.

Functionalism claims that each mental state is associated with a series of 'typical' successors – states which it is particularly apt to bring about. The view holds that to be in a particular mental state requires only that one is in a state liable to produce these successors.[73] Again, we can use a technical notion of Shoemaker's – that of a 'conditional power' – to clarify what this involves:

> An object has power P conditionally upon the possession of the properties in set Q if it has some property r such that having the properties in Q together with r is causally sufficient for having P, while having the properties in Q is not by itself causally sufficient for having P. Thus, for example, a knife-shaped object has the power of cutting wood conditionally upon being knife-sized and made of steel; for it is true of knife-shaped things, but not of things in general, that if they are knife-sized and made of steel they will have the power to cut wood. When a thing has a power conditionally upon the possession of certain properties, let us say that this amounts to its having a *conditional power*. (Shoemaker 1980, p. 115)

Put more simply, a sentence ascribing a conditional power to an object relays information what it would be able to do were it to possess the properties upon which the power is conditional.

The functionalist says that each mental state is associated with a causal role, constituted by a series of conditional powers.[74] Each of those conditional pow-

---

73. As Shoemaker puts it, "no state would count as the belief that it is raining unless it satisfied [the relevant] characterization, and any state that satisfied it would automatically count as that belief" (Shoemaker 1984, p. 92).

74. Shoemaker is attracted to the view that all properties are individuated by the conditional powers which they convey to their objects – "the contributions [they are] capable of making to the causal powers of things that have them" (Shoemaker 1980, p. 115). If he is correct about this, functionalism about the mind is an application of a more general theory about the nature of properties to the specific case of mental states.

ers specifies circumstances under which one of the state's typical successors will be produced. Thus, for example, the belief that it is raining is liable to cause one to take an umbrella with one to the shops conditional on a desire to keep dry, and the belief that an umbrella will ward off the rain (Shoemaker 1984, p. 92). The functionalist therefore claims that to ascribe that belief to someone is, in part, to ascribe to her the causal power to take an umbrella to the shops, should those circumstances arise.

I hope that this gloss provides enough by way of an overview of functionalism. We can connect this to the ideas of the previous subsection. First off, we can define psychological continuity as "the playing out over time of the functional natures of the mental states characteristic of persons" (ibid., p. 95). This is to say that one person is psychologically continuous with an earlier person if and only if her states are successors of the earlier person's mental states (i.e. they have been produced by the activation of conditional causal powers associated with her earlier states). Functionalism thus provides a clean definition of the connections which Shoemaker takes to be constitutive of a person's mental states; these connections are those to which we appeal in defining the characteristic successors and predecessors of our mental states.

Shoemaker's account identifies what is clearly a significant ongoing pattern of causal influence. Such a pattern would, I think, be worth picking out and identifying independently of any claims about personal identity. Given that it is to be independently identified, it is not unrealistic to think that psychological continuity is a relatively natural notion. If so, then one may well think that we have good reason to endorse a psychological account of personal identity; the account's high degree of naturalness, alongside its fit with some of our intuitions, make it a very eligible interpretation of our speech about persons.[75] I will argue that some overlooked complexities make it difficult to see how this thought can be sustained. First, though, I want to refute Shoemaker's claim that functionalism entails his preferred account of personal identity.

---

75. As Shoemaker is at pains to point out (see, in particular, his 1997), an account of personal identity in terms of psychological continuity coheres with an intuitive picture of substance, on which something is a substance if its later states exhibit a high degree of counterfactual dependence upon its earlier states (in a predictable, lawlike manner).

3.2   The Entailment Claim

Before introducing functionalism, Shoemaker writes:

> What I propose to do now is to consider a widely held (and widely disputed) theory about the nature of mental states . . . and to see what that theory implies about the nature of personal identity. What it implies seems to me a version of the psychological continuity view. (Shoemaker 1984, p. 91)

Call this the 'Entailment Claim'. For this claim to be true, there ought to be a valid argument leading us from functionalism about mental states to a psychological account of personal persistence. Since Shoemaker, at this point in time, claimed that we could travel via teletransportation (ibid., pp. 110–111), and does so on the basis of his account, this argument should support the wide psychological view. In this section, I argue that the functionalist need not accept the wide psychological view. I begin by explaining how Shoemaker motivates the entailment claim. I then note that the argument does not entail that we would survive teletransportation. I explain Shoemaker's response, and argue that it does not succeed.

3.2.1   The Argument

Shoemaker emphasizes that functional definitions of mental states particularly focus on the difference they make to the future states of their subject (through interaction with her other mental states and the stimulation provided by her environment):

> But of course, it is in conjunction with other mental states of the same person that a mental state produces the effects it does; and its immediate effects, those the having of which is definitive of its being the mental state it is, will be states (or behaviour) on the part of the very same person who had the mental state in question. (ibid., p. 93)

If Shoemaker is correct, a future mental state is produced by the conditional powers definitive of a presently existing mental state only if the future mental state and the contemporary mental state have the same subject.[76] Since psycho-

---

76. Zimmerman helpfully labels powers of this kind – powers which a subject exercises through losing or gaining a property – 'identity-entailing powers' (Zimmerman 2009, pp. 680–682). Shoe-

logical continuity is defined in terms of the exercise of these conditional powers, psychological continuity entails personal identity. The argument can be summarised as follows:

1. x is only psychologically continuous with y if x's mental states are the products of the conditional powers definitive of y's mental states (or vice-versa).

2. x's mental states are the products of the conditional powers definitive of y's mental states (or vice-versa) only if x and y are identical.

C. x is only psychologically continuous with y if x is identical to y.

Although this might initially seem to nudge us in the direction of the wide psychological view, closer reflection will show that it does not. In order to support this view, we need an account of the conditions under which the conditional powers definitive of a mental state are exercised which (a) does not appeal to the idea of personal identity; and (b) does not require that these powers are exercised via their usual causal parthways. In the remainder of the section, I outline an objection to (a), and detail Shoemaker's reply. Following this, section 2.3 argues that, even if we accept Shoemaker's defence of (a), we should nevertheless hold that (b) cannot be met consistently with using psychological continuity to define personal identity.

3.2.2 The Circularity Objection

The objection I have in mind is nicely framed as a circularity worry. I agree with Shoemaker that mental states are defined largely by generalisations about the changes which their subjects are liable to undergo. Given that we have defined psychological continuity in terms of these changes, it follows that two subjects are psychologically continuous only if they are identical. However, this does not show that personal identity can be analysed in terms of psychological continuity. It could equally be the case that identity is a necessary condition for psychological continuity; one could conclude that Shoemaker's analysis shows that in order to know whether two subjects are psychologically continuous, we first need to resolve whether they are identical. This criticism suggests that one

---

maker's observation amounts to the claim that functionalists define psychological properties in terms of their identity-entailing powers.

must possess an independent account of our persistence conditions in order to define psychological continuity. It follows from this that one cannot (without circularity) define our persistence in terms of psychological continuity.

This circularity worry is analogous to another which plagues the traditional analysis of personal identity in terms of memory. Butler writes:

> [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. (Butler 1736, Dissertation I/p. 258)

The worry here is simple. We can agree with the traditional analysis that every event that one remembers is from one's own past. However, one might claim that the explanation of this fact is not that memory constitutes personal identity, but rather that memory is defined in terms of personal identity. According to this analysis of memory, one only has a genuine memory of a past event if that memory corresponds to an experience which one once had. Thus, it is a necessary condition on really remembering (as opposed to simply seeming to remember) that one remembers events from one's own past. If this is correct, then one cannot define personal identity in terms of memory without moving in a circle.[77]

To summarise, one might worry that it is impossible to establish that cases like Teletransportation involve psychological continuity unless we already know that we would survive them. As a result, one cannot use Shoemaker's analysis of psychological continuity to argue for any particular account of our persistence without assuming the point at issue. If this criticism is correct, we can conclude that the functional definitions of mental states do not entail the wide psychological account of personal identity.

3.2.3 Responding to the Circularity Objection

Of course, Shoemaker is sensitive to this concern. His response is best appreciated against the background of an important feature of functionalist theories of

---

77. The orthodox way to respond to this objection originates with Shoemaker's 'Persons and their Pasts' (1970). We define a state – 'quasi-memory' – which is exactly like memory, except that it is not a necessary condition upon quasi-remembering some past event that one actually experienced that event (see also Parfit 1984, pp. 219–223; Shoemaker 1984, pp. 80–88). This notion is criticised by Evans 1982, pp. 235–248, McDowell 1997, and Wiggins 2001, pp. 212–225. For a defence, see Roache 2006. We shall consider an alternative strategy below.

mind. It has often been observed that a definition of any mental state will mention its interactions with other mental states and processes. Thus, for example, Shoemaker writes:

> The belief that it is raining is defined (according to the functionalist) as the state which (among other things) leads to the taking of an umbrella when combined with (among other things) the desire to keep dry; and the desire to keep dry is defined as the state which (among other things) leads to the taking of an umbrella when combined with (among other things) the belief that it is raining. (Shoemaker 1984, pp. 98–99)

Though the functionalist defines belief and desire in terms of one another, it is not objectionable for her to do so. This is because we can easily determine what mental states a subject has by reference to her behaviour – the product of the combination of mental states that she possesses. For example, if someone takes an umbrella when she leaves my house, it is reasonable to guess that she desires something which she believes the umbrella will help her attain. Given this, we may form hypotheses about which beliefs and desires she is likely to entertain. Further, given certain background assumptions (e.g. that she knows the function of umbrellas, that she can hear the rain beating down upon the roof), we may conclude that it is most likely that she desires to keep dry, and believes that the umbrella will keep off the rain. If needed, this hypothesis can be further tested against her future behaviour. In doing this, we are working backwards from observed results to the values of several hidden variables; there is nothing objectionable about this process (or about the definitions which we rely upon in order to engage in it).

If functionalism is correct, the definition of any mental state comes as part of a wider 'package deal' – a collection of interdependent definitions which together outline a theory of the mind as a whole. Shoemaker proposes that we include the definition of psychological continuity within this package deal. Thus defined, psychological continuity is the relationship which holds between people identified at different times whenever the mental states and behaviours of the later person are appropriately sensitive to those possessed by the earlier person (i.e. they are of the right kind to be counted as direct or indirect descendants of her earlier mental states).[78] Under this proposal, all that is necessary for psy-

---

78. Shoemaker gives a particularly clear overview of this position on pp. 581–586 of his 2004a.

chological continuity is that there exist correlations of the right sort between the states and behaviours of subjects identified at different times.

This proposal creates the need for a small terminological clarification. At this point, Shoemaker has provided us with two definitions of psychological continuity. According to the first, psychological continuity involves the exercise of causal powers whose effects occur within their bearer. On this definition, it is an analytic truth that two subjects cannot be psychologically continuous unless they are identical to one another. The second definition, in contrast, looks to define psychological continuity in terms of correlations between the mental states possessed by subjects at different times. This definition does not automatically rule out that distinct subjects could be psychologically continuous; whether or not this is so will depend upon what types of correlations could exist between distinct subjects' mental states. If Shoemaker is correct, then these two definitions are co-extensive – both pick out the relation of personal identity. In the next section, I will argue that Shoemaker is not correct on this point; there are cases which, according to the second definition, involve psychological continuity between distinct individuals (whilst, according to the first definition, there is not psychological continuity at all). For the sake of clarity, in what follows, my use of the term 'psychological continuity' will correspond to the second definition, on which psychological continuity is defined just in terms of correlations between mental states existing at different times.

With psychological continuity defined in this way, one need not determine whether or not two subjects are identical before concluding that they are psychologically continuous. Thus, the circularity worry is avoided. Further, it is easy to see that Teletransportation preserves psychological continuity; the post-teletransport person's psychology very closely resembles that of the pre-teletransport person (and, where they differ, the post-teletransport person's features are accounted for by the combination of the circumstances in which they find themselves, and the functional definitions of the pre-teletransport person's mental states). As we have seen, Shoemaker proposes that psychological continuity is sufficient for personal identity. If his proposal is sustainable, it therefore entails the wide psychological view.

Before we consider objections to Shoemaker's analysis, we should consider

---

Similar proposals are suggested in Slors 2001, and Wiggins 1980, pp. 160–163, though Wiggins' later rejects the picture developed here (c.f. in particular his 2004a, and 2004b).

whether this analysis validates the entailment claim. It is easy to see that it does not. Shoemaker has shown how he can bypass circularity worries, producing a functional definition of a causal relationship that might hold between mental states. He then proposes that we define personal identity in terms of this relationship. There are two steps here, and both are optional for the functionalist. First, the functionalist could refuse Shoemaker's response to the circularity worry, instead choosing to incorporate an existing account of personal identity into her definition of psychological continuity. This would have the result that two subjects are only psychologically continuous when they are, according to the existing account, identical to one another, and the mental states of one are successors of the other's. Second, the functionalist could agree with Shoemaker's 'wide' functionalist definition of psychological continuity, but refuse to identify the relationship thus defined with that of personal identity. In either case, she would be able to accept an account of personal identity other than the wide psychological view.

To summarise, Shoemaker suggests that we can functionally define the relation that holds between mental states when they belong to the same person. In doing so, he avoids the charge that we must know whether or not two mental states belong to the same person before we determine whether one is caused by the conditional powers conferred by the other; the mere existence of systematic correlations between mental states suffices for their subjects to be psychologically continuous (and therefore, if Shoemaker is correct, identical). Nevertheless, it remains optional for functionalists to adopt this definition, and optional for them to agree that the notion he defines is co-extensive with personal identity. It follows that the entailment claim fails; functionalists need not adopt the wide psychological view. The following section extends this criticism by showing that Shoemaker's functional definition of psychological continuity needs modification if it is to provide a sustainable account of personal identity.

3.3   The Coherence of Shoemaker's Account

We have shown that functionalism does not, by itself, entail any particular account of personal identity. Nevertheless if Shoemaker's proposal is sustainable, it provides a neat, clear version of the wide psychological view. Further, this definition corresponds to a significant set of causal relations, which we should wish to identify independently of our concern with personal identity. These relations

have interesting and more or less uniform effects. Thus, were Shoemaker's proposal sustainable, I take it that this would be strong evidence in favour of the wide psychological view. Unfortunately, it is demonstrable that psychological continuity, as Shoemaker defines it, cannot be coextensive with personal identity. Though fixes are available, it is far from clear whether they will support the wide psychological view. We may therefore conclude that Shoemaker gives us little reason to accept this view.

3.3.1 Teletransportation and Fission

If the wide psychological view is correct, then it is theoretically possible for a person to travel by teletransportation. As noted above, Shoemaker's views sustain this verdict. Teletransportation produces a subject whose mental states duplicate those of the original. This subject will subsequently develop many of the mental states which the pre-teletransport person would have developed had her body not been destroyed.[79] Given this, we can say that these subjects are psychologically continuous, and hence – on Shoemaker's account – identical.

In the last chapter, I noted that it is controversial whether a person could survive teletransportation; whilst our initial impressions may support this claim, things are significantly less clear upon reflection. In this subsection, I want to take this criticism one step further; it can be proved that psychological continuity (as it has so far been defined) does not suffice for personal identity.

To see this, it is worth reflecting upon two cases. The first is the familiar example of fission:

Fission After someone's body has been scanned and destroyed, her blueprints are sent to two different machines, each of which assembles a duplicate. The upshot is two new brains and bodies, each replicating the pre-teletransport brain and body.

Though views differ, I am most inclined to say that two people exist at the end of this process – one for each body.[80] The pre-teletransport person is psy-

---

79. Further, where there is a difference, the states developed by the post-teletransport person will be equivalent to those that the pre-teletransport person would have developed had she found herself in the former's position.

80. A variety of creative alternatives have been devised. The most popular of these is the Lewisian claim that fission involves multiple occupancy (D. Lewis 1976); when fission occurs, it separates two persons who, up until that point, have shared their body and brains (being composed of the

chologically continuous with both of these later persons. However, since they are distinct, she cannot be identical to both. If we agree with this description of the case, we must deny that psychological continuity is always sufficient for personal identity.[81]

It will be helpful to supplement the example of fission with another thought experiment. This case, which I introduced in the previous chapter (p. 54), is the branch-line case; your brain and body are scanned and replicated, but your original body is not destroyed, and continues to support conscious life. What is interesting about this thought experiment is that it has the same formal structure as the fission case – a single person comes to be psychologically continuous with two people existing at a later time – but we are inclined to judge the case differently. It has the same formal structure as fission because the scanning and copying process, by which one's brain and body are replicated, is intended to be the very same type of process as that involved in Teletransportation. As such, it preserves psychological continuity. At the same time, your original body continues to sustain a conscious life as usual, without any disruption to psychological continuity. Thus, you come to be psychologically continuous with your replica, and with the person housed in your original body. Most people would believe that you are identical to the latter person; you remain, unscathed, within your original body, despite your replication.

These cases show that psychological continuity, if understood purely in terms of correspondences between the mental states of persons existing at different times, is not sufficient for personal identity. This does not entirely refute Shoemaker. For he may hold that psychological continuity, as defined above,

---

same temporal parts up until that point). Perry 1972 adds a third subject, who thereby comes to have two bodies. Dainton suggests that a single subject becomes scattered, by coming to have a non-linear personal timeline (Dainton 2008, pp. 378–393). Johansson 2010 explains that it could be determinate that I survive fission, whilst indeterminate which fission product I am. Gallois 1998 provides a novel analysis of identity over time, geared to cases such as this. Finally, Wollheim provides a delightfully idiosyncratic solution on which what happens to the subject (whether she comes to have two bodies, or ceases to exist) can depend upon her stage of psychological development (Wollheim 1979, pp. 224–232).

81. Of course, we could always reject this description, and accept one of the alternative responses developed in the previous footnote. Here is not the place to consider these responses in detail. I shall therefore content myself with only one point. This is that all of these responses are willing to identify the replica with someone who existed prior to replication. Whilst this is not quite so egregious a judgment as the claim that the original subject ceases to exist, it does strike us as quite odd; we are intuitively inclined to think that the branch-line case leaves the original subject unaffected, and just creates someone entirely new in addition. If so, then the response I provide to the no-branching requirement may also apply to such responses as the four-dimensionalist's acceptance that fission cases involve multiple occupancy.

only suffices for identity if certain other circumstances hold. We will examine this suggestion in what follows.

3.3.2 Solutions

If the above is right, we cannot claim that psychological continuity always suffices for personal identity. Shoemaker is sensitive to this worry, and has a response:

> It may be, however, that the relation we define by this procedure is not quite the relation of copersonality ...So let us simply dub this relation 'psychological unity'. Synchronic psychological unity will always be copersonality. And diachronic psychological unity will be copersonality as long as no 'branching' (e.g., fission) has occurred in the development of the succession of states. (Shoemaker 1984, pp. 100–101)

The general style of this response consists in adding necessary conditions to our analysis of personal identity. The result is an account which says that psychological continuity is sufficient for personal identity whenever these additional conditions hold. Responding in this way provides further ammunition to those who deny the entailment claim; since one introduces these additional constraints after functionally defining psychological continuity, the opponent of the psychological view may well accept Shoemaker's functionalist account of psychological continuity and nevertheless deny that it has anything to do with personal identity. Given that we have already argued that the entailment claim is false, let's set this fact aside. Instead, we will just consider whether Shoemaker's proposal can be refined to give a plausible account of personal identity. I will survey three possible modifications, and argue that they do little to motivate the wide psychological view.

Option 1 – Non-branching requirement

We begin with Shoemaker's preferred modification – the addition of a non-branching requirement. In fission, a subject existing at one time is psychologically continuous with two subjects existing at a later time. These subjects are numerically distinct from one another. This proves that psychological continuity, as we have defined it, does not suffice for the identity of persons. Nevertheless, one might respond that psychological continuity does suffice for personal iden-

tity provided that it does not take a branching form; according to this proposal, two subjects are identical if and only if they are psychologically continuous, but there exists no third subject psychologically continuous with one of them, but not with the other.[82] Fission violates this additional clause. Thus, by including this clause, we can circumvent the problems raised by fission.

Non-branching criteria are controversial, since they dictate that the identity of persons existing at different times depends on whether or not some third person exists who stands in a specified relationship to the person identified at the earlier time. This seems to make answers to questions about personal identity trade on overly extrinsic considerations. This has been taken to conflict with the plausible 'only x and y' principle. Harold Noonan presents a neat formulation of this rule when he writes:

Only x and y [W]hether a later individual y is identical with an earlier individual x can depend only on facts about x and y and the relationships between them: it cannot depend upon facts about any individuals other than x or y. Otherwise put, what the principle asserts is that whether x is identical with y can only depend upon the *intrinsic* relationship between them, it cannot be determined *extrinsically* (Noonan 1989, p. 152)[83]

Though I am sympathetic to this objection, I do not think that it is decisive. First, I would be happy to say that there may be some cases in which rejecting the Only x and y principle is our least worst option.[84] Second, though I am convinced that the answers to questions of identity cannot trade on *extrinsic* considerations, it is far from clear to me that the fact that an entity's career has branched is extrinsic to its relationship to future entities. Shoemaker writes:

> If during a certain interval there is no branching in a psychologically continuous series of person states, that seems an intrinsic fact about

---

82. Some interesting complexities are involved in formulating a non-branching requirement. See, in particular Gustafsson 2019.

83. Though this is the simplest formulation, there are some difficulties with it (c.f. Shoemaker 1984, pp. 115–118, and Noonan 1989, pp. 152–154). Noonan presents a revised version of the principle on p. 164. The principle itself is inspired by Williams' 1956–1957, pp. 238–240 and his 1960. For other formulations, and discussions of this principle, see Garrett 1990; Hawley 2005; Heller 1987; Noonan 1985c; Noonan 1985b; Noonan 1985a; Wiggins 2001, pp. 95–99; Wiggins 2012, pp. 4–5.

84. For example, Madden 2016a gives a compelling account of our intuitions about brain transplants. His analysis, however, requires us to adopt the closest continuer view, discussed in the next subsection of our discussion. Thus, though his account is attractive, it requires us to reject the Only x and y principle.

> that series and so about the individual whose career it is. So taking non-branching psychological continuity as a *sufficient* condition of identity does not offend against the rule. It is less clear whether it is violated if we take such continuity to be a *necessary* condition of identity; if there is branching at some point, and this is taken to show that the segment of the series up to that point and the segment of it from that point on are the careers of two different persons, it is perhaps unclear whether it is an intrinsic fact about the earlier person that there was an episode of branching at the end of his career or an intrinsic fact about the later person that there was such an episode at the beginning of his career. (Shoemaker 2004a, p. 588)

Put otherwise, when branching occurs, it occurs as a result of a causal process which (at the very least) causes a new entity to come into existence (and may also cause another to cease to exist). It is unclear why it is not an intrinsic fact about the new entity that it came into existence in such a fashion (e.g. as the result of the assembly of material by a teletransporter). But, if the fact that the new entity came into existence in that way is an intrinsic fact about its career, then we can say that it cannot be identical to anything that existed before it came into existence; doing so is not relying on anything other than relationships between it and the careers of anything with which it might be identified. Given this, it might be doubted whether a non-branching requirement does violate the principle that facts about identity cannot depend upon factors extrinsic to the relationship between two entities.

Even if we do not know quite what to say about the Only x and y rule, this is not to say that it does not highlight something objectionable about solving the problem of fission by introducing a non-branching requirement. To see this, we should begin by considering how this added requirement would deal with the branch-line case, in which a subject's psychological states are replicated without interfering with her body or brain. Here, the creation of a replica causes branching in the functionally mandated effects of the original person's mental states. The no-branching clause therefore rules that she ceases to exist. This strikes me as quite plainly the wrong result; intuitively, the original subject continues to exist, unaffected by her duplication elsewhere. Further, given the uninterrupted functioning of her brain, it strikes me that it is quite fair to complain that the non-branching requirement allows questions of identity to depend upon factors quite extrinsic to the relation between the objects identified. The non-branching requirement therefore has significant costs. We should look for alternatives.

§3.3: The Coherence of Shoemaker's Account

Option 2 – The Closest Continuer theory

In *Philosophical Explanations*, Nozick promises to capture the intuitions that we could survive teletransportation, that we would not survive fission, and that we would not cease to exist in the branch-line case. According to his 'closest continuer' theory of personal identity:

> y at t_2 is the same person as x at t_1 only if, first, y's properties at t_2 stem from, grow out of, are causally dependent on x's properties at t_1 and, second, there is no other z at t_2 that stands in a closer (or as close) relationship to x at t_1 than y at t_2 does. (1981, pp. 36–37)[85]

Our response to the branch-line case reveals that bodily continuity contributes to the closeness of the relationship between persons existing at different times. As a result, we can say that the original person continues to exist, housed within her original body, despite the existence of her replica. We can hold this consistently with allowing that ordinary teletransportation creates psychological connections which are 'good enough' for survival when no better candidate exists. The closest continuer theory therefore validates the claim that psychological continuity can suffice for personal identity in the absence of its usual cause, though it concedes that the presence of this cause does help one's claim to have survived.

The closest continuer theory shows us how to reconcile intuitions that are ordinarily taken to be in tension with one another. The theory is clever. However, in evaluating the view, we should consider whether it is more costly to jettison some of our intuitive reactions than it is to hold them in unison. I will present five reasons for thinking that this is not the case. First, though, we should note that, just as with the non-branching criterion, the closest continuer theory is controversial, and seems to conflict with the Only x and y principle. As before, I am unsure how to evaluate this criticism. The five reasons I suggest will therefore be slightly less theoretical in nature.

The first thing that I want to do is to suggest that it is not very costly to revise some of our intuitive reactions to thought experiments. One source of

---

85. This is strictly only a necessary condition on personal identity. There are two reasons for this. First, even an entity's closest continuer may fail to be close enough for identity. Second, an entity's closest continuer at one time may fall on a branch that, at a previous moment in time, was not as close as one of its competitor. Thus, for example, if the original person in the branch-line case suddenly dies, we would not say that she suddenly switches location to that of her replica even though her replica would be her closest continuer existing at that time.

evidence for this is the fact that presentation of the branch-line case is liable to dramatically affect how we view teletransportation. Once faced with the former case, it is difficult to feel confident that one would survive teletransportation; it seems to me much easier to conclude that one would not.

Second, I want to identify another advantage to rejecting some of our initial responses to thought experiments about personal identity. The advantage is that we can thereby bring our beliefs about persons into line with those about other items – such as the clothes on our body – which we do not believe could survive teletransportation (however easy it is to tell a story on which they appear to do so).[86] Again, this suggests that there is very little cost to rejecting our first impressions about Teletransportation.

My third point concerns the cost of accepting the closest continuer view. The theory is surprising, and marches out of step with our first impressions concerning which intuitions can be held consistently with one another. Unger provides a neat summary of our usual philosophical inclinations:

> We may have an intuition that [Teletransportation] is essentially the same as [Branch-line]. Whether the cases stand or whether the cases fall, they stand or fall together. This additional reaction is one of our *second-order* intuitions about cases concerning our survival . . . [I]n addition, we have reliable *second-order comparative* intuitions. As an example of this, we respond that, as far as survival is concerned, each of [Teletransportation, Branch-line, and Fission] is much more like the other two than it is like ordinary cases of survival. Judiciously treated, such comparative intuitions may help justify us in being more confident about our more straightforward second-order intuitions.

This leads to an argument against the closest continuer approach. The theory is generous to our 'first-order' intuitions about particular cases, but ignores many of our second-order, and second-order comparative, intuitions. This

---

86. Unger gives an example in which we might be inclined to say that a hunk of mud gets teletransported:

> Suppose that there is, in this story, a conspicuous *hunk of mud*, dense with a rare mineral. Perhaps held in a space explorer's gloved hand, the hunk, as well as what is holding it, may be exposed to an information-extracting machine. As a side-effect, this device casts the hunk's material bits to the wind, along with the bits of anything holding the hunk. On the other side, owing to the transmission of suitable information . . . there thus appears just the right sort of hunk of mud, constituted of new matter, held in just the right sort of gloved hand. (Unger 1990, p. 85)

seems to me a significant cost, particularly given our already-noted willingness to revise our first impressions about Teletransportation.

Fourth, we should note that Shoemaker expresses some uncertainty about the claim that preservation of the brain is a 'better' realiser of personal persistence than teletransportation. He claims that our strong inclination to judge that you would survive in the branch-line case, rather than undergoing fission, is 'parochial', asserting that no metaphysical reason can be given to believe this (Shoemaker 1984, pp. 130–132). In some ways, this is quite a puzzling comment; nothing seems clearer than that the scanning process of the branch-line case leaves the original subject unaffected (and so cannot cause her to cease to exist). Given this, it is difficult to see how our reaction to this case rests upon idiosyncrasies in our concept of personal identity – as Shoemaker's comments suggests.

Although I doubt that our judgment about the branch-line case is merely parochial, I think that we can get a handle on Shoemaker's concern. Our functionalist account of personal identity began with an account of the connections between mental states existing at different times. This account abstracted from the actual realisers of our mental states. Further, in holding that we could survive teletransportation, we showed little interest in the preservation of our brain and body. Put simply, our starting point has prioritised an account of persons as psychological entities, and downplayed all other concomitants of their actual existence. To prioritise the continued existence of the brain at this stage would betray this. If we are interested in psychological continuity, whatever its realisers, and we think that teletransportation preserves this, why do we gain by introducing a (non-psychologically grounded) distinction between better and worse realisers of psychological continuity?[87]

A fifth, though related, comment concerns the naturalness of the relationship tracked by a closest continuer account of personal identity. I suggested above that Shoemaker's account is appealing because it connects personal identity to a type of causal relationship that we already recognise – the relationship between psychological states and their successors. It thereby promises to provide a highly natural account of personal identity. Judged by the standards of the previous chapter, this makes Shoemaker's account a very eligible interpre-

---

87. I will, in effect, deal with an objection of this kind – holding that we have no reason to distinguish between different realisers of psychological continuity – in the next chapter.

tation of our practices of re-identifying persons. Unfortunately, however, the closest continuer account seems to frustrate these expectations. For one entity to be the closest continuer of another, the two must stand in some more specific relationship which dictates how closely the former entity continues the latter. Further, this more specific relationship might suffice for identity in one situation, when it constitutes the closest continuing relationship, but not in another, when it is outcompeted by a closer relationship. Thus, for example, the closest continuer account dictates that a single type of causal process can underwrite personal identity in Teletransportation, whilst it does not do so in Branch-line. Given this, the closest continuer account will sometimes treat very different underlying relationships alike in being sufficient for identity, whilst distinguishing between different instances of the same general type of relationship. It is plausible that this counts as evidence that being a person's closest continuer is not a very natural relationship in which two entities may stand. If so, this feature of the account militates against its eligibility as an interpretation of the persistence conditions of persons.

I do not think that these objections to the closest continuer account are decisive. However, they do show that the account has unattractive features. Since we are modifying Shoemaker's account in any event, we should consider the merits of more radical alternatives.

Option 3 – Going Narrow

Both of the options considered above agree that we could survive Teletransportation. Thus, they both agree that a functionally defined relationship between psychological states existing at different times sometimes suffices for personal identity. If the realiser of this functionally defined relationship matters at all, it does so only when two or more entities compete for the title of identity with a third, with whom each is psychologically continuous. The final option I want to consider rejects this equanimity regarding the various different realisers of psychological continuity. It says that psychological continuity only suffices for personal identity when it has the right sort of realiser (in our case, the continued functioning of the brain). As we shall see, this approach has a number of merits.

To begin to characterise this view, we should consider what it rules out. If the envisaged restriction is to help with Fission and the Branch-line case, it seems that it should require us to reject the verdict that we survive Teletransportation;

it is intuitive that the three cases, as described above, involve the same type of realiser of psychological continuity – viz. a process that involves scanning and replicating a subject's body and brain. What reason do we have for thinking that psychological continuity does not suffice for personal identity when realised in this way?

Teletransportation involves the duplication of a person's brain and body, based upon information gathered by scanning them. We have already encountered the objection that post-teletransport person stands in too extrinsic a relationship to the pre-teletransport person for the two to be identical. According to this objection, we are not the type of thing that can be transferred from place to place by the mere transfer of information.[88] Agar presents an interesting take on this objection, arguing that teletransportation involves a breakdown in psychological continuity:

> The relationship the teletransporter stands in to psychological patterns is quite different to that of neurophysiology. Rather than supporting or constituting the emergence of a desire to eat from a state of hunger, teletransportation constitutes the state of hunger emerging from a state of hunger, or perhaps the desire to eat emerging from the desire to eat. The teletransporter is not even a temporary replacement for the mind's functional architecture, because it plays no role in accounting for the diachronic patterns constitutive of folk psychology. (Agar 2003, p. 65)

Agar observes that teletransportation causes my present mental states to have successors only because these states are first copied. The copies of our mental states are then causally responsible for the generation of future successor states. Agar suggests that this process of copying should be considered a 'non-psychological intermediate' between the pre-teletransport and the post-teletransport person – a process which only indirectly contributes to the bringing about of the successors of a subject's mental states (it does not bring about successors to my current mental states, but merely causes there to exist future copies of those states, which bring about their successors). This is a plausible thought, and again captures a sense in which the connection between the pre-teletransport and the post-teletransport persons is of the wrong kind to consti-

---

88. In connection with this, see van Inwagen's 1997, in which he argues against the psychological theory on the basis that it neglects to consider constraints upon acceptable persistence conditions for material objects.

tute identity; it passes outside of the subject herself, and of her realisation of her mental states.

Given the above comments, we might claim that psychological continuity only suffices for personal identity when it involves the subject herself changing, and bringing about her mental states directly, and does not involve the intervention of external processes which (a) could occur whilst the subject remained unaffected, and (b) impose causal intermediaries between her mental states (or their realisers), and their successors (or the realisers of their successors). Teletransportation fails to meet this constraint, and so, the account under consideration denies that it realises psychological continuity in the correct way for a subject to persist.

The above only provides a rough gloss on what is likely to be a very technical constraint on the realisers of psychological continuity. I do not want to attempt a full specification of this constraint. Instead, I shall say two substantive things about it. The first is that this constraint corresponds closely to the idea of an immanent causal connection – a type of connection by which an object's later states are caused by its former states, rather than by external forces acting upon it.[89] This idea is often discussed in connection with causal theories of persistence through time. Thus, even without specifying the required constraint fully, we can see that there is an appetite for accounts of this kind. Second, I think that we can say something about what is required for psychological continuity to have the right kind of realiser.

If personal identity requires psychological continuity, it also requires that this continuity is not realised in too extrinsic a way, such as via intermediaries which do not, themselves, realise psychological states – such as the process of copying.[90] This constraint is met by the usual cause of psychological continuity – the continued functioning of a subject's brain and body. Further, it is plausible that this constraint entails that any new material that comes to constitute a subject can only realise psychological continuity in the right kind of way if it is

89. This notion is helpfully characterised in Zimmerman 1997. A useful, less technical, gloss on much the same notion is given by Unger's discussion of 'systemic energy' (1990, pp. 129–130).

90. In developing this constraint more fully, one would have to navigate complexities raised by cryogenics. The advocates of this technology claim that it is theoretically possible to 'super-freeze' a deceased subject's organic matter, and then latter to 'super-thaw' her body when the technologies exist to bring it to life. The dominant reaction is to suppose that this technology, were it medically possible, would resurrect the original subject. To uphold this verdict, the present constraint would need to explain why a dead and super-frozen brain can still be said to realise a subject's psychological states.

first assimilated into an existing realiser of that subject's mental states.[91] Thus, Agar's refinement of the functionalist view seems to entail the narrow psychological view, on which we persist only if there is sufficiently continuous physical realisation of our capacities for mentality. Though, as we have seen, this view is compatible with the replacement of our neural material with inorganic substitutes, or the realisation of a subject by spatially distant, but communicating, parts, it nevertheless constitutes a significant step away from the wide psychological view.

I have suggested that one's mental states must be continuously realised in order for relationships of psychological continuity to be relevant to personal identity. This view responds well to the deficiencies of the previous responses; it distinguishes between the branch-line case and fission, just as we are inclined to do. It also avoids the concern that processes of teletransportation are too 'extrinsic' to count as good realisers of psychological continuity. However, it remains to be seen whether it fully resolves the problems raised by Fission and the Branch-line case.

We should begin with a challenge. There is no reason to believe that teletransportation is the only means of producing branching psychological continuity. Thus, it is not clear that the requirement that our psychologies are continuously realised is able to ward off all cases in which a single subject comes to be psychologically continuous with two subjects existing at a later time. The following scenario appears to combine branching psychological continuity with the continuous realisation of mental states:

Brain Division My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine. (Parfit 1984, pp. 254–255)[92]

---

91. Agar suggests a counter-example, in which a teletransportation device converts "thought from orthodox physical realization to some bizarre alternative form – perhaps as pure energy, or pure radio signals" (Agar 2003, p. 66), thus creating appropriate successors without the need for non-psychological causal intermediaries (c.f. also the 'falling elevator model' of resurrection floated by Zimmerman 1999; Zimmerman 2010). I think, though, that he is right to suggest that cases like this strain intelligibility; we can therefore set them aside without loss.

92. Parfit provides a neat articulation of a case that was developed by David Wiggins in *Identity and Spatio-Temporal Continuity* (1967, pp. 52–53). Here, one might also consider Williams' case of

Chapter 3: Functional Subjects

We must concede that the suggested modification does not entirely avoid problem cases involving branching psychological continuity. This shows that even continuously realised psychological continuity is not sufficient for personal identity. Have we made no progress?

I want to make two comments which highlight how the debate has been moved along. First, we should note that the Brain Division case is significantly different from Fission and the Branch-line case; the former involves division of the realisers of a subject's mental states, whilst the latter two cases involve their replication. I think that this impacts upon our intuitions about these cases. Whilst it is very easy to say that the original subject remains behind, unaffected, in the branch-line case (and that she ceases to exist in the fission case), it is not so easy to find a satisfactory response to Brain Division; it is hard to say what happens to the original subject. Thus, it strikes me that it is a point in favour of the narrow view that it deals successfully with fission and branch-line cases, whilst it is not a significant demerit that it should have difficulty with Brain Division; we might expect to face individuative difficulties when faced with interference with the realisers of a subject's mental life.[93]

Second, it should also be noted that I did not completely reject either the closest continuer theory or the non-branching requirement;[94] I allowed that both might play a role in a theory of the individuation of material things.[95] Rather,

---

a man who splits in two, like an amoeba, and what he has to say in response to the difficulties raised (Williams 1960, pp. 47–48).

93. C.f. the various cases raised by Unger in Ch. 6 of his 1990, all of which rely upon separating the neural realisers of a subject's mental states, and then playing with the way in which they communicate.

94. In this, the dialectic is slightly different for me than it is for David Wiggins. Wiggins initially, in 1967, accepted something very like the narrow view:

> [We should] analyse *person* in such a way that coincidence under the concept *person* logically required the *continuance in one organized parcel of all that was causally sufficient and causally necessary to the continuance of essential and characteristic functioning, no autonomously sufficient part achieving autonomous and functionally separate existence.* (ibid., p. 55)

The view presented here excludes splitting by something like a no-branching requirement. Later, however, Wiggins rejects no-branching requirements, taking them to allow facts of identity to be affected by overly extrinsic considerations. Thus, eventually, we find that pp. 225–236 of Wiggins 2001 presents an argument against Shoemaker's view brain transplant case (introduced on p. 113 of the next chapter) based upon the prospect of a case like Brain Division. Unlike Wiggins, I would not be quick to conclude that the possibility of split brains invalidates the narrow psychological view.

95. Indeed, given that the majority of material entities can be split in half, or have pieces shaved

I simply argued that they provided inappropriate solutions to cases in which branching psychological continuity is produced by duplicating someone's brain and body. This leaves room for either theory to play a role in determining what happens to a subject when the realisers of her mental states are divided from one another. Though these theories cannot be the whole solution to the problem of branching psychological continuity, they might yet provide part of the solution.

I conclude that cases such as Brain Division do nothing to discredit the thought that our mental states must be continuously realised if psychological continuity is to be relevant to the question of our persistence. The response is attractive, and it does not have the demerits that the non-branching requirement and the closest continuer view have when considered by themselves. It therefore seems to me that the narrow psychological view presents the best hope for responding to Fission and Branch-line whilst retaining an account of personal identity inspired by functionalism. Since this imposes significant constraints on the realisers of personal identity, I therefore conclude that Shoemaker is wrong to look to functionalism in order to motivate the wide psychological view.

3.4   Conclusion

I do not take the above to show that Shoemaker's account of personal identity is untenable. I think that it is ingenious, and that he provides a clearcut account of the notion of psychological continuity to which proponents of the wide psychological approach often appeal. Perhaps that is all that Shoemaker requires of it. However, as noted above, what I take to be promising about Shoemaker's analysis is that it provides an account of a pattern of causal influence which we should, in any event, recognise. The above argument shows that the wide psychological theorist cannot straightforwardly state that she takes our persistence to be secured by this functionally defined pattern. A number of fixes are available. However, I have argued that the most appropriate fix undermines the wide psychological view. Shoemaker's preferred account of personal identity does not, therefore follow in any straightforward way from functionalism in the philosophy of mind.

---

off of them, I suspect that one will have to refer to a theory of this kind (although, for an alternative, much more stringent, style of response, see ibid., pp. 99–102). Fission cases, and the problems that they raise, are easily constructed for almost any kind of material object. Thus, that such cases can be constructed does not, by itself, count against a theory of the identity of any kind of object – what is at issue is whether we can convincingly respond to such cases without giving up the theory.

Chapter 3: Functional Subjects

In what follows, I pursue the idea that accounts of personal identity should pay close attention to the realisers of our mental states. The guiding thought is that our psychological states stand in relationships of unity, and that the existence of these relationships cannot be explained except by reference to enduring relationships between the realisers of our mental states. Since these unity relationships are required if there are to exist any persons, it strikes me that it is reasonable to incorporate these facts into our account of personal identity. This further motivates the view introduced in this chapter.

The Material Self

This chapter presents a final argument against the wide psychological view. The conclusion of the argument is that one should not produce an account of personal identity which abstracts entirely from the fact that our psychological states are realised in enduring physical systems. Of course, very few people seriously deny that our mental states are realised in this way.[96] However, its significance is under-appreciated. What I shall argue, using Barry Dainton's work as a foil, is that even if our persistence can be understood in terms of connections, or potential connections, between psychological entities, one cannot explain the fact that these connections come in enduring and unified packages without looking outside the psychological, and to their realisers. Reflection on the nature of our mental states suggests that it is, at least, typically necessary that they should be continuously realised. Given that we have failed to motivate the wide psychological view, and that this chapter finds reason to emphasize the fact that our mental states are typically continuously realised, it is not a large step to the claim that we persist only if our mental states are continuously realised – the view floated at the end of the previous chapter.

The first two sections introduce some preliminaries. Section 4.1 contextualises 'Mentalism' – the view that an account of the nature of persons can confine itself to the psychological. 4.2 then gives a brief summary of Dainton's 'phenomenal view' of the self. 4.3 introduces a novel thought experiment in order to question what is required for the existence of enduring perspectives like ours – complexes of mental states which are internally unified, and causally isolated from one another. In 4.4, I consider a number of responses, ultimately settling on the simplest – that our mental states are associated with realisers which stand

---

96. For some examples, see Foster 1991, and the essays in Lavazza and H. Robinson 2014.

in stable relationships to one another. 4.5 defends the claim that this should impact upon our accounts of personal identity, developing the observation of the previous section into an argument against the wide psychological view, and considering responses.

4.1   The Pure Ego Theory

To set up the main conclusions of the chapter, it will be useful to begin with a precursor to the wide psychological view. Here, our starting point will be a collection of views that are rarely discussed in contemporary philosophy. These are what came to be known as 'pure ego' views, though in a less recondite turn of phrase they might also be called 'soul' views.[97] As I use the phrase, the pure ego theorist accepts three claims:

Mentalism: If mental events M and N belong to the same person, we can understand what makes this so without referring to anything other than psychological particulars, properties, and relations.

Centering: Mental events M and N belong to the same person in virtue of standing in a particular relationship to a third entity–the 'centre' of the person's mind.[98]

Distinctness: The centre of a person's mind is not another mental event; it is a substance.

Put less formally, the pure ego view says that each experience is closely associated with a mental substance which constitutes the core of its subject's mind. Once one has accepted this position, two further issues must be addressed. First, one must characterise the relationship between pure egos and persons. Presumably, the pure ego is a part of the person. Should we state that it is identical to the person, or that it is only a proper part of the person? Second, one must develop a picture of the relationship between mental events and pure egos – a topic which I feel is best addressed in light of an overall picture of the ontology

97. See, for example, Swinburne 1984 for a theory that is arguably a version of the pure ego view.

98. Alternatively, if the centre of a person's mind is itself composed of mental states (e.g. a mass of bodily feeling as in James 1890, pp. 296–305, and Ayer 1963), two mental states belong to the same mind if each either stands in the particular relationship to its centre, or is among the centre's constituents.

of mental events.[99] Both of these decision-points raise interesting questions, and investigation of them may prove relevant to an overall assessment of the pure ego view. These issues, however, lie outside of the scope of our discussion; I raise the example of the pure ego theory in order to set the ground for an alternative. I shall assume, in what follows, that we are right to reject this theory.

4.1.1   An Alternative View

Today, the pure ego theory has largely fallen out of fashion. Pure egos are mysterious entities, and it has seemed to many that it is superfluous to posit them in order to explain how more than one mental event might belong to a single person. Particularly influential here has been Hume's claim that we cannot perceive pure egos, and that we therefore ought not to postulate their existence:

> 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. (Hume 1739/1985, Book I, §VI/p.300)

Hume's statement of opposition is of course not uncontentious.[100] However, it has been historically influential, and I wish to consider an alternative tradition which has developed in its wake. This tradition accepts mentalism, but rejects the distinctness claim. Members of this tradition are divided about whether to accept or reject centering. A common feature of all views of this kind is the thought that the self is constituted by a number of independently existing mental items, standing in appropriate 'unifying' relationships to one another.[101]

---

99. What I have in mind here is the choice between taking mental events to be independent particulars (what Broad calls 'existents'), taking them to be non-independent modifications of the centre (what Broad calls 'subsistents'), or taking them to be complexes with the centre as one constituent. See Broad 1925, pp. 558–568 for a taxonomy of pure ego views.

100. Thus, for example, Russell and McTaggart both claim that it is possible to be introspectively aware of oneself (Russell 1910/11, pp. 110–111; McTaggart 1927, pp. 62–63), and Broad makes remarks which suggest that he is sceptical of Hume's contention (Broad 1925, Ch. 6, esp. pp. 280–282).

101. C.f. Hume:

> [E]very thing, which is different, is distinguishable, and every thing which is distinguishable, is separable by the imagination ... since all our perceptions are different from each other, and from everything else in the universe, they are also distinct and

Chapter 4: The Material Self

Hence, even if there is a centre, it is (or its constituents are) of much the same kind as the other mental events which belong to the person. I want to argue that such a view does not provide a sufficiently robust explanation of the unity of our minds.

Historically, there are close affinities between the wide psychological view and this alternative form of mentalism. Philosophers who adopt the wide psychological view have been inspired by Humean scepticism about the idea of a primitive psychological substance, responsible for each of our experiences, and by Locke's denial that our persistence is to be understood by reference to such an entity. Both Locke and Hume are mentalists; there is no question in the accounts that they, and their early successors, develop that the self is to be understood in purely psychological terms. This has left its fingerprint on contemporary advocates of the wide psychological view; though such thinkers will often acknowledge that we must be constituted by some material entity whenever we exist, and that we have physical properties, these facts are entirely inert in their discussion of personal identity. Even if something like distinctness holds, it plays no role in accounting for our persistence through time. Instead, persistence is analysed exclusively through the lens of psychological relations between different stages of a subject's career. I think that this is wrongheaded.

My diagnosis of the issue is that my opponent has rejected the wrong element of the pure ego theory. The pure ego theorist was correct to look beyond mental states in order to explain their belonging to a subject. She was wrong to think that this called for the postulation of yet another mental item.[102] Instead, we must reject mentalism, and explain the ownership of mental states by reference to the physical systems which produce them.

---

separable, and may be consider'd as separately existent, and may exist separately, and have no need of any thing else to support their existence. (Hume 1739/1985, Book I, §V/p.282)

102. An important subtlety: I do not claim, as the pure ego theorist might, that the concept of an experience must be analysed in terms of its centre. My claim is merely that something like the following quotation of Broad's must be true:

[T]he subsistence of certain observable relations between a set of mental events might entail that there was an Existent Centre to which they all stood in a certain common relation (1925, p. 571)

In other words, the fact that a person's experiences are all phenomenal co-conscious indicates that there is a physical system to which they are all related.

4.2 Dainton's Views

I am going to argue that the unity distinctive of a person's mind can only be explained if we look beyond the psychological. To do this, it will be helpful to have an opponent in view. I will introduce Barry Dainton's views for this purpose.[103] Dainton's views are rigorously developed, and explicitly pursue the idea that our mental states are unified by psychological relationships internal to conscious awareness. For this reason, he presents a perfect foil for our discussion of mentalism.

Dainton's view is primarily motivated by the intuition that it is sufficient for you and I to continue to exist that our streams of consciousness continue – that future experiences come into being which stand to our present experiences as our present experiences do to those in our immediate pasts. Given this, he suggests that we define our persistence conditions in terms of phenomenal relationships between experiential items:

> [P]rovided our streams of consciousness flow on, we can easily envisage surviving total ruptures in material and psychological-cum-causal continuity. All this suggested a way forward: why not try to construct an account of our persistence and existence conditions entirely in terms of phenomenal unity? (Dainton 2008, p. 73)

The fact that he thinks we can envisage surviving 'ruptures' in material continuity shows that he is a proponent of the wide psychological view. His view differs from Shoemaker's insofar as he does not analyse persistence in terms of causal relationships between the beliefs, desires, and other mental states that a subject exhibits at different times. Instead, his account focuses on a phenomenal relationship connecting our experiences at different times – what he calls 'co-consciousness'. It is to this that we now turn.

4.2.1 Co-consciousness

Dainton uses the term 'co-consciousness' to describe the "quite distinctive, wholly phenomenal" unity characteristic of our experiences (ibid). Co-consciousness comes in two forms – diachronic, which accounts for a subject's

---

103. I will draw predominantly from Dainton's *The Phenomenal Self* (2008). Dainton also develops his 'C-theory' in his 2004; 2012; 2015; 2016, and, working with Tim Bayne, in Dainton and Baine 2005. An engaging critical response is given by Snowdon in his 2016.

persistence through time, and synchronic, which describes the unity of her consciousness at any moment in time. Both forms of co-consciousness are 'primitive' on Dainton's view, admitting of no characterisation in terms of simpler features of, or relations between, experiences. Nevertheless, by careful use of example, he thinks that he can indicate to the reader what is meant by both terms. Let us begin with diachronic co-consciousness, which will be the more important notion in what is to come.

The key feature of experiences that are diachronically co-conscious is that they fall within a single continuous stream of conscious awareness, either fading into one another directly (direct diachronic co-consciousness), or fading into one another via intermediary experiences (indirect diachronic co-consciousness). When successive events are directly diachronically co-conscious, Dainton argues that there is a sense in which they are genuinely experienced together, in virtue of falling under a single 'specious present' (Dainton 2008, pp. 51–56) – a temporally extended duration consisting of events which fall under one's immediate awareness. Dainton outlines this with a few illustrative examples, which he supposes will be familiar to his readers:

> Think of what it is like to listen to a police siren gradually recede into the distance, or to see a ship slowly drifting across a bay, or to feel the burn of a strong mint on one's tongue, or to run one's finger along a smooth slab of marble. In these cases, and untold others, the continuity in and of our experience is obvious. (ibid., p. 50).

This quotation clarifies the notion of 'direct co-consciousness'. To explain the notion of indirect diachronic co-consciousness, Dainton relies upon a model invented by John Foster (1979, pp. 175–177); we briefly encountered a version of this model in chapter 2. Foster's model has it that any experience has duration. As such, it has a number of temporal parts, at least some of which are temporally close enough to one another that they fall under a single specious present (and so are experienced together). Successive specious presents temporally overlap. As such, though they do not exactly coincide, they share some of their temporal parts. When two specious presents overlap in this way, one does not experience their non-overlapping sections as flowing into (or out of) one another. One is, however, aware of experiences in the earliest part of the first specious present flowing into those contained in the later overlapping section. In turn, these experiences are felt to flow into the later section of the second specious present. This suffices for the subject to experience the events as taking place within a sin-

gle continuous period of conscious awareness. Dainton suggests that we ought to take this form of indirect connection to be sufficient for identity.

The relationship of synchronic co-consciousness characterises the experienced togetherness of simultaneous experiences – those parts of a specious present that do not precede or succeed one another, but instead accompany one another. Yet again, Dainton clarifies this relationship with a number of examples:

> We all know what it is like to a see a cat crossing a road; we all know what it is like to see a cat crossing a road while hearing a plane pass overhead. In the one case there is a visual experience, in the other there is a visual experience and an auditory expeirence, and the latter experiences are experienced as occurring together in a single episode of consciousness. (ibid., p. 34)

Whilst diachronic co-consciousness is not transitive, Dainton suggests that synchronic co-consciousness is:

> [I]t is all but impossible to conceive how [synchronic] co-consciousness could fail to be transitive. If experiences E1 and E2 are unified by co-consciousness they form, in effect, a single experiential state, all of whose parts are fully and mutually co-conscious with one another, and as a consequence it is not obvious that an additional experience could by co-conscious with one part of this state without also being co-conscious with the remainder. (ibid., p. 50)

For present purposes, this brief sketch will suffice to introduce both synchronic and diachronic co-consciousness. What is interesting, at least to Dainton's mind, about co-consciousness is that the phenomenal unity of our experience is evident 'from the inside'. This leads him to the claim that a stream of consciousness is unified by phenomenal relationships which are internal to the stream. Given this, and the claim that we continue to exist if our streams of consciousness continue, co-consciousness represents a promising point of departure for a phenomenalistic account of the self.

4.2.2 The Bridge Problem

A promising point of departure, sure. However, a key problem faces anyone who wants to understand persistence in terms of phenomenal continuity – what Dainton calls the bridge problem (ibid., p. 75–81). Imagine waking up after

a dreamless sleep. Our waking experiences do not directly continue the experiences we had before falling asleep. Thus, these experiences are not directly diachronically co-conscious with one another. Given the absence of experience during our imagined sleep, there is no ongoing chain of experiences connecting the moments before sleep to our first waking moments. Thus, the experiences before and after sleeping are not indirectly diachronically co-conscious either. Yet, nevertheless, the very same person wakes up as previously fell asleep. Diachronic co-consciousness is not, in either its direct or indirect form, necessary for personal identity.

The problem is this:

> Co-consciousness (in its direct or indirect forms) may supply us with a sufficient condition for the consubjectivity of experiences, but (on the face of it, at least) it does not amount to a necessary condition. Since experiences in distinct streams separated by a gap in time are neither directly nor indirectly co-conscious, [co-consciousness] provides no guide as to the circumstances in which such streams belong to the same self. (Dainton 2008, p. 75)

Yet again, Dainton follows Foster's lead in response to the bridge problem (Foster 1979, pp. 179–181);[104] what makes a waking experience belong to the same person as an experience just before sleeping is that the waking experience (or an appropriate counterpart) could easily have been diachronically co-conscious with the experience just before sleep; had the subject stayed awake, the two experiences (or appropriate counterparts) would have occurred within just one stream of consciousness.[105] These experiences are, to use Dainton's turn of phrase, 'potentially diachronically co-conscious', though they are not actually diachronically co-conscious. The obvious solution to the bridge problem, then, is to assert that two experiences belong to a single subject if and only if they are

---

104. It's worth noting, however, that Foster (at least in this iteration – to my mind, his later 1991 revokes this) acknowledges the need for a relationship between mental and physical items in order to account for the unity of the mind.

105. This remains the case even if there is a maximum time period that a subject can stay awake (and even if the time spent asleep exceeds this maximum time period). In such a case, the subject's last waking experience (or an appropriate counterpart) could have occurred within an alternative series of experiences whose last member (or an appropriate counterpart) occurred later than the subject (in fact) fell asleep. The last member of that stream (or an appropriate counterpart), in turn, could have taken up a position in a different stream of conscious experience which extended even later into the period of unconsciousness. By this process, then, it is possible to reunite our subject's waking experiences with her experiences just before falling asleep.

potentially diachronically co-conscious.

With these preliminaries noted, I shall now present a brief summary of Dainton's account of personal identity.

4.2.3 The Phenomenal Account

Dainton's favoured account of personal identity focuses upon relationships between experiential powers – capacities for particular kinds of conscious experience – rather than directly upon inter-experiential relationships (2008, pp. 86–87). This is a direct result of the introduction of potential co-consciousness. It is plausible that it cannot just be a brute fact that two experiences are potentially diachronically co-conscious; something must ground the fact that the earlier experience would, if extended, join up with the later. Instead, Dainton suggests that whenever two experiences are potentially diachronically co-conscious, this is explained by the presence of relationships between a series of 'experiential powers' such that, had those powers been active at the right times, the two experiences would have been (indirectly) diachronically co-conscious.

The ownership of experiential powers, like that of experiences, is defined in terms of the two primary forms of co-consciousness, diachronic and synchronic. To a first approximation, two experiential powers belong to a subject at a single time if they are disposed to produce synchronically co-conscious experiences, two powers belong to the same subject at different times if they are disposed to produce diachronically co-conscious experiences.

To reduce to a minimum the technical apparatus required to convey Dainton's theory, I have provided my own summary of the main principles which determine whether or not two experiential powers belong to the same subject:

1. Powers α and β if both active at t_1, belong to the same subject (at t_1) if and only if the experiences they produce are synchronically co-conscious.

2. Inactive power α belongs to the same subject as power β (whether or not β is active) if and only if:

   a) Were both powers active, they would belong to the same subject by 1.

   b) Both powers are linked by a chain of powers, adjacent members of which belong to the same subject by 1. and 2. a.

   c) Were α and β to be activated in short succession, the experiences they would produce would be diachronically co-conscious.

3. Power α belongs to the same subject at t_1 as does power β at t_2 if:

    a) α is active at t_1, and β is active at t_2, and the experiences they produce at these times are diachronically co-conscious.

    b) Were α active at t_1, and β active at t_2, they would belong to the same subject by 3. a.

    c) Both powers are linked by a chain of powers, adjacent members of which belong to the same subject by 3. a. or by 3. b.

4. If α and β belong to the same subject by 1. or 2., and β and γ belong to the same subject by 3., then α and γ belong to the same subject.

This summary skips over some important details; I do not, for example, consider the possibility that an experiential power might produce more than one non-co-conscious experience at a single time (p. 128). Nor do I consider what happens when the dispositions that an experiential power possesses are probabilistic (50% of the time, it would produce an experience co-conscious with those of power α, 50% of the time, it would produce an experience co-conscious with those of power β) (p. 129), or when they vary depending on which other powers are active and which are not (e.g. pp. 103–111). These raise interesting points. However, they are tangential to the main thrust of this chapter. I want to argue we should look to connections between experiential powers and the physical world in order to make sense of the existence of subjects who, like us, have complex psychologies. When pushed to this position, then it is unclear why we should continue to pursue a purely phenomenological analysis of the self.

4.2.4   An Initial Tension

We can begin with the thought that there is an element of tension in Dainton's account. On the one hand, he begins with the phenomenally obvious – conscious experiences and relationships that are internal to consciousness. On the other hand, these relationships alone are insufficient to bring together the successive phases of a subject's life. In other to fully unite a subject's experiences, we must look to dispositional relationships between experiential powers. Such things, since they are not directly observable, may sit uneasily with those of a Humean bent, whose key reason for pursuing the theory glossed above was its promise to dispense with unobservable entities and relations underlying our experiences.

Even setting this aside, however, there are reasons to think that the tension in Dainton's account causes it trouble.

What are experiential powers? What individuates them? What determines whether or not two powers are disposed to interact? Dainton's answer is unambiguous. In the actual world, at least, experiential powers are realised in matter:

> I am assuming that all experiential powers (in our universe at least) are grounded in some physical system, their power-base, as we can call it. If two token experiential powers existing at the same time are of the same type, to be distinct they must have different power-bases, they must be grounded in different physical things, or different parts or aspects of they same physical thing. (Dainton 2008, p. 120)[106]

A person is associated with a number of experiential powers, embodied in matter, and standing in an appropriate, materially realised, relationship to one another. I am inclined to believe that this fact is ontologically significant; it explains how there could be persons. I am also inclined to think that this should impact upon our persistence conditions. The following sections outline an argument for this claim.

4.3   Fragmentation and Recombination

Dainton does not deny that there is a close relationship between experiential powers and the material world. Nevertheless, his account references only experiential powers and (experientially grounded) relationships between them. It is this which qualifies it as a form of mentalism. I wish to argue that a more illuminating account should mention more than this; certain explanatory questions are best answered by explicitly acknowledging that our material constitution plays a significant role in giving unity to our experiential powers.

The core problem with Dainton's account is that co-consciousness alone only creates a thin kind of unity – one whose presence, and whose persistence, re-

---

106. Indeed, Dainton says much the same about experiences:

> I will work on the less theoretically charged assumption that experiences owe their individuality to their precise phenomenal character, their time of occurrence and their *physical basis*. If E1 and E2 are two simultaneously occurring experiences with the same phenomenal character, to be distinct they must be grounded in different physical systems, or different parts or aspects of the same physical system (Dainton 2008, p. 30)

quires deeper explanation. I will illustrate this through a series of thought experiments. At the outset, though, it is worth noting that I do not claim that these thought experiments are possible (indeed, part of my point is that they are not, at least in the absence of undue interference with the matter which composes us). In fact, I think that the most far flung of these thought experiments is nigh on inconceivable (albeit not significantly more so than Dainton's varieties of virtual reality, which I shall discuss below). My claim is simply that an account of personal identity will benefit from accepting resources other than those available to a mentalist like Dainton.

The first thought experiment is relatively straightforward:

Inactive Power Exchange Certain gustatory capacities – key among them capacities to detect or to enjoy certain flavours – are highly valued. Imagine that I have an extremely refined palate. A wealthy gourmet wishes to enhance her enjoyment of Japanese food. She pays me to undergo a simple procedure; my capacity for tasting umami flavours is swapped for hers. Whilst this procedure occurs, we are each conscious, staring at an operating theatre wall. Afterwards, I go to make a favourite broth. I am shocked by the deadening of the subtle flavours, as well as by how overwhelmingly salty the meal now is. The gourmet, on the other hand, goes out to dinner, and reports finding it astoundingly tasty.

This scenario is a little strange (in particular, the lack of description of the medical procedure might make us wary to draw any conclusions – for now, we can let this slide). A single experiential power is initially potentially coconscious with the remainder of my powers. It then becomes potentially coconscious (indeed, actually so at the dinner) with the gourmet's other experiential powers. This process interrupts neither of our streams of consciousness.

If the change-over of experiential powers is an isolated occurrence, Dainton's account looks pretty good. However, two modifications will serve to tighten the screw on his account. First of all, let us imagine that the transferred experiential power is active throughout the procedure:

Active Power Exchange$_1$ The wealthy gourmet now asks for my capacity to detect tannins in liquids. As an experiment, she proposes that we both drink a glass of wine as we swap capacities. Whilst we stare at the wall, we each

savour our glass of wine. I begin by tasting a dry rioja. Midway through the operation, however, I cease to notice the puckering sensation in my cheeks – the astringency of the wine becomes imperceptible. Simultaneously, the gourmet becomes excited at the detection of features of the wine that were hitherto beyond her.[107]

In this scenario, an experiential power undergoes a transition. At first, it produces experiences which are synchronically and diachronically co-conscious with those produced by the remainder of my experiential powers. In a moment, however, this all changes; it (one and the same experiential power) comes to produce experiences which are synchronically and diachronically co-conscious with those produced by the gourmet's experiential powers. This scenario is a little harder to imagine, but still strikes me as coherent. My experience is akin to what occurs when, half a glass down, the wine's flavour begins to deaden; the gourmet's experience is akin to that which occurs when a trained friend tells me what to look for in a wine. However, we may place the scenario outside the bounds of easy imaginability by imagining that the experiential capacity is not only active throughout the thought experiment, but also that it is the source of its own diachronically unified stream of experience:

Active Power Exchange$_2$ At the same time as the above changes occur, there also occurs an experience as of tasting a good light wine whilst staring at the wall from one position, and savouring the experience as one's last chance to appreciate the wine's more complex notes. This experience flows into that of feeling fuzziness in one's cheeks for the first time, and having one's anticipations of tasting the wine anew met by a fuller knowledge of its character (both of which are accompanied by a perception of the wall from a new position).

Now, I do not dispute that it is hard to imagine such a sudden and all-encompassing change in one's experience (indeed, as noted in footnote 107, it is likely that nothing, not even the experience of the tannicity of the wine, will

---

107. It is important to note that it is no part of this scenario that the gourmet's experience of the wine post-operation will be the same as my experience of the wine pre-operation. Dainton holds (2008, pp. 264–307), and I accept, that experience has a holistic character; the experiences produced by any one experiential power will be affected by the other experiential powers that are active at the very same time.

be entirely constant throughout the experiential transition). It is worth noting, however, that Dainton supports his views by reference to forms of virtual reality which present us with a similar imaginative challenge:

> [W]hen a subject is connected to [a VR-3 device], not only is a completely realistic virtual environment generated [by direct brain stimulation], as previously, but the subject is also furnished with new memories, beliefs, intentions, personality traits, etc. – a whole new psychology, one whose constituent states are not causally dependent upon those they replace . . . On being informed that they can expect to experience nothing more than a few moments of mild anxiety and bewilderment, most VR-3 users understandably opt to remain fully conscious throughout the transition. (Dainton 2008, p. 15)

In this scenario, a single subject's experience persists through a change in many of the experiential powers which generate it. This indicates that what we might call the pure logic of co-consciousness, at least as it is envisaged by Dainton, gives us no reason to object to Active Power Exchange$_2$ (unlike, for example, a scenario in which two experiences are synchronically co-conscious with a third, but not with one another). To finalise the challenge to Dainton's account, let us imagine (or, if this proves impossible, pretend to imagine) a fourth case. This case involves a world in which experiential transitions are widespread, and do not require surgical intervention. In such a world, powers form flimsy alliances, before separating and courting others. Throughout this process, they retain the capacity to generate conscious experience. Call this a world of 'fragmentation and recombination'; it is a world in which each experiential power carries with it a perspective of its own, and conspires with other powers to supply that perspective with a varied and ever-changing flow of experience.[108]

I do not think that such a world is possible. Nor do I think that Dainton is committed to thinking that it is possible. What I do think, however, is that this gives us our best description of a world in which experiential powers are unified only by the relationships which Dainton identifies – relationships of po-

---

108. It is worth noting that we can imagine even more exotic worlds than this. Relationships of diachronic co-consciousness can hold between experiences produced by numerically distinct experiential powers (for example, imagine an experience of a completely red visual field being overtaken by an experience of a completely green visual field). Thus, we could add in relationships of diachronic co-consciousness between entirely disjoint series of experiential powers, and indeed also allow that successive operations of a single experiential power might fail to be diachronically co-conscious.

tential and actual diachronic and synchronic co-consciousness. This world, I will argue, could not contain anything like the subjectivity characteristic of persons. To understand how persons exist, we need to consider why fragmentation and recombination do not occur.

4.3.1 The Cohesion Problem

In fragmentation, a single stream of consciousness 'splits' into a number of descendent streams. This is mediated by the separation of the experiential powers which hitherto sustained the initial stream of consciousness. Further, each successor stream soon forges new connections with other, already active, experiential powers, giving rise to a lattice-like structure of merging and separating streams of consciousness. This subsection argues that complex mental capacities, like the ones that we enjoy, could not exist in a world of fragmentation and recombination.

I begin by clarifying a couple of important differences between the envisaged scenario, and other cases that are frequently discussed in the literature around personal identity. The example of fragmentation and recombination superficially resembles a case which will be more familiar from the last chapter – that of fission. Indeed, Parfit's ingenious 1971 'Personal Identity' contains a case of separating and recombining subjects superficially similar to that our case of fragmentation and recombination:

> [L]et us now turn to a second kind of being. These reproduce by fusion as well as by division. And let us suppose that they fuse every autumn and divide every spring. (Parfit 1971, p. 22)

Parfit's imagined creatures periodically divide. Whenever a creature divides, it produces two descendants with exactly similar psychological features. When two such beings fuse, they produce a descendent whose psychological features are intermediate between theirs.[109] Fission cases have been much discussed in the literature surrounding personal identity, and there is a good case for thinking

---

109. More precisely, Parfit writes:

> Some of [the personality traits of those who fuse] will be compatible. These can coexist in the one resulting person. Some will be incompatible. These, if of equal strength, can cancel out, and if of different strengths, the stronger can be made weaker. And all these effects might be predictable. (Parfit 1971, p. 18)

that they plague any account of our persistence. Thus, one might think that the above case shows nothing new; Dainton may reply dismissively simply by wheeling out his preferred account of fission.[110] I think that this response is unsatisfactory for a number of reasons.

There are two significant disanalogies between Parfit's scenario and the world of fragmentation and recombination. First, Parfit's creatures separate at regular time intervals. Second, they inherit a full suite of psychological capacities from their forebears; each fission descendent inherits a perspective much like her immediate ancestor's. Both of these features are potentially relevant to our assessment of the case; rather than splitting into parts that are then recombined, it is natural to think of Parfit's creatures as reproducing, or indeed as continuing their histories in more than one body.

In contrast, in a world in which recombination frequently occurred, there would be no guarantee that experiential powers should recombine into large groups of interacting psychological capacities, rather than into more primitive clusters of experiential powers. Even if they did, however, it seems difficult to understand how those capacities could form a coherent personal perspective. Persons typically have a number of deeply interdependent capacities; my capacity to experience shame, for example, is plausibly tightly connected with my experience of myself as a social being, and my awareness of my own values (or those of my community), as well as a whole battery of more basic affective responses (e.g. the capacity to feel exposed, and threatened). Were experiential powers simply to recombine at random, there is no guarantee that these different capacities would come together to generate a being capable of an experience like shame. Thus, it would be incredibly unlikely that a world of fragmentation and recombination would allow for the types of psychological states that persons typically possess. This marks it apart from the case of fission.

A further point might drive home the difficulties with the scenario of fragmentation and recombination. Persons do not simply come with a standard battery of psychological capacities; many of their characteristics also derive from a very particular development of their more basic capacities. This introduces fur-

---

110. Dainton's preferred account is excitingly idiosyncratic. Dainton argues that lifetimes need not be linear; a subject might inhabit a branching personal timeline. Given this, he thinks that we should think that a fission case involves a single subject coming to occupy two bodies. Dainton also uses the same mechanism to argue that a single subject (a 'hyper-self') might occupy more than one non-intersecting personal timeline (2015). Another response along the same lines, although notably different in motivation is given by Wollheim 1979.

ther dependencies between different aspects of a person's psychology. Consider, for example, a lone quasi-memory – an impression exactly like that of a memory experience, but produced by a person's experience other than your own.[111] Wollheim and Shoemaker both make the point that it would be difficult for such an impression to be integrated with one's psychology in the way typical of our memories:

> [D]ispositions form, within the psychology of the person, a web or network: they are ancillary to one another, and there seems no method for determining what in the way of other dispositions, what in the way of beliefs, emotions, desires, fears, and other memories, would have to be transferred along with it if the original memory is to run across persons. It seems plausible to hold that, if I am to Q-remember my father's childhood walks, I should also have to have, and to have because he had them, a native speaker's knowledge of German: a capacity to imagine intense cold: a sense of the aspirations of a later nineteenth-century Central European schoolboy: a familiarity, which did not depend upon something I had been told, with the details of my father's family, with the books he would have read, with the thoughts he would be inclined to have when he looked up into the sky, or smelled the smell of soup, with the religion, if any, in which he was brought up, and many other such dispositions which would be backgrounded, if not foregrounded, in the Q-memory. (Wollheim 1984, pp. 113–114)

> Although Jane's seeming memory [of a white Palladian church] is causally due to Paul's experiences in Venice, its causal link to them is not part of a richer causal linkage that constitutes psychological continuity of the sort that characterizes ordinary cases of personal identity (Shoemaker 2004a, p. 581)

In a world of recombining experiential powers, there would be no guarantee that groups of clusters would combine to make a coherent perspective. Yet again, then, it would be incredibly unlikely that such a world could contain entities like us, whose psychological lives meld together a wide range of more basic psychological capacities.

We are not just subjects of experience; we are also persons – a class of subjects distinctive in virtue of the particular psychological capacities they develop. Persons like us can only exist if fragmentation and recombination is, at best, a

---

111. I briefly introduced quasi-memories in footnote 77 of the previous chapter (p. 70).

rare occurrence. I will now turn to explanations of why fragmentation does not occur, before outlining the relevance of this to accounts of personal identity.

4.4 Responses to Fragmentation

It is hard to see how the world of fragmentation and recombination could allow for the type of lives that we now lead; were experiential powers to separate and recombine at random, it should be nigh on impossible for them to combine in such a way as to produce a coherent perspective (or, indeed, very many of the psychological capacities which we enjoy). This raises what I shall call the 'cohesion problem'; an account of our existence must explain how there could be stable, interdependent complexes of psychological capacities (experiential powers included). I will now examine a few attempts at such an explanation, and to argue that our best account points towards relationships between the realisers of our experiential powers.

4.4.1 Experiential Unities

At a first pass, the cohesion problem asks us to explain why fragmentation and recombination do not occur. One possible explanation holds that two groups of experiential powers can only produce diachronically co-conscious experiences if they significantly overlap; if, in other words, both groups contain many of the same powers. If this is correct, fragmentation would not produce a lattice of relationships of diachronic co-consciousness; it would instead simply produce short-lived subjects.

The claim at hand is that a complete, or near-complete turnover of experiential powers would be so disorientating as to preclude the possibility of a single stream of consciousness enduring throughout. There are a few things to say in response to this claim. First of all, it is worth noting that some of Dainton's thought experiments are clearly inconsistent with this principle. As we have seen, in VR-3, a subject's neural structures are rewired to provide her with a new personality. This scenario seems to involve the near-complete turnover of experiential powers.[112] Dainton allows that these experiments would initially

---

112. Even if it does not, Dainton's VR-4 (2008, pp. 18–20), since it leaves the brain almost untouched whilst translating a subject's stream of consciousness into a computer, certainly does involve the complete turnover of experiential powers.

result in great disorientation (p. 17), but nevertheless claims that it is conceivable that one and the same subject could persist through such a scenario. Indeed, it is exactly possibilities like this which initially motivated his account as against its alternatives. As such, I worry that this response would undermine Dainton's account of selves, rather than save it.

This aside, there is a deeper question to ask about the response under consideration. Even if we allow that it is true, and that the world of fragmentation would not involve phenomenal continuity, does this help Dainton's account at all? I think it is clear that it does not. Merely observing that the world of fragmentation would not include continuing perspectives does nothing to explain why perspectives so complex and enduring as ours can exist. The answer to the cohesion problem should be focused on explaining this, not merely explaining why fragmentation is inconsistent with any kind of continued perspective. Hence, the response under consideration is not relevant to the problematic of this chapter. We may set it aside.

4.4.2   Power Individuation

Another response, similar to the above, looks to the metaphysics of powers. It might be claimed that experiential powers only exist in clusters – groups which co-operate to produce a stream of consciousness. If so, a power can leave a group only by being destroyed.[113] Once again, this would mean that fragmentation is impossible; powers would not persist outside of their original groups, and so could not find each other and come to cooperate.

This response falls afoul of precisely the same points as we noted above. It is similarly inconsistent with some of the examples that Dainton gives, such as those concerning simple subjects which consist of only a few experiential powers:

> Maggot has three sensory organs – an eye, a nose, and his skin – each of which provides him with a limited range of simple senses. (Dainton 2008, p. 88)[114]

If he can imagine such subjects, it is difficult to see why he should resist the possibility that one or two experiential powers might separate from a larger

---

113. Or, alternatively, by being a member of a large enough 'splinter group' of capacities that go their own way. This is not relevant to the current context, and so we may set it aside.

114. Dainton also defends the possibility of simple subjects at the beginning of chapter 8 of the cited book.

system of powers, and thereby come to constitute such a subject. It is, therefore, difficult to assess the impact of this response on Dainton's approach.[115] Second, even if it is available to him, it should only increase our wonderment. How could any powers stick around for long enough for there to exist subjects of any degree of complexity? Granted, if any experiential powers are to persist, they will do so housed within stable, long-lived clusters, but how could any such clusters exist? A response based on the metaphysics of powers may be right to claim that the world of constant recombination is impossible, but this simply makes stark just how much explanatory work is required before a proposal such as Dainton's can be considered complete. It will, in other words, weaken, rather than strengthen his project.

4.4.3   Location and Laws

I want to set aside just one more style of response. This proposal is inspired by the following remarks by Broad:

> [I]f we want to make up a theory of mental events and substances analogous to that which we have suggested for material events and substances, we must assume a third determinable Positional Quality which we might call the quality of "Mental Position". (Broad 1925, p. 599)

> Then, if the determinate qualities under the determinable of Mental Position form a manifold of more than one dimension, a mind may be analogous to a body and may have something analogous to size and shape. (ibid., p. 601)

Broad postulates that mental events (and we might extend this to include experiential powers) occupy a kind of 'mental space'; each has a position in that space, by reference to which it is individuated. Experiential powers that are adjacent to one another interact to produce co-conscious experiences, and also adhere to one another; unless external force is applied, a series of adjacent experiential powers will not fragment. Thus, facts about mental space, and about

---

115. A further point: as I noted above, experiences can be diachronically co-conscious despite depending on entirely different experiential powers. If this is the case, then a close cousin to the world of fragmentation could occur. This would be a world in which any experience followed on from, and was followed by, a vast array of experiences, each of which was produced by an entirely different collection of experiential capacities. Though powers would not persist for long periods of time, this world would still contain a frantic series of ever-changing experiences.

the relationships between powers located in mental space explain why fragmentation cannot occur. They simultaneously explain the endurance of one's perspective, and so provide an adequate answer to the cohesion problem.

Interesting as this proposal is, I find the postulation of mental space a little strange; this strikes me as unnecessarily ontologically profligate. Of course, speaking of space here is merely metaphorical; strictly speaking, Broad only postulates a range of so-called 'positional' properties – properties which define for each object a position relative to other entities (just as pitch can be used to locate sounds relative to one another in 'musical space'). Even this, however, introduces us to properties which we have no other reason to postulate. One might wonder whether there is a more parsimonious way to achieve the same result.

In fact there are two. Broad postulates mental positional properties so as to model his proposal upon an analysis of the individuation of physical objects are which dispenses with a substantialist notion of space. We might either modify Broad's proposal to locate experiential powers alongside physical properties, or we might reject entirely the use of 'space' to explain relationships between experiential powers. I shall take each proposal in turn.

The first proposal places experiential powers in 'ordinary' space alongside material entities and events. The reason that fragmentation does not occur is that experiential powers, when they are located close to one another, cohere. This proposal has a definite advantage above the proposal considered above; it does not ask us to postulate an entirely new set of positional qualities. However, one might ask where each experiential power is located, and whether there is an explanation of why it is located there, instead of somewhere else. If we have to locate powers in space, I think that we should locate our experiential powers where we are, and say that they move when we move (indeed, that when they move, they do so because of the movement of our bodies). To say this, however, is to say that powers adhere not just to one another, but also to physical bodies. This is to admit that there is an especially intimate connection between conscious experiences and their physical realisers – one which may itself require explanation.

The alternative is to dispense with the idea of positional qualities altogether. Experiential powers are not individuated by reference to their positions in mental space. Nor do positions in mental space explain how they interact. Nevertheless, we might keep hold of the idea of a type of adhesion between powers

which explains why they do not separate. We might claim that relationships of potential co-consciousness, although purely mental, are enduring; once formed, they are not easily undone. In effect, this would be to postulate a law of attraction between interacting experiential powers. To sustain this proposal, we should say that this law is primitive, and cannot be explained in simpler terms. This is, I think, the best that my opponent can do, and I do not know what can be said against it. However, once we've reached this point, I would simply like to point to the obvious alternative – the view that says that the enduring unity of our mental powers is connected to the enduring unity of a material system which realises them. Overall, I prefer this commonsensical response to any more exotic account of mental space or laws of attraction between experiential powers.

4.4.4 Stabilisation and the body

The problem with the foregoing responses is this: the world of fragmentation and recombination features powers whose unity is unstable – collections of powers break apart and reform at will. The first two of the responses considered above explain why unstable combinations of powers cannot exist. In doing so, they explain why the imagined scenario is impossible. They do not, however, explain how stable collections of powers could exist instead. As such, they fail to explain how our existence is possible. The final response is better, but requires us to postulate spaces or laws for the mental world. I want to consider an alternative way in which experiential powers could combine in a relatively permanent fashion.

It strikes me that there is an obvious response to the cohesion problem. Our experiences are mediated by neural systems, predominantly located in our brains. When stimulated, these produce sensory experience. Whatever the metaphysics of experience, and of experiential powers, it is undeniable that powers are closely connected with the relevant neural structures. And, as brains are stable, long-lasting physical entities, it follows that the experiential powers associated with them will also remain in existence, and continue to stand in the appropriate relationships to one another for a decent amount of time. The risk of fragmentation is warded off (indeed, does not arise) when we recognise the intimate relationship between experiential powers and persisting structures in the physical world.

I above remarked that it is difficult to know what experiential powers are supposed to be, how they are to be individuated, and how they should be disposed to interact, without seeing them as somehow tied to the physical world. I have tried to focus this puzzlement by explaining a key aspect of the unity of our experiential powers which cannot be accounted for from within consciousness. Summarised, the key points are as follows:

1. A person has a complex psychological make-up.

2. In order for beings with complex psychological make-ups to exist, their experiential powers must stand in enduring relationships to one another.

3. If powers stand in enduring relationships to one another, there must be some explanation of why this is so.

4. The best explanation factors in a relationship between the physical world and a subject's experiential powers; powers are associated with (or realised in) enduring physical structures.

Co-consciousness alone cannot provide an explanation of how there could be long-lasting systems of experiential powers. Rather, to do so, we must recognise a close relationship between experiential powers (or, more minimally, the experiential powers of persons) and the physical world. I will now argue that this fact should be reflected in accounts of our persistence.

4.5   Fragmentation and Personal Identity

Dainton's account of personal identity makes no reference to the physical realisers of our experiential powers. I have argued that the fact that our experiential powers are materially realised explains how there could exist beings with complex and integrated psychological lives like ours. Our existence involves more than relations of diachronic (and potential diachronic) co-consciousness between free-floating experiential powers. What implications does this have for an account of personal identity? In particular, how should we evaluate the wide psychological view in light of this finding? The aim of the present section is to argue that it supports a version of the narrow view, according to which our persistence requires continuity in the realisers of our mental states.

Chapter 4: The Material Self

We should begin by acknowledging that there is not going to be any direct argument from the findings of this chapter to the claim that the wide psychological view is false. So far, we have argued that it is a necessary condition upon the existence of human persons that their experiential powers are realised by stable material entities whenever they exist. It does not follow from this that their experiential powers should *always* be realised by the same material entities. Thus, the results identified above are logically compatible with the claim that we can survive Teletransportation or Upload. Nevertheless, I think that the above suggests a good case against the wide psychological view.

The first step to this argument is to recognise that the points made above provide a metaphysically central role to cases in which our experiential powers are realised by stable material systems; even if teletransportation and the like were to become very widespread, these techniques for extending one's psychological life would operate by 'migrating' one's psychology from one such system to another. Given this fact, we can make three points which count against the wide psychological view.

First, if the wide psychological view is correct, then a person's existence can consist in a series of phases in which her mental life is realised by a number of physically continuous systems, linked by processes which 'transfer' her between systems. The proponent of the alternative narrow view will hold that phases of a person's life cannot be connected together in this way; rather, thought experiments like Teletransportation and Upload merely produce someone who psychologically resembles the person subjected to the process. It bears note that this alternative description accurately describes what the processes involved would be designed to do; those designing a teletransporter, or a device for uploading one's consciousness to a computer, would set about their task by trying to find ways to replicate one person's psychology in new material. Given this points about the objectives of the designers of these forms of technology, and the fact that the base case of a person's identity over time involves the realisation of her experiential powers by a unified material system, it strikes me that we should ask what reasons we have to extend an account of personal identity beyond this base case. Put otherwise, these facts suggest that the metaphysical default is to think of personal identity in terms of the conditions which allow for the existence of persons – the physically continuous realisation of their mental states.

§4.5: Fragmentation and Personal Identity

My second point is that the above undermines at least one attempt to argue that we should not differentiate between different ways to generate psychological connections between a person's mental states at different times. This argument might be drawn from the following quotation:

> [T]he force which its being the brain that is transplanted in [a brain transplant] has on our intuitions about that case really does lie wholly in the fact that the brain is normally, slip-ups apart, the carrier of a person's psychology. (Thomson 1997, p. 207)

Of course, Thomson rejects the claim that psychological features are at all relevant to our identity; her objective in the cited paper is to argue that we should be identified with our bodies. Here, however, she claims that if one has to choose between the wide and the narrow psychological views, it is difficult to see what reasons could be given for preferring the narrow view. She appeals here to the philosophical staple of a brain transplant, following the canonical description of such cases by Shoemaker:

> Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the oeprations, however, the assistant inadvertently put Brown's brain in Robinson's head ...Let us call [this man] "Brownson." Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body ...When asked his name he automatically replies "Brown." He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life ...Of Robinson's past life, he evidences no knowledge at all. (Shoemaker 1963, pp. 23–24)

Thomson's thought is that we intuit that Brownson is Brown because the brain transplant procedure ensures that there are psychological connections between these two individuals. Our intuitions about Brown's persistence are guided primarily by the presence or absence of these psychological connections, and not by the fact that his brain is preserved. Given this, the argument continues, it would be arbitrary to adopt the view that it matters how relations of psychological continuity are realised; this would be to make a distinction where there is no difference in the psychological connections in which we are primarily interested.

Chapter 4: The Material Self

Our response to this objection begins by noting that it is not just the norm for our mental states to have physically continuous realisers; this is the metaphysically base case, and is required for there to exist persons at all. Given that we have to refer to the base case in order to explain how persons exist at all, I do not think that it is arbitrary or unjustified to insist that we cannot be transferred between distinct material systems. We have discovered non-obvious necessary conditions for our existence, and I cannot see why an account of our persistence should not be informed by this discovery. When put in this light, I think that the argument presented above no longer seems appealing.

This brings me to my third point. The previous chapter showed that it is difficult to provide a logically coherent account of our persistence whilst confining ourselves to psychological relationships between mental states existing at different times. Armed with the materials of the present chapter, we can say that part of the reason for this difficulty is that such accounts are quick to downplay the way in which our mental states are actually realised. This chapter therefore helps to explain why the previous chapter was correct to draw the conclusion that we should not abstract from the way in which our mental states are realised. I conclude from this that there is strong positive reason to accept that our persistence requires that our experiential powers have physically continuous realisers.

To summarise, complex subjects only exist because there exist stabilised clusters of experiential powers, realised in material systems. Once this is recognised, it enables us to see why the narrow view does not represent a merely parochial preference for the usual above the unusual. Indeed, the previous chapter suggests that acceptance of the wide view has the unfortunate effect of blurring the line between procedures which ensure our survival and those which merely copy us, or create, downwind from us, a number of streams of consciousness. There is therefore a strong presumptive case in favour of the narrow view. I will now consider three ways to argue that we should nevertheless accept the wide psychological view.

4.5.1   Responses

I have argued that it would be nigh on impossible for persons to exist if their experiential powers were not somehow realised. Our minds are not merely unified systems of free-floating experiential powers; their unity crucially depends upon their realisation by systems of material entities. This subsection considers three

ways to deny that this should have any bearing on an account of our nature. I will argue that all three responses can be faulted.

Virtual Reality Revisited

I have already argued that we do not have decisive intuitions in favour of the wide psychological view. Nevertheless, it is worth considering one thought experiment in favour of this view, because of the weight placed upon it by Dainton. This is a form of virtual reality in which a person's stream of consciousness is 'lifted' from her brain, and joined to one which is produced by a machine. In cases like this, it is hard to resist the temptation to believe that one survives. Nevertheless, I think that it is difficult, on reflection, to have a settled verdict on what this case involves.

It is to be imagined that the designers of a virtual reality device have discovered 'the fundamental physical principles responsible for the generation of phenomenal consciousness', and that this allows your stream of consciousness to be 'lifted' from your brain, and joined to one which is produced by a computer:

> [T]he VR-4 program does not install a new psychology into your brain, it completely bypasses your own psychology, and furnishes you with an uninterrupted flow of experience that is entirely computer-driven. Moreover, this stream of consciousness is no longer the product of activity going on within your brain, it too is entirely machine-generated. (Dainton 2008, p. 18)

It is hard to resist the temptation to believe that you would survive using this technology; it seems obvious that wherever there is continuous conscious experience, there is also a single person to which that experience belongs. Dainton suggests that this justifies the conclusion that our continued existence can be accounted for in purely phenomenal terms.

To respond, we should begin by noting that, on Dainton's analysis, to say that one's stream of consciousness is transferred onto a computer is simply to say that that computer comes to realise experiences which are diachronically co-conscious with one's own; a 'stream of consciousness' is not a thing which is transferred from one place to another, but rather a neat shorthand for describing relationships of phenomenal co-consciousness. Further, if it is indeed possible to cause distinct objects to realise diachronically co-conscious experi-

ences, there is no obvious reason to think that relationships of co-consciousness will be well-behaved; for example, it is perfectly conceivable that your stream of consciousness should divide, coming to be diachronically co-conscious with two distinct series of experiences. I find myself quite uncertain about what to say about scenarios in which my stream of consciousness divides or is joined to many others. Given this, I think that it is wise to doubt that Dainton's example of virtual reality provides decisive reason to accept the wide psychological view.

More can be said here. It is easy to think of 'streams of consciousness' as analogous to persisting objects – identifiable particulars which trace out a unified career. If we were able to think of them as such, there would be no reason to object to the claim that a person persists just in case her stream of consciousness continues. Unfortunately, Dainton's account does not allow us to think of streams of consciousness in this way. I would therefore submit that our intuitions about Dainton's cases of virtual reality come from implicitly assuming that the notion of a stream of consciousness is thicker, and more robustly unified, than his theoretical materials allow. Once again, this suggests that we should be wary about drawing conclusions from the thought experiments he puts before us. Let's therefore turn to other, more theoretical, grounds for favouring the wide psychological view.

Parsimony

I have argued that it is not arbitrary, or unmotivated, for an account of personal identity to incorporate restrictions upon the way that psychological connections between a person's mental states can be realised. Nevertheless, it may still argued that it is a prima facie cost for an account to incorporate such restrictions if this can be avoided. The thought here is that the wide psychological view may be favoured by considerations of parsimony; it provides a simpler, less complicated, account of our persistence by abstracting away from the details of the realisation of our mental states.[116]

116. One way in which we can demonstrate the comparative simplicity of this account is that it need not engage with questions such as those raised by Thomson in the following passage:

> Suppose a psychology-transplanting was caused by transplanting a liver. (The drugs injected in the donor to anaesthetize him before the operation caused his brain to imprint on his liver in such a way that, when transplanted, the liver caused reprogramming of the recipient's brain.) Was that good enough? …Or was it not good enough on the ground that transplanting a liver doesn't *normally* caused psychology-transplanting? (Why should that matter?) (Thomson 1997, p. 207)

§4.5: Fragmentation and Personal Identity

To dispense with this argument quickly, we should note that parsimony is a virtue only when competing explanations of some phenomenon are (nearly enough) otherwise matched. This is not the case here. A purely mentalistic account of our existence ignores a necessary precondition upon the existence of perspectives such as our own. I have argued that this leads to difficulties for the wide psychological view; the view needs significant tweaking to provide a logically coherent account of our persistence. This strongly suggests that the wide psychological view is not nearly matched to the narrow view. I now want to consider one final reason to reject the narrow psychological view in favour of the wide.

Multiple Realisability

The third reason for abstracting from material considerations in an account of personal identity lies in the thought that there might be multiple different types of person, with very different underlying material constitutions. Should this be the case, we might be inclined to pursue an account which ignores their differences, and captures the key sense in which they are beings of the same kind (or, at least, beings with similar persistence conditions):

> Since selves may come in many shapes and forms, ideally we want an account of their existence and persistence that is as general as possible ...Human brains may contain one or more EPs [experiential powers], but the same goes for dolphin brains, alien brains, and immaterial souls, if there are or could be such things (Dainton 2008, p. 81)

This is a relatively common motivation for psychological accounts of personhood; we see it also in the following quotation from Shoemaker:

> [I]t does not appear to be any sort of necessary truth that all persons are human beings. Maybe Locke was right in thinking that a highly intelligent parrot could count as a person. Maybe porpoises could turn out to be persons. Maybe there are creatures elsewhere in the universe ('Martians'), biologically unrelated to us, who ought to count as persons (Shoemaker 1984, p. 112)

In both cases, we seem to be faced with an argument of the following form:

1. Different selves are physically constituted differently.

2. In describing a property, we should do so without reference to what differs between its various instantiations, and account for what is common to them all.

C. We should describe selves without reference to their physical constitution, thereby accepting the wide psychological view instead of the narrow one.

I do not find this a compelling reason to excise facts about material realisation from an account of personal identity. Two responses are available. I will begin by casting doubt on the second premise. Subsequently, I will demonstrate that the argument is invalid.

We should, of course, grant that where a property is common to a number of, otherwise quite different, entities, our account of that property ought to abstract somewhat from the specific differences between them. However, it does not follow that we should always overlook these differences entirely; it may be, for example, that these differences account for differences in how the common property is manifested.[117] If this is so, then it would be sensible to acknowledge that the property comes in various kinds, and to map how these kinds differ from one another. Applied to the present case, this is to say that it is at least a logical possibility that we should provide slightly different accounts of the persistence of slightly different types of person (where these different accounts are keyed to differences in the realisers of their psychological characteristics). I note this as a preliminary point; as we shall see, it is not the main flaw in the argument from multiple realisability.

Unfortunately for its proponent, the argument from multiple realisability is invalid. It may well be that various kinds of person exist, and that our accounts of their persistence should abstract from many significant differences between them. However, if all persons are materially realised, then it is common to all of these kinds of person that they have some kind of material constitution.[118] If so,

---

117. For an incredibly stimulating book which uses considerations of this form to mount an opposition to functionalism, and a case in favour of the token identity theory of mental states, see Polger and Shapiro's *The Multiple Realization Book* (2016).

118. Of course, returning to the pure ego theory, one response here might be to insist that our theory of personal identity should recognise the possibility of immaterial souls, and so should not insist that persons are materially realised. The response to this criticism requires some nuance. I would want to argue that Dainton must nevertheless accept that persons are realised by stabilised and insulated collections of experiential powers; this stability is the result of either their material realisation or their inherence in a soul. Thus, he must nevertheless accept that persons are unified partly in virtue of elements which are not experiences and inter-experiential relationships.

then the above argument does nothing to motivate us to ignore the general idea of a person's material constitution in developing an account of personal identity. Instead, it leaves open that we could 'blackbox' the differences between different kinds of person, stating only that persons' experiential powers are realised in a physical system whose continued existence is necessary for their persistence. This would suffice for our argument against the wide psychological view.

Thus, to conclude, I do not think that Dainton can argue that our account of personal identity should skip over the fact that our experiential powers are materially realised. By recognising this fact, we may explain what would otherwise seem puzzling – the possibility of a persisting, robustly unified, perspective upon the world. This is not in tension with recognising commonalities between different types of person.

4.6   Conclusion

Mental states come in causally isolated, and highly integrated complexes. I have argued that we cannot understand how these complexes could exist unless we look to their material basis. Our psychological lives are typically underwritten by the persistence of relatively stable material realisers. Once we recognise this, it is a small step to giving these realisers a more central place in our theory of personal identity – taking their existence to be necessary for our own. Given that previous chapters have failed to find motivation for the wide psychological view, I conclude that this is the correct response. We should reject the wide psychological view, and hold that our continued existence requires the continuous existence of realisers of our mental states.

This chapter has made much of the fact that our mental states are physically realised. I now want to consider what follows from this. Should we accept that we are material? If so, which parts do we have? In the following two chapters, I will argue that there is no reason to reject the (attractive) claim that we are material, and that we have humanoid form; the parts of our bodies are literally parts of us. Chapter 7 concludes by translating the views developed in this thesis into an objection to an existing account of the nature of material entities.

Disjointness

We now turn to the second question with which this thesis is concerned – that of our material composition. Here, it is useful to grapple with empirical findings – in particular with information about how our mental states are realised in the brain. This chapter considers the implications of the finding that some of our mental states depend on different regions of the brain from others. I will investigate whether this finding is inconsistent with two theses about our composition – first, that we have material parts at all, and second that we are the smallest entities to contain the parts involved in the realisation of our mental states. I will argue that neither claim is disconfirmed by empirical findings, though the next chapter will ultimately reject the second thesis.

Section 5.1 introduces what I call 'the disjointness finding' and two views with which it has been thought to be incompatible – 'materialism about the self' and 'thinking subject minimalism'. 5.2 evaluates an argument against materialism about the self presented in Lowe 2006. The argument claims that the disjointness finding shows that no single material entity is necessary for the existence of every one of my mental states. Since I am necessary for the existence of each of my mental states, it follows that I cannot be a material entity. I argue that this argument relies upon substantive theses concerning the individuation of mental states, and of material objects. Lowe does not give us sufficient reason to accept these theses.

I continue, in 5.3, to consider a second argument based on the disjointness finding. This argument turns on the notion of being a subject in a 'strict', non-derivative fashion. We are, needless to say, the subjects of our thoughts in this strict manner. The argument takes the form of a reductio; it is claimed that a series of principles jointly entail that some of our mental states belong, in the strict

sense, to different entities to one another (contra the claim that they all belong to us in the strict sense). I identify the principle which we should reject – the claim that each mental state belongs, in the strict sense, to something composed only by its realisers. I argue that the fact that this principle is flawed fails to discredit the core idea of thinking subject minimalism – that we are the smallest entities to which our mental states can be ascribed.

5.4 brings the discussion of 5.3 to bear upon two objections Olson raises against thinking subject minimalism. The first objection faults thinking subject minimalism for assuming that we can make precise the idea that some mental states are 'directly involved' in our thinking. If this cannot be made precise, it is impossible to say what parts we have. This problem can be bypassed if we accept a more restricted ontology than Olson ascribes to his opponent. The second objection contrasts two versions of thinking subject minimalism and argues that we have no reason to prefer one to the other. Since one of these has unpalatable consequences, we should reject both. I show that the discussion of section 5.3 allows us to distinguish between these two versions of thinking subject minimalism, and to prefer one to the other. Thus, at the conclusion of the chapter, we will have defended the thesis that we have material parts, and developed a specific proposal for determining which parts we have.

5.1   Setting the Stage

The thesis that we have material parts is extremely plausible – especially in light of the finding that our mental states are dependent upon states of the brain. Call this thesis 'materialism about the self'.[119] It may be possible to assess materialism about the self without considering what parts we have.[120] Nevertheless, I want to begin with a concrete proposal to which the majority of materialists

---

119. As defined, materialism about the self is compatible with the claim that we have both material and immaterial parts (being, for example, composed by the conjunction of a specific soul with a specific body). Though it might be argued that it is misleading to call this 'materialism', I can set this criticism aside; it is, as specified earlier, a basic assumption of the thesis that there are no fundamental immaterial substances around to compose us. Further, even if there were, I think that it still would be interesting to defend the thesis that we have material parts against objections.

120. I have in mind here particularly the arguments of Descartes' meditations (see in particular 1641/1998, Meditation VI, p. 78; p. 86). Moving further back, however, we might also consider early Christian arguments against materialism. A good survey is found in R. Martin and Barresi 2006, chs. 3–5.

are committed. The following thesis is shared by all common-sense versions of materialism about the self:[121]

Minimal Parthood If M is a mental state of mine, then every one of M's realisers is part of me.

Minimal Parthood makes use of the technical notion of a mental state's 'realisers'. Given the wide variety of quite different views in the philosophy of mind, and the philosophy of perception, we should pause to specify just which entities we have in mind. I intend Minimal Parthood to pick out a commitment common to almost all views that take us to be material. Thus, we should dictate that something is a realiser of our mental states only if almost any view that holds that we are material will agree that it is among our parts. Let us stipulate that we will understand 'realiser' in an extremely narrow sense, on which a mental state's realisers are restricted to those parts of the brain whose configuration and activity are particularly relevant to explaining its existence. This is a stipulation, and is not intended to beg any questions about the nature of our mental states. I assume that these things will be realisers of our mental states on any possible view of mentality, but I grant that they may be only a subset of the realisers of my mental states on some reasonable views in the philosophy of mind. For example, in the philosophy of perception, there are those who say that the entities that we perceive are amongst the constituents of perceptual experience.[122] These things are quite plainly not among our parts. Thus, although these views about mentality are (I think) reasonable, not everything that they identify as constitutive of our mental states will count as amongst their realisers under the narrow sense of realisation that we have stipulated.[123]

This is only a very loose and imprecise way of pinning down the notion of realisation, and of stating its significance. In 5.4.2, I shall consider whether it is

---

121. Two non common-sense views are Chisholm's claim that we are identical to a single material particle (c.f. Chisholm 1978, p. 31), and Madden's 'remote thought hypothesis' (Madden 2011), according to which there are situations, such as Johnston's 'remnant person' case (Johnston 2007, pp. 45–47; Johnston 2016), in which a person's thinking is done for them by something which is not, at that time, amongst their parts.

122. See, for example, Brewer 2011; M. G. F. Martin 2006; Nudds 2009; Genone 2016.

123. Equally, though, I think that it is fair (at this juncture) to assume that many of the parts of our bodies do not count as realisers of our mental states; I intend to allow for controversy about whether or not these entities are amongst our parts, and so we must deny that they realise our mental states.

objectionable if we cannot give a precise characterization of the notion of realisation (or, equivalently, say exactly which things realise our mental states). For now, however, we can set aside this criticism. I will assume only that neuroscientific findings to the effect that different regions of the brain subserve different cognitive functions can be recast as claims that those regions realise our mental states (whilst other regions of the brain do not). With this in place, we have more than enough content to develop the arguments which this chapter will examine.

Minimal Parthood entails that we contain as parts the totality of the realisers of our mental states. Call these our 'thinking parts'.[124] Note that Minimal Parthood is not intended to provide a full inventory of our parts; it tells us that certain entities are amongst them, but it is silent about whether we have other parts besides.[125] If materialism about the self is to be true, one would expect it to be possible to provide a principled account of which other parts we have (if any). The second view that we shall discuss – thinking subject minimalism – is presented in the spirit of giving such an account.

At a first pass, thinking subject minimalism says that a person has no parts other than her thinking parts. This is how Olson understands the view when he introduces it via the following claim:

> [T]he parts of a thinker must all be in some sense directly involved
> in its thinking. (Olson 2007, p. 88)

We should note in passing that Olson introduces the term 'thinking subject minimalism' to label a potential motivation for a view that he opposes – the view that we are brains, or 'thinking brains' (ibid., pp. 87–91; Olson 2015b, pp. 49–51). Hud Hudson is a proponent of a view of this kind, for this very reason (2001). Other than him, however, it is not entirely clear who accepts the view that Olson characterises. For example, both Jeff McMahan (2002, pp. 88–94) and Derek Parfit (2012) accept the view that we are identical to brain-sized entities that do our thinking. Indeed, Olson takes Parfit to assume thinking subject minimalism

---

124. Of course, 'thinking' here is not understood predicatively; I do not mean to say that these parts themselves think, but rather that they are intimately involved in our thinking (compare the 'rowing' in 'rowing boat').

125. Indeed, it is worth noting in passing that Minimal Parthood is also silent about our persistence conditions. It is compatible with this view that what are now realisers of my mental states could cease to be part of me without my ceasing to exist, and even that they could come to realise some other person's mental states (through, for example, being transplanted into another person's head).

when he discusses his views (c.f. Olson 2015b, p. 49). As we shall see, though, neither Parfit nor McMahan assert that we have no parts other than those involved in our thinking. This will be relevant in what is to come, where I offer a plausible 'moderate' reformulation of thinking subject minimalism, which agrees that we are identical to the smallest entity to contain all of our thinking parts, but does not require that we have no other parts. This moderate formulation avoids some of the reasons Olson gives for rejecting thinking subject minimalism, but is compatible with the motivations Parfit and McMahan give for their views.

Minimal Parthood says that our thinking parts are part of us. Despite the controversy about its interpretation, we can at least say that thinking subject minimalism attributes to us as few other parts to us as possible. This chapter aims to argue that neither is discredited by the following observation:

Disjointness There exist mental states of mine, M and N, such that M's realisers and N's realisers are (at least partly) disjoint; either some of M's realisers are not among N's realisers, or vice-versa.

Put otherwise, Disjointness says that some of my mental states are only realised by a subset of my thinking parts, rather than by all of them. This thesis is plausible, and seems to be supported by the fact that damage to parts of the brain often only disrupts specific types of cognitive functioning, whilst leaving others entirely unaffected.[126] Disjointness also seems to be presupposed by the attempt to differentiate between the functions of different areas of the brain; to say that some part of the brain has the 'function' of carrying out some cognitive task is quite plainly to say that its parts are especially involved in the realisation of that task.[127] I shall assume that Disjointness is true in what follows.

I have identified two claims about our composition, and one empirical observation about the neural realisation of our mental states. This suffices to intro-

---

126. For example, Shammi and Stuss 1999 suggest that damage to the right frontal lobe disrupts the ability to appreciate humour more than damage to other areas of the brain. I choose this paper specifically because of its title – 'Humour appreciation: a role of the right frontal lobe'. This title clearly indicates that the writers take the right frontal lobe to be particularly involved in realising states of amusement.

127. It is, of course, possible that many of the specific cognitive tasks identified by cognitive scientists – tasks such as the detection of the edges of perceived surfaces – do not match up one-one with the types of mental state we characteristically ascribe to persons. Nevertheless, it is plausible that neuroscientific results show that some parts of the brain are involved in the production of some, but not all, of our mental states.

duce the background to this chapter. I shall begin with a very surprising complaint – that Disjointness entails that Minimal Parthood is false, and is therefore inconsistent with common-sense versions of materialism about the self.

5.2 Lowe's Argument against Materialism

This section considers an argument against materialism about the self developed by E. J. Lowe in his 2006. Lowe's case consists of an argument that we have a property – that of being necessary for the existence of each of our mental states – that no material object has. This contention is supported by a subsidiary argument, based upon Disjointness, which justifies the claim that no material object has the property in question. Lowe summarises the former of these two arguments as follows:

> It does not appear that either my whole body, or any particular part of it, can qualify – in the way that *I* do – as the unique subject of all of *my* thoughts and other conscious mental states, for no *one* bodily entity is necessary for the existence of *all* of those mental states of mine, even if each of them depends for its existence upon *some* bodily entity. (ibid., p. 10)

The following seems to capture the core of the argument:

1. I am necessary for the existence of each of my mental states.

2. No material entity is necessary for the existence of each of my mental states.

C. I am not a material entity.

The argument is valid. Its first premise is ordinarily supported by reference to the claim that mental states are individuated by their thinkers; its dependence upon me is part of what makes any mental state the state that it is. This is a plausible view, and Lowe explicitly relies on it in discussion elsewhere.[128] Although there may be some who would reject this claim about the individuation of mental states,[129] I do not know how to evaluate their arguments. I therefore propose

---

128. See, for example, Lowe 2012, p. 149 and Lowe 1996, pp. 25–32.

129. One source of support for this objection is Hume's mistrust of the idea that there could be necessary connections between so-called 'distinct existences' – cited in footnote 101 on p. 92.

to shelve the criticism. Instead, I will point out flaws in Lowe's argument which are more easily validated.

If we accept the first premise of Lowe's argument, that leaves the second to consider. It is fair to say that this premise is not obviously true. Lowe supports it by appeal to a subsidiary argument, based on Disjointness:

> Now, even if it is granted that each of my conscious thoughts and feelings depends on some particular neuronal activity going on in some part of my brain and could not exist in the absence of such activity, it seems clear that there is no *one* part of my brain such that *all* of my conscious thoughts and feelings depend upon neuronal activity going on precisely *there*. Nor, of course, need *all* of my brain exist in order for me to enjoy each and every one of my conscious thoughts and feelings, even if it is true that without a brain I would enjoy no conscious thoughts and feelings whatsoever. For, clearly, I could still have many conscious thoughts and feelings even if my brain were reduced in various different ways by the loss or destruction of various different parts of it ... Hence, I conclude that I cannot be identical either with my brain as a whole or with any specific part of it. (ibid., p. 10)

This argument is intended to rule out two hypotheses – first, that I am identical to some part of my brain, and second that I am identical to my brain itself. I take it that the following captures the spirit of the argument:

1. The existence of a mental state's realisers is sufficient for its existence.

2. As such, nothing which has parts other than the realisers of a mental state is necessary for its existence.

3. By Disjointness, some of my mental states have disjoint realisers from one another.

4. Hence, nothing both contains all of the realisers of my mental states and, for each of my mental states, only has that state's realisers as parts.

Scepticism about the idea that mental states are individuated by reference to their thinkers led to development of so-called 'no-ownership' views, as critically discussed in chapter 3 of Strawson's *Individuals* (1959, pp. 96–98). For discussion of Strawson's argument, see Snowdon 2009b (in particular, §4). Though not a no-ownership theorist, William Carter suggests that one person might acquire another person's mind (and mental states) in his 1989.

C. Nothing both contains all of the realisers of each of my mental states and is necessary for the existence of each of my mental states.

I have three comments to make about this argument. First, Lowe supports its second premise by appeal to the first. However, I do not think that premise 1 does give us sufficient reason to accept premise 2. Second, if 2 is not entailed by 1, it is hard to see why we should accept it (at least without further information about the specific ways in which our mental states are realised by parts of our brains). Third, it might be queried why the materialist should accept premise 1. I will take these points in turn.

5.2.1 An Invalid Transition

It is relatively clear how Lowe means to support premise 2. We are to imagine isolating the realisers of one of our mental states from the remainder of the brain (and stimulating those realisers in just the way they are now being stimulated). According to premise 1, this would suffice for the existence of that mental state, though nothing would then exist which contains both that mental state's realisers and other parts besides. Thus, it is not necessary for the mental state's existence that there exist something with parts other than its realisers.

Unfortunately for Lowe, this does not suffice for the truth of premise 2. This premise asserts that each of our mental states can exist in the absence of every entity which actually has parts other than its realisers. For this to obtain, there must not exist any entity which actually has parts other than the realisers of my mental state, and which could come to be composed by those realisers alone. Put otherwise, Lowe must justify the claim that isolating the realisers of one of my mental states would destroy, rather than reducing in size, anything which possesses parts other than those realisers. The envisaged scenario does not, by itself, motivate this additional claim.

In theory, one could support Lowe's argument by appeal to the thesis that it is impossible for any object presently existing to be composed by a proper subset of the parts that now compose it. In practice, one is only likely to hold this view on the back of a stronger thesis, such as the claim that any object presently existing can only exist if it is composed by exactly the parts that it now has.[130] This stronger thesis is extremely implausible; it stands in the face of the claim

---

130. Chisholm sets out this position in both chapter 3 and appendix B of his *Person and Object* (1976).

that material objects – particularly living things – can persist through material turnover. Equally, though, the weaker thesis ought to be rejected; it has the un-palatable consequence that an object will cease to exist if any parts are cut away from it. Since Lowe's argument is presumably intended to target all materialist views, it should not rest upon such a surprising metaphysical claim.

This point can be pressed. When Lowe introduces his argument, he claims that he could have thoughts and experiences even if his brain was 'reduced' in various ways. At first sight, however, to speak of a brain being reduced is to speak of the brain's continuing to exist, but losing some of its parts. Further, I previously quoted a passage in which Lowe concedes that he could not enjoy thought if he did not have a brain. On the face of it, this concession amounts to the claim that whenever Lowe's mental states exist, they do so in virtue of the activities of a brain that (at that time) belongs to him. This suggests that reducing the brain as Lowe envisages either does not cause it to cease to exist, or that it leaves a different (presumably newly created) brain in place of the original. The latter reading is extremely strange. I doubt that Lowe intends it. Thus, I conclude that the former reading should be preferred, though it undermines Lowe's argument by suggesting that our brains can come to be composed by fewer parts than now compose them.

We can spot where things go wrong by considering how Lowe makes his case. He begins by saying that the existence of his conscious thoughts and feel-ings does not require that of 'all of his brain', by which he means that it does not require all of the parts of his brain. This is extremely plausible. He moves from this to the claim that the 'brain as a whole' is not required for the existence of the experiences in question. This phrase is ambiguous, and can either mean that his brain is not required at all, or that it is not required that his brain contain all the parts that it does at present. Only the former claim supports his argument, but we have not been given good reason to believe this claim. It seems that there is a significant gap in Lowe's argument.

I now want to consider under what circumstances it would be reasonable to conclude that my brain could not survive being 'reduced' to just the parts required for one of my conscious mental states.

5.2.2 Carving up the Brain

As we have seen, the second premise of Lowe's argument claims that no entity which actually has parts other than the realisers of a mental state is necessary

for its existence. Though this does not follow from premise 1, it may still be true. We dismissed the general claim that no entity can lose any of its parts; this is a controversial view, which the materialist may reasonably reject. I cannot think of any other obvious and well-represented views about the nature of material objects which would support premise 2. If we are to make a case for this premise which is acceptable to all materialists, I will therefore presume that the case will have to rest on empirical findings – information about what is actually required for neural material to realise one of my mental states – rather than exotic metaphysical theses (such as the claim that my brain cannot lose any of its parts). This subsection considers what those findings could be.

It seems to me that empirical findings could only support premise 2 if they suggested that our mental states are realised by relatively small, localised portions of the brain.[131] If this were not the case, and large regions of our brains were to realise our mental states, then it would be sensible to conclude that the existence of the realisers of any one of our mental states would suffice for the existence of the brain itself (albeit in a damaged state).

Lowe does not give any evidence that the realisers of our mental states are small and localised. Further, I do not know how this could be proven. It is, of course, true that cognitive scientists are able to identify specific regions of the brain which are particularly involved in tasks such as pattern recognition, one's sense of the passage of time, motion perception, memory recall, and so on. However, I believe that the empirical results in question are typically obtained by observing that subjects who suffer damage to some portion of the brain are unable to carry out the associated tasks, or that that region of the brain is unusually active when the cognitive task is carried out (whilst other regions are no more active than usual). Both types of evidence strongly suggest that the regions in question are necessary for the performance of that task. They do not, however, prove that those regions could carry out the task on their own.[132] Thus,

131. Indeed, even though this would help Lowe's case, it would be insufficient to prove it. Imagine that some particularly central part of the brain (say, for example, the brain stem) is involved in all of our mental activities. If so, one might then hold that the conjunction of this part and enough additional tissue to realise any of our mental states is sufficient for the existence of the brain (in however damaged a state).

132. Kanwisher 2010 develops some evidence, based on these techniques, that many regions of the brain are remarkably specialised. However, as she concedes (p. 11169), much more work is needed to show how the functions of these different regions translate into facts about our mentality. She has in mind here the idea that we need to more precisely characterise the functions of these regions, and spell out what aspects of our experience they are responsible for. However,

they do not prove that the regions would suffice for the existence of the mental state were they isolated and correctly stimulated. There is some work to be done here, and no reason to think that premise 2 will be borne out.

5.2.3   The Individuation of Mental States

Let's now consider premise 1 – the claim that the existence of a mental state's realisers are sufficient for its existence. I have four comments to make about this premise.

First, it is not, in general, true that the existence (and appropriate configuration) of the entities which actually constitute or realise some entity are, in every possible scenario, sufficient for its existence. Suppose for example that I cease to exist, and that countless decades in the future, and perhaps by complete chance, the microscopic particles responsible for my existence are rearranged just as they are now. It is highly contentious that they would then constitute me, rather than a duplicate person with exactly the same material realisers.

We can take this criticism further, based on the way in which I have narrowed down the notion of realisation that interests us. As I stated before, we shall take the realisers of our mental states to be things that are plausibly parts of us – parts of the brain and so on. The existence (and appropriate stimulation) of the realisers of a mental state would only suffice for its existence if one's being in a particular brain state sufficed for being in that mental state. There are plausible views that deny this.[133] It seems to me that the first premise of Lowe's argument requires one to reject these views. Nothing I have said so far is intended to require one to take such a position.

My third comment is that there are possible circumstances in which Lowe may be forced to deny that the existence of a mental state's realisers suffices for its existence. Imagine, for example, that the regions associated with two mental states are entirely disjoint, and so that they can be separated from one

---

in giving examples (e.g. does the area of the brain responsible for face perception also represent the sex, age, and race of the faces perceived?), I think she also highlights the question of how the 'representational states' associated with these areas of the brain map onto the types of mental state that we attribute to ourselves. I raised this concern earlier in footnote 127 (p. 125).

133. Thus, for example, disjunctivists claim that two subjectively indistinguishable experiences, associated with the same pattern of nervous stimulation, can differ in their fundamental kind. See for example, M. G. F. Martin 2002; Snowdon 2005; Snowdon 1990b. See also chapters 2 and 3 of Williamson 2000. I will draw upon Williamson's work in the next chapter.

another and nevertheless continue to support conscious experience.[134] Were this to occur, Lowe would most likely want to claim that at most one of these brain regions was associated with the original subject. However, this turns out to be incompatible with the claim that the existence (and appropriate stimulation) of the realisers of our mental states suffices for their existence. I will briefly explain why and consider a response.

As we have seen, Lowe claims that none of our mental states can exist without belonging to us. Hence, if the original subject is only associated with one of the isolated brain regions, it follows that the other brain region cannot support the same mental states as it did earlier, despite containing all of their realisers. If so, Lowe would be obliged to reject the first premise of his argument. Worse still, one imagines that this type of case would provide the best evidence for the claim that the existence of the realisers of one of my mental states does not always suffice for that of my brain; if the realisers of distinct mental states could be separated in this way, one presumes that only one set of realisers would constitute the brain when separated. Thus, the circumstances which force Lowe to reject the first premise also seem to be just those which are best suited to justify one of the claims that the argument's second premise presupposes.

Of course, Lowe need not claim that the existence of the realisers of one of our mental states is actually sufficient for the mental state's existence; his argument would work if this were a principle that he denied, but materialists about the self were forced to accept. However, it is far from clear that materialism about the self entails this commitment. Indeed, the above thought experiment suggests that there are possible circumstances in which even the materialist may wish to reject this principle. Once again, it seems that Lowe has not provided us with sufficient reason to accept the first premise of his argument.

This brings me to my fourth, concluding point. It is far from clear how we should individuate mental states – what suffices for their existence, and when we should say that they continue to exist. Here, the materialist may well take a leaf from Lowe's book, saying that mental states are individuated by reference to their subjects. If so, then the existence of a mental state's realisers only suffice for its existence when they also compose its subject. If the materialist accepts this,

---

134. This is, of course, an instance of a fission case – similar to the Brain Division case mentioned in chapter 3 (p. 85).

they leave no room for Lowe's argument to succeed, except by directly showing that no material object is necessary for my existence. I do not know how this could be done. I conclude that Lowe's argument from necessity is unconvincing.

5.3   Derivative and Non-derivative Thinkers

Lowe's argument fails because it is not clear what is required for the existence of my brain, or for the existence of the mental states realised within it. If we are to consider the implications of Disjointness for our composition, it would be better to consider an argument which dispensed with appeals to scenarios in which the realisers of my mental states are separated from one another. This section will consider an argument of this kind, based on the following principles:

Exclusion If M is a mental state, then M is at most derivatively attributable to any entity with parts other than its realisers.

Primacy If M is a mental state of mine, then M is not at most derivatively attributable to me.[135]

 The argument takes the form of a reductio:

1. By Disjointness, there is some object which is a realiser of one of my mental states, but is not a realiser of another of my mental states.

2. By Minimal Parthood, if something is a realiser of one of my mental states, then it is a part of me.

C_1. I have a part which is not a realiser of one of my mental states.

3. By Exclusion, if I have a part which is not a realiser of one of my mental states, then that mental state is at most derivatively attributable to me.

C_2 At least one of my mental states is at most derivatively attributable to me.

4. By Primacy, none of my mental states are at most derivatively attributable to me.

135. This commonsense thesis is particularly associated with Chisholm, who writes:

> If there are thus two things that now hope for rain, the one doing it on its own and the other such that its hoping is done for it by the thing that now happens to constitute it, then I am the former thing and not the latter thing. (Chisholm 1976, p. 104)

Chapter 5: Disjointness

As far as I know, no-one presents this argument as a serious objection to the view that we have material parts. However, Olson does present a similar argument against a view committed to Exclusion which he calls 'specific minimalism':

> [G]iven what we know about the division of labor in the brain, specific minimalism would make it unlikely that any being ever engages in more than one type of psychological task. Some see, some hear, some remember, some reason; but nothing does all these things, or even any two of them ... What we take to be one person with general mental activities is in reality a vast colony of numerically different specialists, each of which performs only a single sort of mental task. (Olson 2007, pp. 95–96)

Olson takes this argument to discredit thinking subject minimalism. There are complications to Olson's discussion, and Olson's formulations of thinking subject minimalism and specific minimalism differ very slightly from one another. Therefore, it will be useful to discuss the argument presented here on its own terms before engaging with Olson's concerns in section 5.4.

Since C_2 and premise 4 contradict one another, and the above argument is valid, one must reject one of the principles used in the argument's four premises. To my mind, it seems that Exclusion is the principle in most need of justification; even if we set aside thoughts about the realisation of our mental states, this principle conflicts with the extremely plausible thought that we have parts, such as our fingers and toes, which are not realisers of our mental states. Here, it will be of general interest to unpack two principles that together entail Exclusion, both of which might be resisted.

In my view, the strongest possible justification for the exclusion principle appeals to the following two subsidiary principles:

Existence If M is a mental state of mine, then something exists which has all and only M's realisers as parts.

Narrow Minimalism If M is a mental state of mine, and x and y are entities which share M's realisers as parts, then if x is a proper part of y, y is a subject of M at most derivatively – because it has x, or some part of x, as a part.

Put together, these principles entail Exclusion, and are entailed by the conjunction of Exclusion, Minimal Parthood, and Primacy. I will argue that neither

principle is particularly compelling. However, there is an appealing alternative to Narrow Minimalism. This alternative principle is interesting, because it seems to me to capture a commitment shared by all who accept views in the vicinity of thinking subject minimalism.

5.3.1   Existence

Let's start with Existence. I have no intuitions about the conditions under which aggregates of neural tissue compose a further entity (save that they do so when arranged to form a brain). It is therefore far from clear to me that each of my mental states is associated with an object composed only of its realisers. We should therefore consider how this principle might be justified.

One such justification would be inspection of the brain. Neuroscientists divide the brain into distinct regions. It is plausible that these things really do exist, and are not just representational devices to convey information about brain function. We could imagine it being discovered that each of our mental states was realised by all and only the parts of one such region of the brain. This seems unlikely. I see no reason to think that each of our mental states will be realised by all of the parts of a dedicated region of the brain. Further, I see no reason why each of our mental states should only by realised by entities located within the boundaries of a single brain region; it would not be surprising were many of our mental states to depend upon much more diffusely realised networks of mental states. So, I suspect that empirical findings are not going to be much help in justifying Existence

Alternatively, one might justify Existence by accepting a highly liberal ontology, such as the view that any collection of parts whatsoever compose a further object.[136] On this view, it is guaranteed that the realisers of any of our mental states will compose something, no matter how they are distributed through the brain. It is beyond the scope of this thesis to provide a full analysis of views of this kind. However, it is useful to note that such views are far from obviously true, and there is no reason to think that all materialists will be committed to them. When I discuss Olson's objections to thinking subject minimalism, I will

---

136. Kit Fine, whose views we will discuss, adopts an even more liberal account, on which for any collection of objects, and any relational property they bear, or any possible function from times to objects, there is a corresponding object (Fine 1999). Koslicki 2007, pp. 150–156 argues that this is a cost for his views.

present two points at which it will be to the advantage of his opponent to reject this view.

I conclude that it is feasible to reject Existence. What are the consequences of doing so? First, though denying Existence is logically compatible with accepting Exclusion, this combination of views is unappealing; when conjoined, these theses entail that some of our mental states cannot be non-derivatively attributed to any of the entities which contain all of their realisers.[137] I will therefore assume that if we reject Existence, we should also reject Exclusion. In contrast, denying Existence need not impact upon the truth value of Narrow Minimalism. Nevertheless, I will now argue that we should reject Narrow Minimalism.

5.3.2 Narrow Minimalism

Even if we reject Existence (and, with it, Exclusion), Narrow Minimalism may still have objectionable consequences. Imagine that there exists an entity which contains the realisers of some, but not all, of my mental states, and which is a proper part of everything that contains the realisers of all of my mental states. Given that this entity exists, nothing contains the realisers of all of my mental states without having a part which contains the realisers of some of its mental states. Applying Narrow Minimalism, we find that nothing that contains the realisers of all of my mental states is a non-derivative subject of the states whose realisers are part of our imagined entity. Under these circumstances, we would have to reject either Minimal Parthood or Narrow Minimalism. I suspect that it is the latter which we should reject.

So, Narrow Minimalism might well have negative consequences which would justify us in rejecting it. Is this all that is wrong with it? I think not. The way to see this is to fall back on the claim that it is sufficient for mental states to belong to a single subject that they interact with one another in the right sort of way.[138] We attribute mental states to subjects collectively, rather than piecemeal. As currently formulated, however, Narrow Minimalism con-

---

137. Of course, one could maintain this combination of views, arguing on this basis that materialism about the self – which is incompatible with their conjunction – is false. However, this would be dialectically odd. We have already noted that Exclusion is far from intuitive. We suggested Existence as a potential motivating factor. If this principle is rejected, it is very hard to see what would justify one in nevertheless accepting Exclusion.

138. Olson, in his discussion of thinking subject minimalism, argues that this principle (which he calls the 'psychological individuation principle' is in tension with the view that we have material parts (Olson 2007, pp. 137–139). I respond to Olson's objection in 5.4.3.

siders the relationship between material objects and individual mental states, rather than collections of interacting mental states. We might take this to be a mistake, and state that one should not attempt to attribute mental states to entities in isolation from the other states with which they overact. As such, we may be able to trace Narrow Minimalism's possible objectionable consequences to a misguided method for determining the non-derivative subjects of our mental states.

Though this line of thought suggests that we should reject Narrow Minimalism, it does not suggest that there is anything wrong with the guiding thought that we are identical to the entity with the fewest parts extraneous to the production of our mental states. The following revised principle is true to this motivation:

Wide Minimalism If M is a mental state of mine, and x and y are entities which share M's realisers, then if x is a proper part of y, and x also contains the realisers of every mental state that interacts with M, y is a subject of M at most derivatively – because it has x, or some part of x, as a part.

This principle is compatible with Minimal Parthood and Disjointness; it entails only that we should be identified with one of the smallest entities composed of all the parts involved in generating each of our mental states. If the principle is to be plausible, I think that we would have to accept the further commitment that there is a single 'minimal' object which contains the realisers of our mental states (rather than a series of partly overlapping entities, each of which does so). It strikes me that it is quite plausible that this is true, and that the brain is the minimal part in question. I will assume that Wide Minimalism is workable, and should be preferred to Narrow Minimalism.

To sum up, the argument from derivative thinkers relies upon a controversial view about which material objects exist. It pairs this with what I have argued is a misguided principle for determining which material objects could qualify as the proper subjects of one's mental states. Given this, we should conclude that we can reject both of the principles presupposed by Exclusion. I want to conclude this chapter by considering the relationship between Wide Minimalism and thinking subject minimalism.

5.4 Thinking Subject Minimalism

If we are material objects, then there should be a principled way to determine which objects we are. Thinking subject minimalism is one such way. As Olson characterises the view, thinking subject minimalists are committed to thinking that there is an entity composed by just the realisers of our mental states. This view is stronger than Wide Minimalism, which only claims that there is a smallest entity containing those realisers, but does not specify exactly which parts it has. The first part of this section argues that at least some of those who have views in the vicinity of thinking subject minimalism are not committed to so strong a view as the one characterised by Olson. Subsequently, I will demonstrate that the more moderate view that they accept is resistant to a couple of objections presented by Olson.

5.4.1 Characterising the View

As Olson characterises it, thinking subject minimalism holds that all of the parts of a thinker are directly involved in its thinking. Put otherwise, the claim is that if we are material objects, then we have no parts other than the realisers of our mental states. Olson suggests this principle as motivation for the view that we are our brains, and argues that it ought to be rejected. I want to suggest that we can find a different principle, closer to Wide Minimalism, at work in the writings of at least two of the people whose views he discusses – Jeff McMahan and Derek Parfit – and in those of one who he does not – Ingmar Persson.[139] These thinkers do identify us with the smallest entity to which our mental states can be ascribed, but they do not claim that we have no parts other than the realisers of our mental states. As it is very similar in commitments to the view that Olson criticises, I shall also call the view that I identify 'thinking subject minimalism'.

Thinking subject minimalism, as I understand it, consists in the conjunction of three principles, each of which has been introduced above. These are Primacy, Minimal Parthood, and Wide Minimalism. Together, these principles entail that I am identical to the smallest entity which contains the realisers of all of my mental states. However, if we reject Existence, they do not tell us exactly which material parts we have. In order to discover this, we need to supplement thinking subject minimalism with an account detailing which objects contain

---

139. See, McMahan 2002, pp. 88–94 Parfit 2012, and Persson 1999.

the realisers of our mental states, and what parts they have. In other words, as I characterise the view, thinking subject minimalism only tells us which parts we have when combined with a wider ontology.

I will shortly suggest that, thus characterised, thinking subject minimalism bypasses Olson's objections. First, however, I want to highlight how Derek Parfit motivates a view akin to the one that I have identified. We will evaluate Parfit's argument in subsection 6.2.2 of the next chapter.

In his 'We Are Not Human Beings', Parfit discusses a range of cases which are intended to show that we are not biological organisms. One of these is the 'surviving cerebrum' case:

Surviving Cerebrum [W]hat is removed from my body is not my head, but only my cerebrum, which is then kept functioning by an artificial support system. The resulting entity is conscious, as the neuro-physiological evidence shows. There is also some device which enables this conscious being to communicate with the outside world, since the brain activity involved in certain voluntary mental acts enables this being to spell out the words of messages to us, and some other device enables us to send replies. In this way you have conversations with this conscious being, who claims to be me, seems to have all my memories, and starts to dictate the rest of my unfinished book. (Parfit 2012, pp. 11–12)

According to Parfit, we could survive as artificially sustained cerebra, whilst no organism could be reduced to such a state. This entails that we cannot be organisms. However, if we hold this view, we must face up to an influential objection to views that distinguish us from the organisms located where we are – the too many thinkers problem.[140] Parfit summarises the argument as follows:

> Most human animals . . . have thoughts and other experiences. So if Lockeans distinguish between persons and human animals, their view implies that whenever any person thinks some thought, a human animal also thinks this thought. Every thinking of a thought has two different thinkers. That conclusion seems absurd. (ibid., p. 7)

In the face of this problem, Parfit must either deny that human animals think, or he must claim that it is not problematic to say that there are two thinking

---

140. This problem is also sometimes labelled the thinking animal argument. It forms the staple argument for animalism – the view that we are human animals. For formulations of this argument, see Carter 1989, pp. 10–12; Olson 1995, pp. 165–166; 2003a; Snowdon 1990a, pp. 90–92; 2014b, pp. 92–94.

things where we are. He takes the latter route, beginning with the following claim:

> Animals digest their food by having a part, their stomach, that does the digesting. Animals sneeze by having a part, their nose, that does the sneezing. These facts do not create a Too Many Digesters or Too Many Sneezer's Problem. Human animals think, we can similarly claim, by having a part that does the thinking. (Parfit 2012, pp. 14–15)

I take it that Parfit's claim here is that it is not objectionable if there are, in a sense, two thinking things located where we are, so long as one of those entities thinks derivatively – in virtue of the other thinking on its behalf. Further, there is one particularly obvious way in which a human animal has properties derivatively – when the properties are possessed non-derivatively by some of its parts. Indeed, given the specialisation of a human animal's organs, examples of this kind are ubiquitous. Thus, we can avoid the too many thinkers problem in a principled way by accepting that we are parts of human organisms, who derive their psychological properties from us. If Parfit's argument succeeds, Wide Minimalism emerges readily from attempts to distinguish us from the organisms located where we are.

We will have to wait until the next chapter before evaluating this argument. For now, however, note that Parfit nowhere appeals to the claim that there exists anything composed of all and only the realisers of our mental states (or to anything other than our commonsense ontology). Thus, although I think that Parfit does accept Wide Minimalism, I do not think that he would accept Olson's formulation of thinking subject minimalism.

Parfit is not the only thinker to appeal to a distinction between derivative and non-derivative thought, or to use it to suggest that we are much smaller than the organisms to which our brains belong. Thus, for example, McMahan and Persson both float similar principles:

> My organism is conscious only in a derivative sense, only by virtue of having a conscious part. Similarly, when I blow the horn of my car, the car makes a noise only in the sense that one of its parts makes a noise. There is only one noise; and there is a clear sense in which there is only one noisemaker: the horn ... Nor is the organism as a whole involved in the experience of consciousness except by containing that which is conscious. (McMahan 2002, p. 93)

> A human animal does, indeed, think and have experiences, person-
> alists should concede, but, they can add, it does so in virtue of there
> being something within it which does the thinking and experiencing.
> *Primarily*, it is this proper part or constituent which does the think-
> ing and experiencing, and the organism only does so *derivatively* by
> having it as a part. (Persson 1999, p. 521)[141]

In both cases, I think it is fair to attribute Wide Minimalism to these thinkers, but I do not think that they are committed to the claim that there exists an object composed of all and only the realisers of our mental states. I now want to consider two objections that Olson raises to thinking subject minimalism, as he defines it, and to suggest that the more modest view that I have presented avoids those objections.

5.4.2 Olson's first objection

Olson's first objection is to the idea that we have any way of saying just which objects are directly involved in the production of a mental state. This suggests that thinking subject minimalists lack the resources to give a principled answer to the composition question, since they cannot say exactly which parts we have.
Olson presents this objection as follows:

> Saying that all the parts of a genuine thinker must be directly in-
> volved in its thinking is like saying that all the parts of a genuine
> walker must be directly involved in its walking. Because there is no
> saying, even roughly, which things are directly involved in thinking,
> there is no saying even roughly which things, according to minimal-
> ism, are the parts of a thinker. (Olson 2007, p. 93)

We should consider this quotation alongside a concession that Olson makes:

> I don't mean to say that there is no truth at all in the claim that the
> brain, and not the liver or the stomach, is the organ of thought. Many
> parts of the brain are probably more directly involved in thinking
> than any other parts of the organism, just as those who beat the metal
> are more directly involved in making knives than those who sweep
> the factory floor. (ibid., p. 93)

---

141. Persson argues that this view does not neatly account for our attribution of physical proper-
ties to ourselves. He floats the following pessimistic conclusion:

> [I]t may be that what the separability of persons – or, at least, primary mind owners –
> and their organisms here argued for ultimately leads to is not personalism, but the
> *sceptic* or *nihilisitic* view that our conception of ourselves, and the consequent condi-
> tions of our persistence, are incoherent. (Persson 1999, p. 531)

Chapter 5: Disjointness

Given this concession, it seems that Olson would agree that direct involvement is a matter of degree; some objects are more directly involved in our thinking, and some are less directly involved. His objection relies on the claim that there is no hard and fast distinction between those parts which are directly involved in our thinking, and those which are not. According to him, there is no non-arbitrary way to decide which things are directly involved in the production of our mental states.

In response, we should note first that though there may be no principled way to draw an exact line, we may nevertheless be able to distinguish between parts that are definitely directly involved in our thinking (e.g. particular nerve cells), and parts which are definitely not directly involved in our thinking (e.g. muscle fibres in our legs); nothing in the mere fact of continuous variation prevents this. Given this, a difference between our presentation of thinking subject minimalism and Olson's becomes relevant. I suggested above that the more moderate form of thinking subject minimalism presupposes a background ontology. Wide Minimalism is then employed to determine which object within this ontology is the non-derivative subject of our mental states.

The key to the response I have in mind is the idea that there may be significant compositional differences between any two objects countenanced by a restricted ontology. If so, it is theoretically possible that no object in our ontology both contains all of the objects that are definitely directly involved in our thinking and lacks some parts which might be directly involved in our thinking; on such an ontology, anything that contains everything definitely directly involved in our thinking will also contain everything which could reasonably be judged to be directly involved in our thinking. If so, then the margins of imprecision in the notion of direct involvement will be irrelevant to the determination of the smallest entity to contain our thinking parts; however this imprecision is resolved, the same entity will be picked out as the minimal subject of our thoughts.

Thus, a thinker like Parfit, who identifies us with our brains, can respond to Olson by claiming that every way of clarifying the notion of 'direct involvement' highlights the brain as the smallest unit composed of everything that is directly involved in a person's thinking. Of course, to defend this response, one would need to give us reason to believe this claim. Though I do not want to consider this point in detail, at least one reason suggests itself. It is often supposed that a lone brain could, if supported as in Surviving Cerebrum, sus-

tain a conscious perspective indistinguishable from the one that you now possess; the brain doesn't lack anything that is necessary for conscious experience. This shows that the parts of the brain have an especially intimate relationship to thought which the other parts of one's body do not.[142] It is not obviously mad to mark this distinction by saying that the parts of the brain are directly involved in one's thought, whilst the other parts of one's body are not.[143]

As long as our ontology is not populated with a large number of entities which only slightly differ in their composition, I assume that this suffices to defend thinking subject minimalism, as we have characterised it, against Olson's first objection. Of course, it is necessary that we should be able to, within broad limits, tell when some parts are directly involved in generating a mental state, and when they are only peripherally involved. Though Olson would, I presume, claim that the imprecision of the notion of 'direct involvement' prevents even this,[144] I think that he would need to say more to make his case. I therefore conclude that Olson's first objection misses its mark.

5.4.3   Olson's second objection

Olson's second objection is that there is no reason to prefer a 'friendly' formulation of thinking subject minimalism, such as Wide Minimalism to Narrow Minimalism, on which many of our mental states are non-derivatively attributable to different material entities from one another. We have already seen that we should reject Narrow Minimalism. If Olson is correct that Wide Minimalism is no better motivated, this suggests that we should reject it as well.

---

142. I consider a similar argument in subsection 6.2.1 of the next chapter.

143. There is a second possible response, but it is a little more complicated. Imagine that only three entities contain all the realisers of our mental states – our brains, our heads, and the humanoid entities located where we are. It might be argued that any precisification of the notion of 'direct involvement' which dictates that the brain lacks parts directly involved in one's thinking will rule that the head also lacks some such parts (i.e. any type of part which plays a significant role in thinking, and which the brain lacks, will have instances outside of the head). If so, the thinking subject minimalist can either say that we are brains or that we are humanoid. If we are humanoid, we have very many parts which are not directly involved in our thinking. It might be argued that, since we introduced the idea of 'direct involvement' to pick out the minimal entities responsible for our mental states, the latter precisification is too liberal about the notion of direct involvement. This leaves the view that we are brains.

144. Indeed, he does so when discussing Parfit:

> The problem is that there is no principled way of drawing even a vague boundary around the walking part of me. And thinking is unlikely to differ in this respect from walking. (Olson 2015b, p. 50)

Chapter 5: Disjointness

Olson presents the objection as follows:

> Call an activity such that whatever engages in it in the strictest sense must be composed of just those things directly involved in its engaging in that activity a *minimalistic* activity. Thinking-subject minimalism says that thinking or mentality in general is minimalistic. Specific minimalism says that specific types of thinking, such as remembering, are minimalistic. (Given how the brain works, thinking in general and remembering will not both be minimalistic.) Which is it? (Olson 2007, p. 97)

Let's begin by clarifying the labels Olson uses. 'Specific minimalism' names the conjunction of Narrow Minimalism and the claim that, for each mental state, there exists an entity composed of just the realisers of that mental state (i.e. the claim we earlier labelled Existence). 'Thinking subject minimalism', as we have seen, consists of Wide Minimalism conjoined with the claim that there exists something composed of all and only my thinking parts. I have already argued that we need not accept these claims about which entities exist; unless we adopt a fairly plenitudinous ontology, we have no reason to believe that there will be entities composed of just the parts involved in the production of one or more mental states. Rejecting this assumption may have implications for Olson's objection.

If we deny that there are entities composed of just the realisers of some mental state (or collection of mental states), it is logically possible that (the division of labour in the brain notwithstanding) the smallest entity to contain all of the realisers of any one of my mental states also contains all the realisers of my other mental states. If so, then (as Olson might put it) the smallest entity to engage in any specific form of thinking would also be the smallest entity to engage in thinking in general. If so, we would not need to choose between the two forms of minimalism on offer; they would have just the same implications about our composition. This would undercut Olson's argument. Nevertheless, this response strikes me as somewhat speculative. I shall therefore argue that we have good reason to prefer Wide Minimalism to Narrow Minimalism.

I have two suggestions about how to defend this preference. First, we know that we are the non-derivative subjects of all of our mental states. We also find Minimal Parthood very plausible; it is hard to deny that the realisers of our mental states are among our parts. Given that we have seen that Narrow Minimalism may be inconsistent with the conjunction of Minimal Parthood and Primacy, it

seems that we have good reason to reject it; there exists an argument, based on claims that we take ourselves to know, to the conclusion that Narrow Minimalism is false. No such argument exists to the effect that Wide Minimalism is incompatible with what we know. Thus, there is at least some reason to prefer Wide Minimalism to Narrow Minimalism.[145]

Second, the consequences of Narrow Minimalism are not the only reason to disprefer it to Wide Minimalism. I argued above that we might distinguish the two by appeal to the principle that interacting mental states belong to the same subject. If we accept this principle, we should deny that one can determine the non-derivative subjects of our mental states piecemeal, as does Narrow Minimalism. Again, this gives us a principled reason to prefer Wide Minimalism. It therefore seems that Olson is wrong to think that there is no choosing between the two forms of minimalism.

I have given two reasons for thinking that we can prefer Wide Minimalism to Narrow Minimalism. Olson considers and rejects the second of these, but I do not think that his objection succeeds. He claims that the principle that interacting mental states belong to the same subject – which he calls 'the psychological individuation principle' – is incompatible with materialism about the self. If so, it is incompatible with thinking subject minimalism, and cannot be relied upon to defend it. Olson's argument has two steps. First, he argues that the psychological individuation principle entails that we have mental states as parts:

> [The psychological individuation principle states] that, necessarily, for every person or thinking being ("mind"), there is exactly one mental system, all and only the elements of which are the mental states of that being; and for every mental system, there is exactly one thinking being whose mental states are the elements of that system … How could the number of mental systems necessarily fix the number of anything but mental systems? Remember: the number of mental systems is determined entirely by causal relations among mental states and actions. And it is hard to see how causal relations among mental states and actions could entail both the existence and

---

145. I do think that Olson has a response here, but it is very difficult to see where the burden of proof lies. The response would be to claim that we are trying to assess whether thinking is minimalistic and, in doing so, comparing it with alternative views, such as animalism. Here, it might be thought relevant that we do not know how one could differentiate these forms of minimalism save by their consequences; there is, one might say, no principled reason to accept a workable form of minimalism when we could instead reject it altogether.

the precise number of things that are not made up even partly of mental states or actions. (Olson 2007, pp. 137–138)

The second stage of the argument is to claim that no material object has mental states as parts:

> It is especially hard to see how the psychological individuation principle could be compatible with our being material things. (I assume that no material thing has mental states as parts.) Any material thing that could have mental properties at all, it seems, could be mentally *dis*unified ... In that case, two or more unified mental systems would be associated with one human organism at once. According to the psychological individuation principle, there would therefore be two or more thinking beings – two or more *people* – "sharing" that animal. (ibid., p. 138)

He then continues to argue that if two or more people 'share' an animal, we should deny that they are material beings (ibid).[146] Hence, Olson concludes, the psychological individuation principle is incompatible with our being material things (and therefore incompatible with minimalism).

I think that this argument can be opposed at a number of points. First, it is not clear to me why the claim that persons are individuated by reference to mental states entails that they have mental states among their parts.[147] Second, I appealed to the claim that interacting mental states belong to a single thinker (provided that they are unified in the right sort of way), but I did not claim that mental states only belong to a single thinker if they are interacting; it is compatible with what has been said that we might discover that a single material thing is disunified in virtue of containing the realisers of disjoint sets of interacting mental states, and so that it was mentally disunified.[148] Thus, it is not

146. For a similar argument, see Olson 2003b.

147. We might compare the psychological individuation principle with Olson's preferred alternative – the claim that the number of persons is determined by the number of biological lives – integrated series of biological processes which together constitute a homeodynamic system. This principle is compatible with holding that we do not have processes and biological states as *parts*. It is simply unclear why the claim that one set of things play a role in individuating an entity entails that the former set of things are among its parts.

148. Another point here: Olson claims that we cannot understand how disunified mental states could belong to different material entities except by taking each to be identical to a different part of the brain. He continues "But that presupposes thinking-subject minimalism, and as we saw in chapter 4, minimalism faces no end of trouble" (Olson 2007, pp. 138–139). However, as we have seen, Olson's primary objection to thinking subject minimalism is that it requires, but is

entirely clear why the psychological individuation principle is supposed to be incompatible with our being material.

We can conclude that it is far from clear why Olson thinks that we have no principled reason to prefer Narrow Minimalism to Wide Minimalism. Given this, I conclude that his arguments against thinking subject minimalism are unsuccessful.

5.5   Conclusion

I have argued that Lowe's argument from necessity does not generate strong objections to materialism about the self. I have also argued that thinking subject minimalists need not identify different primary subjects for many of our thoughts. On its best, most moderate formulation, thinking subject minimalism only tells us what we are when supplemented with an ontology, telling us which objects exist and how they are composed.

The aim of the next chapter is to examine thinking subject minimalism, and to argue that we should reject the view. Rather than attempting to spell out an ontology in depth, I shall make some assumptions. First, we should not deny that brains exist, absent very compelling arguments to the effect that they do not. Since no part of the brain both clearly exists and is plausibly the subject of every one of my thoughts, I therefore also assume that if thinking subject minimalism is true, our brains are the non-derivative subjects of our thoughts (rather than some smaller entity which they contain). With these assumptions stated, it is possible to assess thinking subject minimalism by comparing it to the alternative hypothesis that we have humanoid form. I argue that the latter thesis can be defended.

---

incompatible with, the psychological individuation principle because the latter entails that we are not material. Here, however, he's supporting the claim that the psychological individuation principle entails that we are not material on the basis that thinking subject minimalism is false. It seems, therefore, that his grounds for thinking that the psychological individuation principle is incompatible with thinking subject minimalism presuppose that thinking subject minimalism is false. The argument against thinking subject minimalism appears to presuppose what it is intended to show.

The Problem of Overlappers

The previous chapter introduced our thinking parts – those entities which, in a fairly minimal sense, realise our mental states. According to commonsense forms of materialism about the self, our thinking parts are literally part of us. I have argued that empirical findings present no obvious reason to reject this claim. Nevertheless, if materialism about the self is true, we should be able to determine which other parts we have. Here, we encounter a problem, which I label the 'problem of overlappers'.[149] More than one entity contains my thinking parts. An account of our nature should present us with reasons to think that only one of these (non-derivatively) thinks my thoughts. This chapter offers some considerations which cut in favour of the commonsensical claim that we have 'humanoid form'. I pursue this claim in close conversation with a view developed in the previous chapter – thinking subject minimalism – which identifies us with the smallest entities to contain all of our thinking parts.

I begin section 6.1 by setting out the problem. I finesse this problem by considering two ways in which entities which contain my thinking parts relate to my thoughts. According to the terminology that I shall adopt, brain-sized entities 'generate' those thoughts, whilst they are 'housed within' larger humanoid entities. The question I investigate is whether we should be identified with the generators of our thoughts, or the entities within which our thoughts are housed. 6.2 considers three ways to motivate the view that we are brain-sized. The first way attempts to argue that that if we accept that a lone brain could sustain a conscious perspective, we should also conclude that we are brain-sized.

---

149. The problem which interests me has been given a number of labels other than the one I use. Thus, for example, Olson references the 'thinking-brain problem' (2007, p. 76), Blatti and Moran the 'thinking parts problem' (2016; 2018), and Madden the 'problem of thinking parts' (2016).

Chapter 6: The Problem of Overlappers

The second argument is based on the two claims that we are not organisms and that nevertheless we speak correctly when we attribute psychological properties to the organisms located where we are. The third argument, originally presented by Carl Gillett, suggests that many of our mental states cannot be attributed to anything with sensory organs or muscles (and hence that they must be attributed to our brains). None of these arguments are entirely successful (though I acknowledge that one can only fully respond to the first argument by motivating an alternative metaphysical picture).

With the arguments for thinking subject minimalism set aside, I consider two reasons to favour the alternative claim that we are humanoid. 6.3 considers an argument by Bennett and Hacker to the effect that nothing brain-sized can think. This argument rests on a very specific, post-Wittgenstinian, picture of the nature of mental states. This picture is controversial, and I suspect it should be rejected. As a result, it is far from clear that their argument succeeds. In 6.4, I offer my own proposal. I argue that there is a particularly intimate connection between many of our mental states, and parts of the body other than the brain. It seems to me to follow that these states primarily belong to entities much larger than the brain. I conclude that we should reject thinking minimalism, and claim that we have parts other than the realisers of our mental states. This paves the way for the discussion of the next, and final, chapter.

6.1   The Problem of Overlappers

I begin by sketching a challenge for any account which identifies us with some material object. The challenge begins with the philosophical commonplace that a lone brain could constitute a thinker; thus, for example, the lone cerebrum in Surviving Cerebrum (presented on p. 139 of the previous chapter) sustains a conscious subject. Faced with this commonplace, we might draw the conclusion that all that is required for something to think is that it contain thinking parts. Unfortunately, this conclusion has unpalatable consequences. Our challenge is to show how they might be avoided.

To illustrate the problem, it is helpful to introduce the notion of an 'overlapper'. Two entities overlap if they have parts in common. For our purposes, however, a more restricted notion will be of use. Let us say that a person's over-

lappers are just those entities which contain all of her thinking parts.[150] Thus construed, our overlappers include, at least, our brains and heads. They may also include more exotic entities besides – such things as all of a person except her left foot, or the sum of that person's thinking parts and the Eiffel tower. It is intuitive to think that at least some overlappers exist, even if there are some whose existence we might deny.

As mentioned above, it is plausible, when looking at cases like Surviving Cerebrum, to think that we need no parts other than our thinking parts in order to enjoy (at least some) mental states. This generates the following argument:

1. More than one entity contains one's thinking parts.

2. Anything which contains one's thinking parts is a thinker.

C. Where one is, there are a number of thinkers.

The conclusion of this argument is extremely surprising. It seems straightforwardly absurd to say that there are two thinking entities where I am. Call this the 'problem of overlappers'. It is sometimes emphasized that accepting that several of our overlappers think leads to further unpalatable consequences. Thus, for example, Olson writes:

> If you think you're an animal, then your head, which thinks just as you do, ought to think, mistakenly but on the same grounds, that it is an animal. So for all you know, you might be your head. (Olson 2007, p. 216)

Call this the 'sceptical' presentation of the problem. This presentation assumes three things. First, when you think first-personally, each of your overlappers also has a first-personal thought about itself. Second, when you ascribe a property to yourself, your overlappers each ascribe the very same property to themselves, often mistakenly. Third, if one of your overlappers mistakenly ascribes a property to itself, then you cannot knowledgeably ascribe that property to yourself. These are substantive assumptions, and I do not know how to assess them.[151] Nevertheless, I think that these further consequences are not required

---

150. As presently defined, we are our own overlappers. This allows for a simple presentation of the topic of this chapter – the question of which of my overlappers is identical to me.

151. For work that suggests that one or more of these assumptions might be false, see Blatti 2016, pp. 167–169; Brueckner and Buford 2009; Madden 2011; Madden 2016b; Noonan 1998, pp. 312–318. For a response to Brueckner and Buford, see Yang 2013.

to see what is puzzling about the problem of overlappers. It is prima facie absurd to think that there is a large number of thinking things where I am. This chapter considers how best to reject this conclusion.

6.1.1 Evaluating the Argument

The argument has two premises – first, that there are a number of overlappers located where I am, and, second, that containing my thinking parts suffices for something to be a thinker. Since the conclusion seems absurd, we know that at least one of these premises is false. Let's first consider which of these premises we should reject.

Though there may be principled reasons to deny the existence of overlappers distinct from oneself – and, indeed, though this is a popular response to the problem[152] – I will assume that we should deny this premise only as a final resort. It would be extremely surprising to learn that such entities as brains and heads do not exist. Further, those who deny that our overlappers exist frequently also deny the existence of many other things around us – entities such as tables, chairs, and so on. If we can respond to the thinking parts problem without accepting such an outlandish solution, we ought to do so.

This leaves us with the second premise of the argument. This premise is false if there is just one example of an overlapper which does not think. However, more is needed to avoid the unpalatable conclusion that there are many thinkers associated with our thinking parts; we require good reason to believe that only one of my overlappers thinks, not just that one does not. To fully satisfy us, a response to the problem of overlappers must therefore give us some reason to believe that we are the only entities to think our thoughts – preferably by explaining why only things of certain kinds, or with certain parts, can stand to our thoughts as we do.

Here is how I propose to meet this challenge. It is likely that we have a number of overlappers. However, in line with the assumptions presented in the conclusion of the previous chapter, I presume that only two are 'eligible' candidates to be us; we will be either a small, brain-sized entity, or a large one, containing all of what we intuitively think of as the parts of our bodies. My

---

152. See, for example, Merricks 2001; Olson 2007, pp. 215–219. In developing this response, both are significantly influenced by the organicism of van Inwagen 1990.Watson 2016 argues that this response cannot be motivated. Clint Dowland 2016 trials an response of this form, which claims that we are brain-sized, and that humanoid organisms do not exist.

reason for drawing this conclusion is that when I think, entities of both sizes stand in particularly close relationships to my thoughts – relationships in which no other overlappers stand to those thoughts. I will briefly explain the nature of these relationships.

6.1.2 Two Relationships

I contend that both our brains (or brain-sized entities) and larger humanoid entities stand in important (and subtly different) relationships to our thoughts. This subsection will briefly elaborate upon the nature of these relationships before spelling out the consequences of this for the problem of overlappers.

It will be helpful to begin by assigning labels to the relationships in which we are interested. I shall say that our brains (or brain-sized entities) 'generate' our mental states, but that those states are typically 'housed within' a larger entity. I shall explain each label in turn.

To say that my brain generates my mental states is, in part, to say that it contains the parts whose activity is responsible for the existence of my conscious perspective.[153] There is more to say. First, it is plausible to think that my brain's existence is necessary for the existence of each of my mental states (even if other objects or conditions are necessary as well). McMahan suggests an even closer tie between our mental states and the brain when he writes:

> [A]ll of the plausible understandings of the relation between mind and brain agree that the mind is either causally generated by or identical with certain parts of the brain when they are in certain states, and hence that the mind cannot be tracked or traced independently of the brain. The continued existence of the same mind thus requires the continued existence and functioning of certain regions of the same brain irrespective of whether mental states are brain states, functional states, or causal products of brain states. (McMahan 2002, p. 88)

---

153. This is, I think, the dominant reason for accepting the view that we are brains. Thus, for example, Nagel writes:

> On the evidence, the intact brain seems to be responsible for the maintenance of memory and other psychological continuities and for the unity of consciousness. If in addition the mental states are themselves states of the brain, which is therefore not just a physical system, then the brain is a serious candidate for being the self – even though, as I shall admit, it does not meet all the intuitive conditions on the idea of the self. (T. Nagel 1986, p. 41)

Chapter 6: The Problem of Overlappers

McMahan claims not only that the existence of my brain is necessary for that of the mental states that I actually have, but also that my brain's existence is required for any mental states to exist and belong to my mind (whether or not these are identical to the states that I actually have). Put otherwise, it is not just individual mental states, or thoughts, but also my mind as a whole whose existence depends upon that of the brain. Given our previous arguments against the wide psychological view, I think that this is an extremely plausible view. We should accept that our minds are individuated by reference to our brains.[154]

Whilst it is often emphasized that there is an intimate connection between our brains and our mental states, it is less often recognised that there is also an intimate connection between our mental states and entities much larger than brains. Yet, the phenomenology of conscious experience seems to bear immediate witness to this fact. We do not, to use Descartes' turn of phrase, experience ourselves as present in our bodies "in the way a sailor is present in a ship" (1641/1998, Meditation VI, p. 81). Rather, it seems immediately evident that the parts of my body are literally parts of me.[155] My experiences seem to reflect the perspective of something humanoid. Indeed, there are several further ways in which my mental states are particularly oriented towards something much larger than the brain. Michael Burke lists a number of these in his 2003. The third and fourth elements of his list are particularly compelling:

> 3. There is a thinker who is immediately conscious of all and only those tactile and kinaesthetic sensations that are felt in some part of Percy [a humanoid entity], but no thinker who is immediately conscious of all and only those tactile and kinaesthetic sensations felt in some part of Finn [an entity containing all of Percy's parts other than his left finger]. Any thinker who is im-

---

154. There is one case to consider here, which I should like to set aside. This is a case in which the neurons of my brain are replaced, one by one, with inorganic replacements, designed to pass on electrical signals when stimulated. This process culminates in something with no organic parts. One might hold that it is not identical to my brain. Yet, if the process of replacement occurs gradually, and I am conscious after each transition, one might think that my mind thereby survives, though my brain does not. I set this aside because, first, I am not sure what would happen to my mind in a case like this, and, second, I am not sure that the brain (or something brain-sized) could not come to be composed of inorganic parts. Hence, I do not know what to make of the case.

155. Indeed, Descartes seems to agree with this when he continues:

> I am most tightly joined and, so to speak, commingled with it, so much so that I and the body constitute one single thing. (1641/1998, Meditation VI, p. 81)

mediately conscious of sensations felt in Finn is immediately conscious, in a natural, normal, and ordinary way, of sensations felt in something, a finger, that lies *outside* Finn.

4. There is a thinker who has direct voluntary control over all parts of Percy over which any thinker has direct voluntary control, and over nothing that *isn't* part of Percy. But any thinker who has direct voluntary control over parts of Finn has direct voluntary control, in a natural, normal, and ordinary way, over something, a finger, that is *not* part of Finn. (ibid., p. 118)

These observations bear witness to the fact that our mental states stand in sensory and volitional relationships to entities much larger than the brain.[156] When I say that our mental states are 'housed within' humanoid entities, I mean to draw attention to the many respects in which our mental states are particularly attuned towards those entities.

Returning to the problem of overlappers, we should note that it is uncontentious that our thoughts are generated by entities other than those that house them. If we reject the seemingly absurd claim that I overlap with a number of thinkers, we must say that either the entities that house my thoughts do not strictly think, or that the entities that generate them do not.[157] A satisfactory an-

---

156. In passing, we should note that Strawson, in questioning why we attribute our thoughts to material entities, presents a case which separates some of these relationships from one another:

> There is a subject of visual experience, S, and there are three different relevant bodies: A, B and C. (1) Whether the eyelids of B and C are open or not is causally irrelevant to whether S sees; but S sees only if the eyelids of A are open . . . (2) Where A and B may be, however, is quite irrelevant to where S sees from . . . This is determined only by where C is. . . . (3) [W]hich of all the views [available to C] is the view seen by S depends on the direction in which the head and eyeballs of B are turned, wherever B may find himself. (P. F. Strawson 1959, pp. 90–91)

Though this is not the conclusion that Strawson would draw, one may draw the conclusion that this case shows that we should not attribute thoughts to the entity (or entities) within which they are housed, but only to those that generate them. I do not think that one is forced to this conclusion; it would be possible, for example, to say that it is indeterminate what parts the subject has, or to deny that the case presents an example in which a subject's mental states are housed in any body at all.

157. At this point, we should briefly consider the following pessimistic conclusion:

> Brains, the idea goes, have a property that is a good candidate for being called 'thinking', and animals have another property that is an equally good candidate, and that's all there is to it. (Olson 2007, p. 83)

If this conclusion is correct, there is no way to argue that one type of entity thinks and the other does not. Olson does not think that we are forced to this conclusion. I agree. We can find reason to

swer to the problem of overlappers must adjudicate between these possibilities. I now want to consider a number of responses to the problem of overlappers, beginning by considering arguments in favour of thinking subject minimalism.

6.2 Motivating Thinking Subject Minimalism

If thinking subject minimalism is correct, then we are identical to the smallest entity to contain our thinking parts. The view is committed to the claim that there is a unique smallest entity. However, the thinking subject minimalist need not think that this entity has no parts other than my thinking parts. In line with the assumptions floated at the end of the last chapter, I shall assume that thinking subject minimalism identifies us with entities composed of all and only the parts of our brains.

Parfit summarises this view as follows:

> [W]e are not animals, or human beings. We are what McMahan calls the conscious, thinking, and controlling *parts* of human beings. (Parfit 2012, p. 14)

As can be seen from this quotation, thinking subject minimalism identifies us with the generators of our mental states, rather than the entities within which they are housed. This section will consider three arguments to the conclusion that this is the correct response to the problem of overlappers. I shall suggest that all might be challenged.

6.2.1 Lone Brains

Let's begin by considering what arguments can be gleaned from the case of a lone brain, sustaining thought by itself. If the example is metaphysically possible,[158] this shows that one's thinking parts alone are capable of sustaining a conscious perspective. I will consider two ways to amplify this claim into arguments for thinking subject minimalism.

---

differentiate between these two relationships, saying that one is thought, and the other is merely the exercise of a related capacity (in my view, the capacity to 'generate' the thought of another entity).

158. This can, of course, be doubted. See, for example, Geddes 2016, §4.2, and Schechtman 1997.

§6.2: Motivating Thinking Subject Minimalism

The first, less sophisticated, argument runs as follows:

1. No more than one's brain is required to sustain a conscious perspective.

2. Thus, no entity with parts other than those of one's brain is necessary for the existence of one's conscious perspective.

3. I am necessary for the existence of my conscious perspective.

C. I do not have parts other than those of my brain.

This argument is very similar to one that we discussed in the previous chapter – an argument against materialism about the self presented by E. J. Lowe. I argued that Lowe's argument falters because he overlooks the possibility that some entity with parts other than the realisers of one of my mental states could come to be composed by those realisers alone. This argument is similarly flawed; we should not assume its second premise without argument to the effect that nothing larger than the brain could come to be composed only by its parts. Johnston puts this point nicely when he writes:

> The plausible observation behind the [view that we are brain-sized] is that the survival of one's brain can be sufficient for one's survival and may well be necessary for it. But this does not show that we are of the kind *human brain*, i.e., that in our normal unmutilated condition we are pinkish-grey, spongy organs awash in cerebrospinal fluid. For the conditions under which one of us might be held to survive as a mere brain are just the conditions where it would make sense to talk of a *radically mutilated* human being, one reduced to the condition of a mere brain. (Johnston 1987a, p. 79)

Johnston endorses the view that we are (as currently constituted) humanoid entities, whilst holding that we can be reduced to the size of our brains.[159] Premise 2 is therefore falsified by Johnston's views. Thus, if the argument is to succeed, it must be supplemented with reasons to reject Johnston's proposal. I will not consider what reasons might be given here. Instead, I want to move on to consider a more subtle take on a similar kind of argument.

---

159. This picture ties in nicely with the views presented in previous chapters, according to which our persistence requires the continuous realisation of our mental states. I shall make more of this point in the conclusion of this chapter.

Chapter 6: The Problem of Overlappers

The more subtle argument begins with the observation that a lone brain sustains mental states which are not housed within any larger entity. Nor do they seem to bear phenomenological, volitional, or sensory relationships to the brain itself. These mental states seem not to be housed within any entity. Thus, the case of a lone brain seems to show that it is not always necessary that one's thoughts should be housed within a thinker. In contrast, whenever there are mental states, there also exists something which generates those states. This seems to show that the generation of thought is much more intimately connected with the existence of a thinker than the relationship of housing some thought. The argument suggests that we should therefore identify ourselves with the generators of our thoughts, rather than the entities which house them.

To respond to this argument, one must provide positive reason to think that there is an intimate connection between thinking and housing thought. We will provide arguments to this conclusion in section 6.4. For now, however, I want to make note of two general ways in which this could be done.

First, it could be argued that something can only think if it is capable of housing thoughts (or, stronger still, if it has once housed thoughts). At present, my brain-sized overlappers do not house my thoughts. Nor could my brain-sized overlappers house thoughts were they isolated; a lone brain does not house the thoughts it generates. It is reasonable to conclude that my brain-sized overlappers are incapable of housing thoughts (at the very least, I cannot easily see what would make them capable of doing so).[160] Hence, if this argument were to succeed, it would show that we are not identical to our brain-sized overlappers.

Second, it may be possible to argue that some kinds of mental state require the existence of an entity to house them. If we enjoy states of these kinds, and if it can be argued that these states belong primarily to the entities that house them, then this would provide evidence that we should be identified with our humanoid overlappers, rather than with our brains. Since all of our mental states have the same subject, this would again suffice to show that our mental states should be attributed to the entities that house them, and not to their generators.

We will have to wait until section 6.4 to see whether either of these avenues

---

160. In effect, this response follows Johnston's suggestion that we take the concept of mutilation seriously (Johnston 1987a, p. 79). A lone brain cannot think, but a mutilated person can. We explain this fact by reference to what the person would be capable of were she not stripped down to the size of a lone brain (i.e. we take seriously the idea that her present state does not reflect her 'characteristic form').

can be pursued. Now, however, I want to consider an alternative argument for the conclusion that we are brain-sized. This argument, which is explicitly endorsed by Derek Parfit and Jeff McMahan, was briefly presented in the previous chapter. I will argue that it is far from clear that the argument succeeds.

6.2.2 Accounting for Thinking Organisms

We now return to an argument for thinking subject minimalism which derives from comments by Parfit, McMahan, and Persson (though the latter does not ultimately endorse it).[161] This argument presupposes that we are not identical to the humanoid organisms located where we are. It states that we are nevertheless correct to ascribe mental states to these organisms. The argument aims to show that we can best explain the latter fact by accepting that we are small parts of those organisms.

The argument is quite involved, and so it is best presented in stages. First, we begin with the following two claims:

Non-Identity We are not identical to the organisms that contain our thinking parts.

Thinking Organisms The organisms that contain our thinking parts think.

The Non-Identity principle is motivated by thought experiments which appear to show that there are situations that we can survive, but which our organisms cannot. One such example is Surviving Cerebrum (presented on p. 139), in which my cerebrum is stripped away from the remainder of my body and artificially stimulated. As above, it is supposed that my envatted cerebrum could sustain a conscious life, and that this would guarantee my continued existence 'as' an artificially stimulated brain. It is also supposed that no organism could exist as an artificially stimulated brain. Hence, I have a property which no organism has. What goes for me, of course, is true of other human persons as well. Thus, we must conclude that we are numerically distinct from the organisms which contain our thinking parts.

Thinking Organisms is supposed to be a commonplace, reflecting our linguistic practices. We ascribe thoughts to organisms on an almost daily basis; we count dogs, cats, field mice, etc. amongst the things to which psychological

161. C.f. McMahan 2002, pp. 88–94; Parfit 2012; Persson 1999.

predicates can rightly be ascribed. There is no reason why the organisms located where we are should be incapable of thinking, whilst more simple animals, like our pets, can do so. Certainly, an account of personal identity should not entail that it is straightforwardly false that those organisms think.

These two principles suggest a problem, familiar already from section 6.1. If we are not identical to our organisms, and our organisms think, then there would seem to be two thinking things where each of us is – a person and her associated organism. This is prima facie absurd. Yet we cannot deny that we think. Thus, if we are to maintain both Non-Identity and Thinking Organisms, we must somehow argue that it is unproblematic that there are two thinking things where we are. Taking note of a distinction raised in the previous chapter, one can try to alleviate the absurdity by holding that only one of these things thinks in the strictest sense, whilst the other thinks derivatively (in virtue of its relation to the thing that thinks in the strictest sense). Since we presumably think in the strictest sense, our organisms must think derivatively. We can summarise this line of thought with the following argument:

1. If we are not identical to some entity which contains our thinking parts and that entity thinks, then there are two things thinking each of our thoughts.

2. If there are two things thinking each of our thoughts, then one of those things thinks derivatively.

3. We do not think derivatively.

C. The organisms which contain our thinking parts think derivatively.

If our organisms think derivatively, we are owed an explanation of why they do so – of why our thoughts can correctly be attributed to them, despite belonging to us in the strictest sense. Here, we can observe that organisms often bear properties derivatively in virtue of the operations of some of their parts; for example, we digest food in virtue of having parts, each of which digests it for us. It might be suggested that thought is no different; organisms think in virtue of having a part – their brain – which thinks on their behalf. This generates the following final argument:

1. If an entity thinks derivatively, it does so in virtue of having a proper part which thinks its thoughts non-derivatively.

2. We are the only thing to think our thoughts non-derivatively.

C. We are a proper part of everything non-identical to us which thinks our thoughts.

According to this argument, we can best preserve the intuitive claim that animals think by claiming that they do so in virtue of containing us as a part. As Parfit puts it:

> [H]uman animals think by having a conscious thinking part which is a person in the Lockean sense. We can call this the Embodied Person View. (Parfit 2012, p. 17)

The argument is interesting, and suffices to get the problematic of this chapter in sight – the question of what it takes for something to think 'strictly' or non-derivatively. However, I think that several of the moves that it makes are questionable. I will briefly outline why.

The Non-Identity Claim

For the above argument to succeed, it is crucial that we have reason to believe that we are not identical to the animals which share our thinking parts. This premise is supported by two further claims. First, it is sufficient for me to continue to exist that my cerebrum continues to exist. Second, it is not sufficient for the organism located where I am to continue to exist that my cerebrum continue to exist. I do think that both claims are intuitive. However, I have already argued that we might have incompatible intuitions regarding our own persistence. We should not overlook the possibility that this is the case here. We might be wrong either about what it takes for us to persist, or about what it takes for a human organism to persist.

Unfortunately for the Non-identity claim, I believe that this is very likely to be our situation. As evidence, we might consider the philosophical staple of a 'brain transplant' case (c.f. Shoemaker 1963, pp. 22–23, cited on p. 113 of chapter 4). The usual intuition about these cases is that one 'goes with' one's brain, obtaining a new body. To this it is added that animals cannot gain new bodies through brain transplants. If these judgments are both true, then we cannot be animals. However, it is far from clear that they are both true. We should note here that a brain transplant case is very easily described as one in which one person 'gets' another person's brain, and not as one in which a person gets a

161

new body.[162] Thus, it is not clear that we have settled intuitions about what would happen to us in a brain transplant case.

Worse still, it seems that some of our everyday practices for reidentifying persons support the claim that one's persistence does not depend upon the continued existence of one's brain; as I mentioned in chapter 2 (p. 56), we are highly inclined to think that someone can exist in a persistent vegetative state, in which she has no mental capacities.[163] Given this, I think that our intuitions about cases like Surviving Cerebrum do not offer clear evidence for claims about our persistence.

This is not the only direction in which our intuitions about Surviving Cerebrum can be questioned; it is also not entirely clear that the continued existence of our brains does not suffice for that of the organisms located where we are. Indeed, as Noonan notes, one can quite easily elicit the intuition that an animal, such as a dog or cat, would obtain a new body via a brain transplant:

> Brown and Robinson might be two dogs, with quite different personalities, skills and training. If Brown's brain was transplanted into Robinson's body [and the resulting dog were to display Brown's personality and skills] there would be just the same temptation to say that Brownson was Brown as in the human version of the case, but it could not be said that Brownson was the same *person* as Brown. As the same *what*, then, would we be tempted to identify Brownson and Brown? No answer seems to be available except as the same dog, the same canine animal. (Noonan 1998, pp. 305–306)[164]

Though I concede that there may be principled reasons for distinguishing our persistence conditions from those of the animals with which we overlap,[165] I conclude that our intuitive responses to cases fail to show that we should do so.

There is much more to say about the Non-Identity claim, and those who believe we are animals have produced subtle and interesting arguments to the

---

162. Indeed, though this may be idiosyncratic, I find that this description rolls off the tongue much more easily than the alternative claim that one person comes to possess another's body.

163. Of course, as mentioned above, moves can be made to reconcile this judgment with the claim that the survival of one's brain would suffice for one's own survival (see especially Madden 2016a). I am not unsympathetic to these moves. However, I should like to note that, when made, they undermine (if they do not entirely weaken) the claim that we are not identical to organisms.

164. For discussion of this point, see D. Mackie 1999a, pp. 371–373 and Noonan 2001, p. 86.

165. Thus, for example, Shoemaker 2016 suggests we can functionally define two types of continuity; biological continuity, which constitutes the continued existence of a living organism, and psychological continuity, which constitutes our continued existence.

effect that psychological considerations are irrelevant to an animal's survival.[166] Equally, there are interesting, and I think convincing, arguments to the effect that we would survive a case like Surviving Cerebrum.[167] All I mean to do here is to point out that our intuitive reactions to these cases are not entirely stable. If so, the proponent of thinking subject minimalism must provide us with an argument to the effect that we should simultaneously believe both that we can obtain new bodies through brain transplantation and that animals cannot.[168] Considered alone, our intuitions about Surviving Cerebrum fail to show that we are inclined to do so. Much more work is needed here if we are to have reasons to believe the Non-Identity claim.

Thinking Organisms

We can also raise doubts about Thinking Organisms. Two thoughts suggest themselves. First, it is plausible that we are, in some sense or other, animals, just as it is true that we, in some sense or other, have hands, feet, and so on.[169] If so, then we should distinguish two ways for Thinking Organisms to be true:

Strict Thinker There is something that is, in some sense or other, an organism and thinks in the strictest sense.[170]

Strict Organism There is something that is, in a strict sense, an organism and thinks in some sense or other.

---

166. C.f. Olson 1997, pp. 109–111.

167. I have in mind here the type of remnant person argument presented in Johnston 2007, pp. 45–47 and Johnston 2016. This argument is critically discussed by Blatti 2016, p. 176; Madden 2011, pp. 309–315; 2016b, pp. 202–205; Olson 2015a; Olson 2016; Snowdon 2014b, pp. 234–237; Toner 2014; Yang 2019.

168. It is worth noting in passing that the presence of good arguments for both claims does not suffice to show that we should accept both in conjunction (it is, of course, possible to have good arguments for two opposed claims). What is needed additionally is an argument showing either that these claims can unproblematically be held together, or that whatever problems arise if we hold both together are no worse than problems that arise if we reject either claim. This requires detailed comparative work, which is outside of the scope of these thesis.

169. Bailey 2016 uses this observation as the basis of a novel argument in favour of animalism; the position that we are identical to the (strict) animals located where we are. Although I am favourably inclined to Bailey's argument, I think that it should be noted that it is far from clear that the proponent of the view that we are derivatively animals is committed to any context in which we would count the number of animals by counting both ourselves and the strict animals located where we are (see, for example Rea 1998).

170. Shoemaker endorses this claim, whilst rejecting Strict Organism, in his aptly titled 'Thinking Animals without Animalism' (2016).

Chapter 6: The Problem of Overlappers

We are organisms in some sense or other, and think in the strictest sense. This validates Strict Thinker. Thus, at least one interpretation of Thinking Organisms is true. Given this, one might wonder whether there is any need for the other interpretation to be true; perhaps whenever we ascribe mentality to animals, we are talking about things like us, and not about biological organisms in the strictest sense possible. This suggests that we may be able to reject Strict Organism – the intended reading of Thinking Organisms – without flatly denying that the sentence 'organisms think' expresses a true claim.

If it is insisted that Strict Organism is the only eligible interpretation of Thinking Organism, one might simply respond that we are wrong to believe that organisms think. Of course, it would be absurd to straightforwardly deny that my dog's body houses a psychologically endowed entity. However, it might be argued that this entity is not identical to the organism whose thinking parts it shares; if cases like Surviving Cerebrum do indeed show that we, who think, are not organisms, then I think that it might fairly be suggested that we erroneously think that our organisms think because we mistakenly identify them with ourselves. Generalizing more widely, one might likewise claim that we are mistaken to believe that any organisms think – in each case mistaking them for a psychologically endowed entity which they house. If so, then upon realising our mistake, we ought to withdraw the claim that organisms think. This would be disastrous for the argument presented above. Thus, although I find Thinking Organisms very plausible (and do not seriously wish to deny it), I think that there is room to resist it.

The Explanation of Derivative Thought

A third questionable move is the first premise of the second argument – the claim that entities which think derivatively do so in virtue of having a proper part which thinks in the strictest sense. This certainly provides one explanation of the claim that our organisms think derivatively, but it is not our only option. Thus, for example, constitutionalists like Lynne Rudder Baker hold that we are constituted by, and therefore share our parts with, the organisms which contain our thinking parts.[171] When one entity constitutes another, Baker holds that they share their properties; one entity has those properties derivatively, and the other

---

171. See especially Baker 1997; Baker 2000; Baker 2016; D. Robinson 2016. See also the papers in Rea 1997.

has them non-derivatively. We think non-derivatively. If we are constituted by organisms, Baker's views therefore entail that they think derivatively.

Constitutionalism is controversial, and has the consequence that two entities can have the same microphysical properties whilst differing in their persistence conditions (and in the properties that they have non-derivatively).[172] However, Baker's position avoids some of the counter-intuitive consequences associated with thinking subject minimalism; she can claim that we have hands and feet, and are capable of seeing one another.[173] I conclude that an argument is needed to secure the claim that the best explanation of the existence of thinking organisms entails that we are proper parts of the organisms located where we are. Parfit and McMahan stop short of providing such an argument.[174]

In summary, I am unsure how much weight to give the argument that we should take ourselves to be brain-sized in order to account for our tendency to ascribe psychological properties to organisms distinct from us. It strikes me that the case for our distinctness from organisms has yet to be made out, that it is unclear how to interpret our ascriptions of thought to organisms, and that it is unclear why taking ourselves to be so small is the best way to account for the claim that animals can think. I conclude that the argument fails. I now want to consider one third, more empirically grounded, argument to the effect that we should be identified with brain-sized entities.

6.2.3 Gillett's Argument

I will now consider an argument to the effect that nothing humanoid could possess many of our mental states. This argument is due to Carl Gillett, who suggests that 'rich' psychological properties (those with cognitive content, such

---

172. For discussion, see Burke 1992; deRosset 2011; P. Mackie 2008; Paul 2010; Rea 1995; Rea 2000; Sutton 2012; Zimmerman 2002.

173. Of course, the thinking subject minimalist can reply that we speak truthfully when we say that we have hands – we have hands derivatively in virtue of being causally connected to the hands of something larger than us. There are interesting, quite involved, arguments to the effect that this proposal is at odds with our readiness to assert sentences which simultaneously ascribe mental properties and properties of our bodies to ourselves. See, for example, Stanley 1998. This reinforces the conclusion that we should reject the view unless very strong arguments cut in its favour.

174. Indeed, it is noteworthy that Parfit, at least, takes us to be things constituted by normal brains – so-called 'embodied persons' (Parfit 2012, pp. 15–17). Thus, he seems to accept the metaphysics of constitution in describing his views. It is therefore unclear why he gains any advantage by claiming that we are proper parts of the organisms within which we are located, rather than entities constituted by those organisms.

as thinking about Venus, or desiring cheese) are primarily ascribable to brains, rather than organisms. This is particularly interesting because I will soon argue for the opposite claim on the basis of very similar considerations.

Gillett takes his conclusion to follow from empirical work into the nature of psychological properties:

> [R]ich psychological properties are taken to be produced by stimulations of the sensory organs and to produce stimulations of the muscles as well as other physiological effects, along with being produced by, and producing, various other rich psychological properties of the same individual. The important points about such neuroscientific hypotheses for my illustrative purposes here are, first, simply that scientific work now posits rich psychological properties, and, second, that these hypotheses in the neurosciences take rich psychological properties to have such characteristic productive roles. (Gillett 2014, p. 45)

The idea here should be familiar from our discussion of functionalism in Chapter 3; an entity is in a particular mental state just in case it is in a state which plays a particular causal role (i.e. it is liable to bring about the types of effect characteristic of the mental state in question). Gillett claims that neuroscientists characterise rich causal properties by reference to their interactions with our perceptual organs, and with their effects on muscular movement. He then argues that this entails that our brains have those psychological states, rather than the organisms to which they belong:

> [C]ompositional levels of individuals within the same compositional hierarchy are productively closed; thus, individuals only productively interact with other individuals at the same levels. Parts and wholes within a compositional hierarchy therefore do not productively interact. Consequently, we can provide a reason why *brains* are taken to instantiate rich psychological properties in the sciences, rather than organisms. (ibid., p. .45)

To say that two entities do not productively interact is to say that they do not causally interact; changes in one's properties do not produce changes in the other's. The claim that entities at different compositional levels do not productively interact therefore amounts to the claim that objects do not cause changes in their parts, nor do their parts cause changes in them. With this clarified, we can formalize Gillett's argument as follows:

§6.2: Motivating Thinking Subject Minimalism

1. There are no causal relations between wholes and their parts.

C_1. Thus, if something is in a state caused by stimulation of its sensory organs, it does not have those organs as parts.

2. I am in perceptual state P.

3. To be in perceptual state P is (in part) to be in a state caused by stimulation of one's sensory organs.

C_2. My sensory organs are not parts of me.

Gillett's argument is highly interesting. However, I think that his intended readings of premises 1 and 3 are controversial. I begin with the former.

It is rather odd to claim that objects do not cause changes in their parts (and that their parts do not cause changes in them). Surely flipping the light switch in my kitchen causes the circuit of which it is part to become live. Similarly, it is hard to believe that when a gorilla paces, it does not initiate the movement of its legs. If Gillett's argument is incompatible with such obvious examples as these, it is wholly unconvincing.

Despite these examples, there is a reading of Gillett's first premise that I think we should grant. To understanding this reading, it is useful to look to one of the sources he cites in support of his argument:

> In our view, the phrase 'top-down causation' is often used to describe a perfectly coherent and familiar relationship between the activities of wholes and the behaviors of their components, but the relationship is not a *causal* relationship ... Rather, in unobjectionable cases both phrases describe *mechanistically mediated effects*. Mechanistically mediated effects are hybrids of constitutive and causal relations in a mechanism, where the constitutive relations are interlevel, and the causal relations are exclusively intralevel. (Craver and Bechtel 2007, p. 547)

This significantly clarifies what is involved in the claim that entities at different compositional levels do not productively interact. Craver and Bechtel's view is not the wild claim that we cannot say that parts cause changes in wholes, or vice-versa, but instead the quite reasonable claim that to do so is to speak about causation in a loose, unprincipled way; strictly speaking, changes at one level constitute, or are constituted by, changes at another level, rather than causing

or being caused by them. Thus, flipping the switch causes charges in the other components of the circuit. This, in turn, amounts to the circuit being live. It is permissible to say that flipping the switch causes the circuit to become live, but this is not to assert that a genuine causal relationship holds between the parts and the whole. The first premise of Gillett's argument is true if we read it as saying 'strictly speaking, there are no causal relationships between wholes and their parts'.

This clarification suggests a question: when neuroscientists say that mental states are caused by stimulation of the organs, or they cause muscular contraction, are they using the notion of causation loosely or strictly? In order for premise 1 to be relevant to their claims, they must be speaking strictly. However, I think that we can make a case for the claim that they are only speaking loosely; premise 3 should not be read as claiming that our perceptual states are caused, in the strict sense, by stimulation of our sensory organs. If I am correct, premises 1 and 3 of Gillett's argument are mismatched; the notion of causation required to validate each premise is incompatible with that required to validate the other.

To begin our case, note that there is something a little artificial about Gillett's description of the productive roles of our mental states. Gillett states that rich psychological states 'produce stimulations of the muscles', and that they are produced by stimulation of the sensory organs. Now, it is true that when I intend to do something, I ordinarily act by means of bodily movements, and that these involve muscular stimulation. However, descriptions of intentions and actions are ordinarily pitched at the personal level of behaviour, rather than the subpersonal level of muscular movement; an intention is an intention to *do* something or other, not for one's muscles to move in a particular way. This suggests that our intentions are borne out by something other than muscular movement. Indeed, it seems to suggest that our intentions are borne out by behaviours that have muscular movements as parts, rather than as strict causes. This picture has the advantage of corresponding more naturally to what we would ordinarily say about the functions of intentional states. Unfortunately, it seems to suggest the opposite conclusion to Gillett's argument.

The above seems to suggest that many of our psychological states are individuated by reference to causal interactions with the behaviour of something humanoid. Since they cause this behaviour, they cannot belong to a part of the humanoid entity. Nor, of course, do they belong to some other entity in its environment. Rather, what we have here is a causal relationship between different

states of a single entity; that the entity is in a certain perceptual state (and in a certain desiderative state) causes it to engage in a certain behaviour. If we accept this account, we will claim that the productive roles hypothesized by neuroscientists (according to Gillett) – those which make reference to stimulation of parts of humanoid entities – are associated with realisers of mental states, rather than with the mental states themselves. When we say we say that mental states cause muscular movement, we are exploiting a loose notion of causation – the 'top-down' causation described by Craver and Bechtel. This alternative model validates the claim that we have parts other than our brains. That this account provides a less artificial description of the productive roles of our mental states is a point in its favour.

I conclude that Gillett's argument does not successfully show that brains think. Instead, it seems to me to suggest the opposite conclusion. It seems that the 'productive roles' associated with mental states do not refer to the causal relationships in which they stand to different parts of our body, but rather to states of an entity composed of those parts. I conclude that the case against thinking subject minimalism looks quite good.

6.2.4 Assessing Thinking Subject Minimalism

To summarise: thinking subject minimalism answers the problem of overlappers by ruling that, strictly speaking, only the brain-sized generators of our thoughts think. We have considered three ways to argue in favour of the view that we are brain-sized. Though the latter two arguments fail, I have not responded directly to the first set of considerations; these highlight that we need to produce an argument in favour of the view that we should be identified with the entities that house our mental states. In the next section, I will consider one way to respond to thinking subject minimalism – Bennett and Hacker's argument to the conclusion that nothing brain-sized could be said to think. I argue that Bennett and Hacker are not successful in making their case. 6.4 pursues an alternative strategy, arguing that many of our mental states are primarily attributable to things larger than our brains.

6.3 Against Thinking Brains

The aim of this section is to consider whether one might outright deny that brains think, without comparing their claims to think with those of larger en-

tities which have them as parts. This is not an unusual response to the problem of overlappers. For example, Moran writes:

> [O]ur intrinsic conscious properties are grounded by solely intrinsic microphysical properties. Moreover, there are large proper parts of human persons that share these intrinsic microphysical properties. However, these large proper parts do not instantiate any intrinsic conscious properties, despite instantiating their (conditional) microphysical grounds. This is because such properties can only be instantiated by things of a certain kind. (Moran 2018, p. 386)

Similarly, Blatti observes:

> [I]t is not [brain-sized] proper parts, but the wholes of which they are constituents who act in ways suitable for description, interpretation, explanation, justification, rationalization; for praise and blame; for judgments concerning responsibility and culpability; for attitudes of approval and reproach; and so on. It is in these engagements with the whole being that we credit *it* – not its thinking parts – with various cognitive and affective capacities and their exercise. (Blatti 2016, p. 173)

I think that both Blatti and Moran are correct; brains are the wrong type of thing to think, and this is borne out by our ordinary ascriptive practices. However, it is hard to see how to argue for this claim; the fact that we do not ordinarily assert that one's brain thinks does not, by itself, should that we would be mistaken to do so.[175] This section considers a rather severe assessment of the claim that brains think; ascriptions of mentality to brains are not merely false, but senseless and confused. This line of thought is forcefully set out by Bennett and Hacker in their *Philosophical Foundations of Neuroscience* (2003). I shall reconstruct their argument, and argue that, though it is cogent, they have not given us sufficient reason to adopt their conclusion.

The core of Bennett and Hacker's argument can be gleaned from the following quotation:

> We can observe whether a person sees something or other – we look at his behaviour and ask him questions. But what would it be to

---

175. To be fair to Moran, this is not his object. Instead, he aims to show that, given that we think, but our large proper parts do not, one should accept that what properties a thing has can depend upon the kind to which it belongs.

> observe whether a brain sees something – as opposed to observing
> that it is the brain of a *person* who sees something. (Bennett and
> Hacker 2003, p. 90)

Their answer, of course, is that there could be no such thing; it is impossible to ascribe psychological predicates to brains on the basis of observation. They claim that if one cannot ascribe mental states to some entity on the basis of observation, one cannot ascribe mental states to them on any other basis. If they are correct, it follows that brains cannot be ascribed mental states.[176]

This argument relies upon a particular view about the nature of mental states. The lynchpin of this view is the idea that there exists a tight conceptual link between a mental state and the behaviour characteristic of one who is in that state. This link is such that one can ascribe mental states to other people simply through observing their behaviour, without the need for inference (ibid., p. 82).[177] Indeed, on this view, possession of a mental state is understood, in part, in terms of the types of behaviour in which one is liable to engage. It follows from this that something incapable of behaviour of the correct kind cannot be ascribed the mental state in question.

Looking inwards, we do see that the brain produces much of our behaviour. However, it is strange to say that the brain itself engages in behaviour. Further, insofar as we do ascribe behaviours to the brain, these consist in patterns of neural activity. To ascribe a psychological state to a person's brain on the basis of this activity, one would first need to inductively establish correlations between different types of brain activity and the person's mental states and behaviour. This is to concede the absence of a conceptual connection between the neural activity and mentality, undercutting the view of mental states that Bennett and Hacker advocate. Hence, their conclusion:

> *The brain is not a logically adequate subject for psychological predicates*.
> Only a human being and what *behaves* like one can intelligibly and
> literally be said to see or be blind, hear or be deaf, ask questions or
> refrain from asking (ibid., p. 72)

Bennett and Hacker are surely correct to say that we can ascribe mental states

---

176. An important precursor to this argument can be found in Kenny 1984. Dennett 2007 and Searle 2007 present responses to Bennett and Hacker.

177. Strawson presents a very similar view in chapter 3. of *Individuals*. There, he notes the existence of behavioural "criteria of a logically adequate kind" for the third-personal ascription of mental states (P. F. Strawson 1959, p. 106).

to other people simply through observing their behaviour. Thus, we should accept that there is a close conceptual connection between behaviour and mentality. However, we need not accept that this connection is quite so tight as to rule out the ascription of mental states to brains. I begin by noting two criticisms of Bennett and Hacker's views, before suggesting an alternative which is not subject to these criticisms.

First, Bennett and Hacker's position follows in the wake of a constellation of anti-Cartesian arguments from the early 20th century. According to the dualist, much of our bodily behaviour results from the activity of an unobservable, and separable, locus of mental activity. This locus of activity, or 'soul', is capable of thought when separated from any observable body. If correct, Bennett and Hacker's position shows that the idea of such entities is nonsense. We cannot observe Cartesian souls. Hence, to think that we can ascribe mental states to such entities is to and so neglect the conceptual connection between psychological predicates and observable behaviour. Though Hacker, at least, accepts this consequence,[178] I do not think that it is correct. Cartesian dualism is a view with a long pedigree. It is easy to convey, and many find it easy to believe; indeed, some psychologists suggest that something very much like dualism is our 'native' conception of persons.[179] I am inclined to think that it would be very surprising were it to turn out that a view which has such sway (and which is so easily characterised) were simply incoherent. Put otherwise, Bennett and Hacker's argument, if correct, requires us to revise our beliefs about what it is coherent to imagine (or, indeed, to believe). In order to set such limits on what is conceivable, it must not even be coherent to deny that their views of mentality are correct. Unfortunately, I think that we can find plenty of reasons to doubt that Bennett and Hacker are right.

There is a second response to consider. I do not always need to see a person, or to see her body move, in order to attribute mental states to her. Thus, for example, in communicating with someone over the internet, or by text message, I can frequently form and test beliefs about her feelings, opinions, and desires.[180] Indeed, I can do this even in conversation with a stranger, whose

---

178. Smit and Hacker defend this claim on pp. 1086–1087 of their 2014.

179. See, for example, Kuhlmeier, Bloom, and Wynn 2004 and Bloom 2004.

180. It is worth noting that this example involves a form of communication similar to that available to the brain in Surviving Cerebrum. Parfit writes that this brain can interface with the outside world through some device that enables it to "spell our words of messages to us" (Parfit 2012,

bodily behaviour I have never observed. This suggests that it is not necessary that a subject have a body in order for me to attribute mental states to her; all that is necessary is that she should interact with me in the right sort of way.

Of course, in the envisaged case, I still observe something that reflects my conversational partner's behaviour; I read the messages that I am sent. However, the crucial point is that my conversant's behaviour is exhibited through objects which are not part of her. Thus, it suffices for the ascription of mental states to someone that I see an object which manifests her behaviour, but I need not actually see her. The proponent of the view that we are brain-sized can say that this is precisely what happens when we 'see' someone act; we see behaviour of hers manifested through something (a body) upon which she acts. Thus, the claim that we need to observe someone's behaviour in order to attribute mental states to them does not clearly rule out the brain view.[181]

These two points do not decisively refute Bennett and Hacker's views. However, they do put pressure on them; we have at least some reason to be sceptical of the case that they present against the claim that brains can think. At this juncture, we should consider what positive reasons we have to uphold their position. As I read it, Bennett and Hacker's position is primarily motivated by the claim that there is a conceptual connection between mentality and behaviour. Unfortunately for them, it is easily shown that we need not accept their views in order to do justice to the idea that behavioural evidence is 'logically sufficient' for the ascription of psychological predicates to persons. I will round out this section by explaining how other views can also accommodate this claim.

First, let's be clear on what is involved in the idea of 'logical sufficiency'

---

p. 12). Thus, other than the mere fact that it is a lone brain, rather than an embodied person, producing the text that I read, Parfit's imagined case is exactly the same as this one. If I can ascribe mental states to the person with whom I am communicating via text, it is very unclear why I should not also be able to ascribe mental states to the lone brain in Surviving Cerebrum.

181. We should consider a response here. The response is that our practices of ascribing mental states to others depend upon a 'base case' in which we can observe someone, and interpret her behaviour as an indication of her mental state. Cases like that of the instant messenger are derivative, and rely upon our competence at interpreting one another when interacting face to face. Thus, one might insist that our inability to see brains, or to interpret them as engaging in behaviour, precludes us from ascribing mental states to them. The problem with this response is that it is not entirely clear that the claim about the base case is true. An example, which was suggested to me by Rory Madden, might help. Consider a range of short-sighted intelligent crustaceans. These entities cannot see each other, but nevertheless communicate by means of giant flags, with which they spell out words in semaphore. It is far from clear that these creatures would be unable to ascribe mental states to one another, despite being unable to see one another's bodily behaviour.

under consideration. Bennett and Hacker rightly observe that it would be a mistake to think that we only ascribe mental states to others on the basis of establishing empirical correlations between certain states of mind, and certain kinds of behaviour. It is not an empirical discovery that someone who is elated is prone to smile and talk excitedly; there is some sense in which the propensity to behave in these ways is just part of what we mean when we say that someone is happy. The key point that I want to press now is that this fact is accommodated by views of mentality on which terms for mental states are introduced to label the causes of certain kinds of behaviour. It will be helpful to begin with the concrete example of functionalism in the philosophy of mind.

Functionalists claim that mental states are individuated, in part, by their role in causing behaviour; to be in a particular mental state is, amongst other things, to have a property liable to cause the behaviour characteristic of that state. They hold that these causal connections are central to the definition of any mental state; it is not possible to be in a given mental state except insofar as one is in a state poised to play the correct functional role. Functionalists therefore hold that there is a conceptual link between one's behaviour and one's state of mind. This allows them to hold that we can infer someone's state of mind directly from their behaviour; the state of happiness, for example, is individuated by reference to the behaviour that it is liable to cause.[182]

We can generalise further. Take any view on which the reference of our terms for mental states is initially fixed by describing their behavioural effects. Such a view will uphold the conceptual connection noted by Hacker and Bennett; given the circumstances under which we introduced the terms in question, they must refer to the causes of certain kinds of behaviour, no matter what these causes are. It is easily shown that such views do not preclude my brain from being the subject of mental states; since my brain is amongst the causes of my behaviour, it is open for us to hold that my mental states are, in the first instance, states of

---

182. Similarly, in discussing another view – central state materialism – Keith Campbell says:

> [T]he Causal Theory of mind has two strands: that the various mental events and processes are postulated causes of segments of behavior belonging to various recognizable patterns, and that the mental causes are given their names in virtue of their postulated connection with those behavior patterns. The first strand admits the view that the mind is something inner, separate, and standing behind behavior …The second strand, that mental terms get their meaning by reference to the behavioral effects of the mental states they denote, preserves the truth of Behaviorism that there is a conceptual connection of mind with behavior. (K. Campbell 1970, pp. 80–81)

my brain.[183] It is not at all evident how this claim is unseated by the fact that we cannot observe the brain, and or by the claim that the brain does not itself behave, but only plays a causal role in the production of behaviour. Thus, the claim that there exists a conceptual connection between behaviour and mentality does not lend support to Bennett and Hacker's claims (though, of course, it does nothing to show that the brain thinks either – it is simply neutral on the topic of the present debate).[184]

I conclude that Bennett and Hacker fail to make their case that the ascription of psychological properties to brains is in breach of the logic of such predicates. Further, since a causal view of mentality is not obviously false, and dualism seems logically coherent, I think that we have good reason to be sceptical of their arguments. So, though it may be correct to refuse to ascribe psychological predicates to brains, I do not see how it can be claimed that this practice is senseless.

6.4 Derivative Attributions

I think that we are unlikely to find a good reason to deny that brains think by reference to Bennett and Hacker's views about mentality. It is therefore necessary to look for another way to defend the thesis that our thoughts should be ascribed to humanoid entities, rather than to our brains. One observation should help here. It seems to me that many of our ascriptions of psychological states to persons depend upon the fact that those states are housed within humanoid entities. If we are to ascribe these states to our brains, we therefore do so partly in virtue of relationships between them and entities larger than those brains. This strongly suggests that the psychological states are strictly attributable to those

---

183. Of course, functionalists typically require that a mental state's behavioural manifestations occur within the subject of that state; to have a mental state is to be in a state liable to cause one to behave in the right sort of way. My aim in this passage is just to note that this is not compulsory.

184. One might respond here that it is misguided to think that mental states stand in causal relationships to behaviour. Gilbert Ryle famously lampooned this 'quasi-mechanical' account of psychological vocabulary (Ryle 1949, , c.f. in particular pp. 8–13). Instead, he claims, psychological predicates serve to convey information about the properties that an object will come to possess under certain conditions. Though Hacker and Bennett do not mention Ryle by name, I suspect that they may have something like Ryle's picture in mind when they claim that the 'criterial grounds' (the logically good evidence) for the ascription of a psychological predicate are partly constitutive of its meaning (e.g. Bennett and Hacker 2003, pp. 82; 83). It is fair to say in response that their arguments do not obviously succeed in discrediting causal analyses of mental states.

larger entities, and only derivatively attributable to anything brain-sized.

To see how the argument works, we might return to the following thought:

> Surely it couldn't be the case that you think only in a derivative sense: that you think only insofar as you have a part that thinks strictly speaking. As Chisholm said, if there are now two things thinking your thoughts, one doing it on its own and the other such that its thinking is done for it by something else, you are the one that thinks on its own. (Olson, p. 32)[185]

As we saw in both this and the previous chapter, thinking subject minimalists often appeal to a principle of this kind. However, this only supports the view that we are brains if our brains think on their own. I want to argue that this is not always the case. Many of our mental states can only be attributed to us on the basis of co-operation between our thinking parts and other parts of a larger entity. The suggestion that we are the entity that thinks on its own would therefore suggest that we are this larger entity, rather than its minimal thought-generating part.

6.4.1   Precursors

Let's begin by considering some views similar to the one I will present. As we have seen, Michael Burke elaborates a number of ways in which our psychological states particularly target the humanoid entities within which they are housed. He takes this to strongly support the view that consciousness is a 'maximal' property, where:

> [K]ind/property/term/concept C is *maximal* just in case necessarily, no identity-sufficient part of a C is itself a C ... Something is an *identity-sufficient* part of a C just in case the particles composing the part would immediately compose *that very* C, if the complement of the part suddenly (as a rare, uncaused quantum event) ceased to exist. (Burke 2003, pp. 112–113)[186]

185. The Chisholm referred to here is Chisholm 1976, p. 104. It's worth noting though that Chisholm's concern is not with parthood, but with the claim that we are constituted by distinct thinking things (à la Locke's thinking substances), or that we constitute thinking things to which we are not identical. He thinks that both suppositions should be denied.

186. Burke further defends his views in Burke 2004. Ted Sider endorses the maximality principle in Sider 2003 (in response to Merricks 1998a). Madden 2015a argues for the related claim that conscious subjects must be 'integral wholes' (i.e. such that one can pass between any two of its parts without passing through an exterior surface, and that one cannot likewise pass from its

§6.4: Derivative Attributions

We should note in passing that this does not solve the problem of overlappers unless every part of us which could support consciousness on its own is also an identity-sufficient part of us. Thus, to accept Burke's response, one must hold either that nothing constituted by a lone brain could think, or that our brains are identity-sufficient parts of us. I find the latter extremely plausible, though (as we have saw when discussing the claim that we are not organisms) it can be doubted. Indeed, this connects to two points made earlier. The first, and more obvious point, is the suggestion of chapters 1–4 that our persistence requires the physically continuous realisation of our psychological lives. The second point was made earlier in this chapter, when discussing the example of lone brains. I suggested that our best hopes for a response may lie in seeing brain-sized thinkers as potentially larger than brains. One presumes that, inversely, one will have to see at least some larger thinking entities as reducible to the size of brains in order for this response to succeed. Thus, Burke's suggestion integrates well with the positive claims of this thesis.

Burke's view is appealing. However, I do think that more is needed to support the claim that our thoughts are primarily ascribable to the entities within which they are housed, rather than to the things which generate them. Here, it will be helpful to turn to an argument tentatively developed by Rory Madden in his 'Thinking Parts' (2016), and 'Animal Self-Awareness' (2017).

The basis of Madden's argument is the idea that there may be kinds of thoughts which we can think, but which our other overlappers cannot. More specifically, Madden suggests that humanoid entities may have the capacity to think about themselves first-personally, whilst brain-sized entities may lack this capacity. The reason for this is that only the former meet a necessary condition for the capacity for first-personal thought – that of self-acquaintance through distinctively first-personal 'reflexive channels'. Madden writes:

> My recommendation is to locate reflexive channels among the monitoring systems of proprioception, kinesthesis, the sense of balance, visual self-location (and indeed inspection of bodily sensations). For surely a case can be made that the evolved function of these sensory channels is to monitor the state of the whole organism. That's what these systems are for. The object presented to consciousness by these sensory systems is the whole organism. (Madden 2017, p. 12)

parts to the parts of an entity distinct from it). Kovacs discusses the maximality principle over pp. 280–282 of Kovacs 2010, and Madden over pp. 196–198 of his 2016b.

Chapter 6: The Problem of Overlappers

If Madden is correct, then the capacity to think first-personal thoughts can only be attributed to the humanoid entities presented to us by these self-monitoring systems. Since we do think first-personal thoughts, it follows that we are identical to these entities, rather than to any of their parts. In passing, we should note that this argument does not directly support Burke's 'maximality principle'; rather, it supports the claim that we are to be identified with the entities within which our mental states are housed. This seems to me plausible.

I find this line of argument extremely attractive, and have no serious doubts to raise about it. Nevertheless, I think that it would be helpful to consider whether supplementary arguments can be given to the same conclusion, without relying upon such a special (and philosophically vexed) instance of thinking as first personal thought.

Here, we might usefully reflect upon some comments made by Aquinas, in consideration of a structurally similar debate about whether we should be identified with our souls (the seats of our intellectual powers), or with an entity composed of both a soul and a body. Aquinas writes:

> [I]t is evident that certain operations, whereof the soul's powers are the principles, do not belong to the soul properly speaking but to the soul as united to the body, because they are not performed except through the medium of the body – such as to see, to hear, and so forth. (Aquinas 1265–1274/2017, Q. 70, art. 1)

If Aquinas is correct, some of our mental states can be attributed to our souls only insofar as they are unified with our bodies. He later argues that this shows that we cannot be identified with our souls:

> [I]t has been shown above that sensation is not the operation of the soul only. Since, then, sensation is an operation of man, but not proper to him, it is clear that man is not a soul only, but something composed of soul and body. (ibid., Q 75, art. 4)

Aquinas' argument begins with the claim that some mental operation is not conducted by the soul in isolation, and concludes that it must therefore be attributed to something with parts other than the soul. This is precisely the form of argument that I mean to pursue. I do not think that it will be possible to argue that brain-sized entities are simply incapable of the types of thought I will consider. However, I do think that I will be able to argue that they are at most derivatively capable of such mental states. The next subsection pursues this

line of argument with respect first to perceptual states, then to singular thought, and – finally – to ascriptions of mentality in general.

6.4.2 Ascription-based arguments

The aim of this subsection is to argue that some kinds of mental states can only be attributed to us insofar as our brains interact with other parts of the bodies to which they belong. I take this as strong evidence that the mental states in question can be attributed to our brains at most derivatively, and that they primarily belong to the humanoid entities within which they are housed.

Let us begin where Aquinas does – with the example of perceptual states. It is clear that what Aquinas says about the relation between perceptual states and the soul can also be applied to the brain; if my brain perceives some object in its environment, it does so partly because my sensory organs are attuned to the object's presence. Thus, it is not the case that the brain perceives by itself. Rather, the brain perceives by interacting with a larger system which functions to generate perceptual states. That system belongs to a large humanoid entity, which houses the perceptual states produced. I conclude that perceptual states are ascribed to the brain derivatively.

I would like briefly to comment on the relationship between this argument and wider issues in the philosophy of mind, relating to the analysis of perceptual states. Due in large part to the influence of Timothy Williamson, it is sometimes held that mental states such as seeing are primitive.[187] This means that they cannot be analysed via the conjunction of an epistemically neutral 'internal mental component' (such as a sensational or representational state, which can be described whilst remaining neutral about whether or not it it reflects how things actually are), and an 'external factor' relating to conditions in one's immediate environment. The argument trialed here is consistent with this position; to say that seeing involves an interaction between one's eyes and one's brain is not to say that it consists in the causation of an epistemically neutral mental state by patterns of retinal stimulation. Indeed, if it were not consistent with this position, then Williamson's knowledge-first project would be a non-starter; it is simply true that seeing involves a causal process by which the brain decodes information about its environment from retinal input.

187. See, in particular, chapter 3 of Williamson's *Knowledge and its Limits* (2000).

Chapter 6: The Problem of Overlappers

Similarly, in fairness to my opponent, it should not be assumed that in thinking of mental states as belonging primarily to brains, they are committed to denying that knowledge is primitive; they can reject the presence of an epistemically neutral mental state by saying that the fundamental description of the mental states ascribable to a brain depend upon whether or not it is interacting correctly with the world, via its sense organs. Thus, a knowledge-first epistemology does not have obvious implications for the debate at hand.

Taking our perceptual states to be primitive has no obvious implications for the debate under discussion in this chapter. What about the alternative position, according to which seeing involves an epistemically neutral 'purely internal' component? Here, there are two things to say.

First, the argument is not obviously inconsistent with the position. If seeing consists in the obtaining of a neutral 'internal state' (seeming to see, having a sensory impression etc.) under the right external conditions – conditions which guarantee that the internal state mirrors reality – then it will remain the case that these external conditions include states of the perceptual system outside of the brain. Thus, if one ascribes sight to one's brain, one still does so on the basis of its interactions with other parts of its humanoid overlapper.

Second, the argument is nevertheless less convincing on this analysis than upon the view that knowledge is primitive. If seeing is a primitive state, some of our mental states belong to brains only in virtue of their relationships to our sense organs. If, however, one thinks that seeing can be analysed into more basic components, one might respond that seeing is not, itself, a mental state. According to this view, the mental state involved in seeing is 'really' a more primitive representational state, and to say that someone sees something is just to say that the state occurs under certain conditions.[188] If so, then one might still

188. The following quotation from Williamson is especially clear in setting out this possibility:

> Someone might expect knowing to be a state of mind simply on the grounds that knowing p involves the paradigmatic mental state of believing p. If those grounds were adequate, the claim that knowing is a state of mind would be banal. However, those grounds imply only that there is a mental state being in which is *necessary* for knowing p. By contrast, the claim that knowing is a state of mind is to be understood as the claim that there is a mental state being in which is necessary and *sufficient* for knowing p ... On the standard view, believing is merely a state of mind but knowing is not, because it is factive: truth is a non-mental component of knowing. (Williamson 2000, pp. 21–22)

J. Nagel 2013 also provides an extremely helpful overview of what is at stake here.

hold that the brain thinks on its own, though how we describe its mental states depends upon their relationship to the parts of a larger entity.

Because the argument from perception is not entirely convincing, it will be helpful to sketch two ways in which it might be generalised. There is no hope of providing an argument that will convince the proponent of every possible position. However, I want to show that it is not just perceptual states whose ascription depends upon perceptual relationships between us and our environment. To resist this style of argument, one would have to adopt a narrow, and quite unappealing, view of the ontology of mental states.

Singular Content

To expand upon the argument raised above, I would like to consider how the having of perceptual states integrates with a wider picture of the nature of our mental states. I begin by considering the notion of singular content.

There is an intimate connection between the content of many of our mental states and the perceptual states which we have had in the past. Typically, we think about what we have once seen, heard, tasted, or felt. Even when this is not so (when, for example, we think about historical figures, such as Julius Caesar), we typically require information about the things about which we think. This information is gathered perceptually; I see traces of someone's presence, listen to the testimony of someone nearby, or so on.

I suspect that what can be said about belief and thinking carries over to other states as well. For example, to be afraid of some entity, one must be able to think about it; typically, this involves the exercise of one's perceptual capacities. Generalising somewhat, it seems to me that any state to which singular content can be ascribed depends upon past exercises of perceptual capacities. If so then to ascribe such mental states to a brain typically requires one to see it as (at least once having been) part of a larger perceptual system.[189] Once again, then, we

---

189. Complications are raised by the example of a brain in a vat. Imagine that the brain has been artificially stimulated from birth to simulate perceptual states. This stimulation does not involve interactions between the brain and perceptual organs. If so, then it is plausible that none of the brain's states are to be ascribed to it in virtue of its relationship to an entity of which it is part. There are two options here. First, we might deny that the brain really thinks thoughts with singular content; it might essay 'mock thoughts' (c.f. Evans discussion of Frege on pp. 10–14; 22–33 of his 1982), which fail to refer because it has never had perceptual contact with entities of the kind that its perceptual systems are designed to track. Alternatively, we might claim that the brain does manage to refer to simulated objects. Even so, this example does not represent the 'normal' operation of the brain, by which we come to be in a position to refer to entities in our environment.

find that the analysis of the psychological states of brains seems to point backwards to the humanoid entities which house our mental states, as was suggested towards the end of subsection 6.2.1.

Thus, the interpretation of many of our mental states is tied up with the past ascription of perceptual states to us. I draw from this the conclusion that anyone who wishes to deny that knowledge and perceptual states are 'really' mental states, claiming instead that they are simply descriptions of more basic representational states, must equally claim that these basic representational states do not 'really' have the singular content that we often ascribe to them; to say that someone believes something, on this view, is simply to adopt a useful convention for describing their mental states. This is a radical proposal, leading (it seems to me) to an attempt to conceive of all of our psychological attributions in terms of the sensations which prompt them.[190] Such proposals have fallen out of favour over the second half of the 20th century, and I doubt that the defender of thinking subject minimalism will wish to commit to such a view of the mind. I therefore conclude that we have good reason to attribute states with singular content primarily to humanoid entities, and not to their smaller overlappers.

Functionalism

A third point relies upon a substantive position in the philosophy of mind. This is a position that we have encountered before – functionalism about the nature of mental states. According to this view, mental states are individuated by reference to their typical causes and effects. If functionalism is correct, what I have said can be extended still further. For any mental state individuated by reference to states which may have singular content – beliefs, desires, etc. – the ascription of that state to a subject presupposes that it is the type of entity which may have states of the other kind. And, as we have seen, to be capable of contentful states, a brain must have either been embodied, or be such that its typical or 'normal' state is that of being embodied. So, even if some non-perceptual psychological states lack singular content, any argument to the effect that they are individuated by reference to their typical interactions with mental states with such content will suffice to show that brains can have such states only insofar as

---

190. I have in mind here something like the phenomenalism of Carnap's *Logical Structure of the World* (1928/2003) and C. I. Lewis's *An Analysis of Knowledge and Valuation* (1946). For one influential objection to views of this kind, see Chisholm 1948.

they are, or have been, the brains of larger entities, with functioning perceptual faculties.

More can be said here. First, it seems to me that the cognitive dynamics of many of our mental states is predicated on the fact that we are able to perceive, and to act on, our environment. Thus, for example, the reason that seeming to hear rain will lead me to think about the location of my umbrella is because auditory sensations typically convey information about the world. Generalising, we might say that to understand and rationalise many of the causal connections definitive of our mental states, one must acknowledge that they paradigmatically reflect a point of view on the world – that of a humanoid entity with working perceptual systems. Yet again, then, the attribution of certain mental states to us draws in the idea of something with parts other than those of our brains.

Though I am less certain about this, there is an additional way in which functionalism may bolster the arguments of previous subsections. This is that a functionalist may be particularly willing to accept Williamson's arguments to the effect that knowledge is unanalysable (and to generalize to the claim that perceptual states, such as seeing, are genuinely mental states). These arguments rely upon the thought that if we ascribe knowledge to an agent, we will typically make different predictions about her future behaviour than we will if we merely ascribe false belief to her (Williamson 2000, pp. 75–80). Imagine for an example that I seem to hear rain falling on the street outside. If it is raining, then I am very likely to get confirmatory evidence as I proceed about my business; I will see rain fall as I look outside, my partner will counsel me to take an umbrella to go to the shops, etc. If I am merely subject to an auditory hallucination, I will likely soon encounter evidence that it is not really raining. The upshot of this is that one's predictions about the future behaviour of someone who seems to hear rain are likely to be more robust (and to extend further into the future) if it is raining than they will if it is not. A functionalist can accommodate this by saying that the causal profile of a genuine perceptual state is more far-reaching than that of a perceptual illusion. If so, they will be in a position to privilege perceptual states as real mental states. As noted above, this will buttress my argument to the effect that some of our mental states are primarily attributable to entities larger than the brain against at least one source of criticism.

Of course, more would need to be said to defend this argument in detail. Functionalism is a controversial theory, and I have done nothing to defend it here. Second, in citing Williamson, I have extended an invitation to the func-

tionalist to think of perceptual states as genuine mental states; I have not given them conclusive reason to do so. Thus, one fears that it may still be possible to view the mind as composed by a 'primitive streak' of mental states for which it is impossible to show the centrality of our perceptual organs. Nevertheless, we do not usually speak in terms of such a primitive streak of mental states' when we read an account of the mind off of our ordinary attributive practices, we find that many of the states that we typically ascribe to ourselves, and to one another, are described in terms of the ways that they are housed. I hope to have at least shown that these linguistic practices present a strong presumptive case for the view that our mental states belong primarily to humanoid entities, containing perceptual systems, rather than to their brain-sized parts.

6.5   Conclusion

In my vicinity, there are a number of entities that seem to have what is required to think. Yet, it seems absurd to claim that there are multiple thinkers where I am. The problem of overlappers asks us to justify ascribing psychological properties to one of these overlappers, whilst denying that the others have such properties (except in a derivative sense). In answering this challenge, we can focus our attention on just two types of overlapper – those that are brain-sized, and those that are humanoid. These stand in interesting relations to my mental states in which none of my other overlappers stand. I have argued that there are no obvious grounds for refusing to predicate mental states of brain-sized entities. Nevertheless, it seems to me that many of our mental-state ascriptions presuppose the existence of something larger than the brain. This suggests to me that the primary subjects of our mental states are larger than the brain, and that these states are only derivatively attributable to anything brain-sized.

This conclusion allows us to draw together two thoughts that have been explored in this thesis. The first of these thoughts is the claim that our persistence conditions are, in some sense psychological, though they require the continued existence of some material entity which realises our mental states. The second of these thoughts is that we are nevertheless humanoid, having parts other than those of our mental states. In the next chapter, I will explain how a picture of this kind might lead us to revise a general theory of the nature of material objects – Kit Fine's theory of embodiments.

De-Fining Material Things

The previous chapters suggested a view of our nature. According to this view, we are humanoid, but possess a privileged 'central part', whose continued functioning is both necessary and sufficient for our persistence. I will now change tack somewhat, asking what implications this should have for a more general account of the nature of material objects – Kit Fine's theory of embodiments.[191] Fine's theory is custom-fitted to an intuitive distinction between 'timeless' and 'temporary' parthood – a distinction to which one might appeal in order to distinguish an object's central parts from those which it can lose without ceasing to exist. Nevertheless, I argue that Kit Fine's theory is unable to account for objects, such as ourselves, which have both timeless and temporary immediate parts. I call such objects 'nucleated wholes'.

Fine's theory is worth engaging with for a number of reasons. First, as I have noted, he makes use of a distinction by which we may try to account for the difference between those parts which are necessary for our continued existence, and those which are not. Second, Kit Fine's theory is rigorously formulated, and successful in its ambition to distinguish an object's bona fide parts from other entities which overlap with it.[192] Third, it is extremely compelling – particularly in its presentation of a 'hierarchical' account of parthood (on which an object's parts are nested within one another, like the members of a set). Despite these factors, the theory of embodiments has received relatively little critical attention.[193]

---

191. Fine's views are presented in most detail in his 1999 'Things and Their Parts'. Fine also discusses the theory of embodiments in his 1982; 2007. Even more recently, Jacinto and Cotnoir's 2019 provides a formal semantics for the theory of embodiments, and Fine's forthcoming 'Acts and Embodiment' uses the theory of embodiments to provide an account of the identities of acts.

192. On which, see pp. 62–65 of Fine 1999.

193. For an exception, see Koslicki 2007, and part 2 of her Koslicki 2008.

Chapter 7: De-Fining Material Things

In this chapter, I develop a novel problem for Fine's account. I claim that there exist material objects – so-called 'nucleated wholes' – whose mereological structure the theory cannot characterise.

Fine's theory is structured around two 'operations of composition', which explain how objects can be generated from their (immediate parts). The operation of rigid embodiment creates a whole from a series of parts and a relation in which they stand. Variable embodiment, in contrast, creates a different type of whole. A variable embodiment exists at a time in virtue of being 'manifested' by a rigid embodiment, whose mereological structure it takes on (at that time). My suggestion amounts to the claim that we, and other nucleated wholes, can be neither rigid nor variable embodiments.

The challenge arises because of the way in which Fine's theory accounts for the holding of relations of timeless and temporary parthood. According to Fine, the operation of rigid embodiment creates an object whose most immediate parts are timeless. As a result, a rigid embodiment's mereological structure is fixed; the rigid embodiment cannot lose or gain immediate parts. In contrast, variable embodiments are subject to mereological change – a fact accounted for by making all of their material parts temporary, rather than timeless. This renders it difficult to account for objects which combine a fixed aspect with a variable aspect – objects like you and me. I call such entities 'nucleated wholes'. Intuitively, we would like to say that a nucleated whole has both timeless and temporary immediate parts. However, the details of Fine's account preclude this. Thus, I claim, Fine's theory is unsuccessful as a general account of the nature of material things.

The first three sections of this chapter present my argument. 7.1 provides an outline of the distinction between timeless and temporary parthood. I then summarise the theory of embodiments in 7.2, highlighting three consequences. First, rigid embodiments have their immediate parts timelessly. Second, rigid embodiments exist only when all of their immediate parts exist. Third, variable embodiments have only temporary material parts. This leads into 7.3, where I introduce the notion of a nucleated whole (using the example of a castle), and argue that Fine's theory cannot account for such objects.

I consider four responses in 7.4. The first two of these responses maintain that nucleated wholes are variable embodiments – first by denying that nucleated wholes have timeless material parts, and second by modifying Fine's theory to permit variable embodiments to have timeless material parts. The latter

two responses both deny that nucleated wholes ever lose or gain immediate parts. They thereby allow us to conceive of nucleated wholes as rigid embodiments. I argue that all four responses should be rejected. I conclude that Fine's theory of embodiments is not, as yet, able to convincingly accommodate nucleated wholes such as ourselves.

7.1 Timeless and Temporary Parthood

Central to Fine's view is a distinction between "two different ways in which one thing can be part of another" (Fine 1999, p. 61). This section will introduce these two species of parthood. Having done so, I will explain Fine's ontological theory and argue that it cannot accommodate material things which are changeable in some respects, but rigid in others – here labelled 'nucleated wholes'.

Fine begins with the notion of temporary parthood. One object is a temporary part of another if it is part of that object in a way that is relative to time – part of the object at a particular moment in time, rather than part of the object with no reference to time whatsoever. As such, it is possible for an object's temporary parts to be part of it at one time, and not at another – as with the newly installed carburettor of an old car, and the molecules in my body (p. 61). It makes sense to speak of when an object came to have its temporary parts, and when those parts ceased to be part of the object.

In contrast, if one thing is a timeless part of another, "it is not appropriate to ask when, or for how long, it is a part; it just is a part" (p. 61). An object's timeless parts are part of it without respect to time at all. As such, there can be no time at which an object exists and its timeless parts are not part of it. Fine gives several examples of objects with timeless parts, including the jacket and trousers of a suit and the slices of bread and ham that make up a particular ham sandwich.

The recognition of timeless and temporary parthood is central to Fine's ontological theory – the theory of embodiment. The next section of this essay will introduce this theory. Following this, I will suggest that it leaves little room for a type of object that we intuitively recognise (roughly speaking, one organised around a central nucleus which is timelessly part of it). I shall then consider some responses on Fine's behalf, and argue that they are unsatisfactory.

7.2 The Theory of Embodiments

The theory of embodiments introduces us to two types of object – rigid embodiments and variable embodiments. These objects are generated by two formal operations of composition – the operations of rigid and variable embodiment – each of which is defined by a series of postulates. This section will provide an informal gloss on each type of object before detailing some of the postulates guiding the corresponding operation of composition. Three observations are key to the argument to come. First, rigid embodiments have only timeless immediate parts. Second, they exist only when their parts exist (and are appropriately arranged). Third, all of a variable embodiment's material parts are temporary.

7.2.1 Rigid Embodiment

The notion of rigid embodiment is foundational to the theory of variable embodiment.[194] We shall therefore begin here. A rigid embodiment is constructed from a number of material objects and a relation in which they stand (the whole's 'principle of rigid embodiment'). Roughly speaking, we may think of the rigid embodiment as an entity consisting of those objects standing in that relation.[195] This is an entity which exists in virtue of the fact that the objects stand in that relation and, as such, exists only when they do so.

Fine gives a neat example of a rigid embodiment – a ham sandwich produced by placing a single slice of ham between two slices of bread. When the slice of ham is enclosed by the pieces of bread, the operation of rigid embodiment will generate a new whole – a ham sandwich – which we can designate using the notation 'Ham, $Bread_1$, $Bread_2$ /between'. Given that the ham sandwich consists in the ham placed between the two pieces of bread, we might expect it to have the ham and the pieces of bread as timeless parts. This is exactly what we find. By postulates R4 and R5 of Fine's system, the ham, both slices of bread, and the relation of betweenness are timeless parts of this whole:

R4: The objects a, b, c, . . . are (timeless) parts of a, b, c, . . . /R (Fine 1999, p. 66)

---

194. This notion is prefigured in Fine's notion of a 'qua object' in his 1982.

195. Indeed, this seems to be exactly how Fine wishes us to think of rigid embodiments: "given some flowers and given the relation of being bunched, there will be a new object – the flowers in the relation of being bunched (what might ordinarily be called a 'bunch of flowers')" (Fine 1999, p. 65).

R5: The relation R is a (timeless) part of a, b, c,... /R. In this case, the relation R itself is taken to be a (constitutive) aspect of the corresponding rigid embodiment (ibid., p. 67)

To introduce a novel piece of terminology, we may call the ham, the two slices of bread, and the relation of betweenness the 'immediate parts' of the ham sandwich; they are part of it without being part of any of its proper parts (except themselves).[196] R4 and R5 effectively state that a rigid embodiment's immediate parts are timeless. This is one of the critical points for the argument to come.

My argument relies upon one more observation about rigid embodiments. As stated above, a rigid embodiment exists only when its immediate parts both exist and stand in the relation specified by its principle of embodiment:

R1: The rigid embodiment a, b, c,... /R exists at a time t iff R holds of a, b, c,... at t (ibid., p. 66)

Thus, for example, the sandwich exists just for as long as the piece of ham remains between the two slices of bread. Again, this observation is critical to the objection I raise.

To summarise, a rigid embodiment's immediate parts are timeless. Further, the rigid embodiment exists only when each of its immediate parts exists (and stands in an appropriate relation to its other immediate parts). With this covered, let us turn to variable embodiment.

7.2.2 Variable Embodiment

Unlike rigid embodiments, many ordinary entities need not retain their immediate components throughout their careers; thus, for example, a car's engine can be removed and replaced, though it is not (let us assume) part of any larger component of the car. To accommodate objects like these, the theory of embodiments must identify a second type of entity – a variable embodiment. The aim of this subsection is to explain the operation of variable embodiment. Yet again, I wish to draw out a key observation for what is to come; I will establish that all of a variable embodiment's material parts are temporary.

Variable embodiments have a more flexible mereological structure than rigid embodiments. Fine accommodates this by the semi-technical notion of a vari-

---

196. Equivalently, they are the objects which the operation of rigid embodiment most directly operates upon.

able embodiment's 'manifestation'. On his view, a variable embodiment, whenever it exists, is constituted by another entity – its manifestation at that time. This manifestation accounts for the variable embodiment's mereological structure at the time in question (Fine 1999, p. 70). The variable embodiment may be constituted by different manifestations at different times, as determined by its 'principle of variable embodiment' – roughly speaking, a function from the times at which the variable embodiment exists to its manifestations at those times.[197] Thus, where /F/ is a variable embodiment with principle F:

V1: The variable embodiment f = /F/ exists at time t iff it has a manifestation at t (ibid., p. 70)

Variable embodiments can readily undergo mereological change in virtue of having different manifestations at different times. Variable embodiment therefore promises to account for those material objects which are not rigid embodiments. I will shortly argue that there exist entities which cannot be represented as either rigid or variable embodiments. First, though, I want to consider how the above interacts with the notions of timeless and temporary parthood.

It is relatively easy to see why, according to Fine, variable embodiments serve as "the ultimate source of temporary parthood" (ibid., p. 71). If a variable embodiment exists at time t, its manifestation at t determines its parts at that time. By postulate V4, a variable embodiment's manifestation at a time is, at that time, a temporary part of the variable embodiment:

V4: Any manifestation of a variable embodiment at a given time is a temporary part of the variable embodiment at that time (in symbols: $f_t \leq_t f$) (ibid., p. 70)

If a variable embodiment has its manifestation as a temporary part, then it would be reasonable to expect the parts of that manifestation also to be temporary parts of the variable embodiment. This result follows from postulate V5a of Fine's system:

V5a: If a is a timeless part of b that exists at t and if b is a part of c at t, then a is a part of c at t (ibid., p. 70)

197. Typically, a variable embodiment's manifestations will be rigid embodiments (per Fine 1999, p. 69). However, there is no reason why some of its manifestations should not be other variable embodiments – a fact picked up by Koslicki 2007, p. 145.

Since a variable embodiment's material parts are all part of its manifestation, it follows from the above that variable embodiments have only temporary material parts – a point which is of importance in what follows.

The above lays out, in broad brush strokes, some of the central tenets of Fine's theory of embodiments. I have highlighted three as particularly important: 1. A rigid embodiment has only timeless immediate parts; it has no temporary immediate parts; 2. A rigid embodiment exists only when its immediate parts exist; 3. A variable embodiment has only temporary material parts. I will now argue that these commitments leave us unable to account for nucleated wholes – material objects with both timeless and temporary immediate parts.

7.3 Nucleated Wholes

On Fine's theory, variable embodiments have only temporary material parts. Hence, if an entity has a timeless (material) part, it follows that it is a rigid embodiment. Rigid embodiments have no immediate temporary parts (although they may have temporary parts through having timeless parts which are themselves variable embodiments). Hence, no entity can have both immediate timeless and immediate temporary material parts. I want to contest this consequence of the theory of embodiment.[198] Such entities do, in fact, exist. Thus, Fine's theory of embodiments is unable to account for every type of material object.

Though I have argued that we are examples of the kind under discussion, it may be helpful to focus on a simpler, less contentious, example. For such an example, let us look to the following notion of a castle:

> A castle comprises a central keep and, if available, a number of surrounding walls. The castle's existence is solely dependent on that of the keep – it is destroyed if and only if the keep is.[199]

198. More accurately, I want to say that we have no reason to accept that any material objects have timeless parts, whilst denying that entities with temporary immediate parts can have timeless parts.

199. This is, of course, merely stipulative. I do not think that it is any harder to accept than Fine's claim that a ham sandwich has timeless parts, and I shall give other examples of nucleated wholes at the end of this section. However, it is perhaps worth noting that castles exist for which this stipulation is simply untrue – Bodiam Castle, for example, has no keep. I hope, though, that it is plausible that some other castles do have their keeps as timeless parts – notably, Restormel castle consists in a shell keep upon a hill, and pele towers, such as Arnside Tower in Cumbria, might be classified as castles whose central keeps are timeless parts. Hence, although I shall not submit any general claims about castle architecture, I take it that this example has enough traction to not be entirely unrealistic.

Chapter 7: De-Fining Material Things

At first sight, a castle's keep seems to be a good example of a timeless material part. In some sense, it is inappropriate to ask when, or for how long, the keep was part of the castle; to ask such a question would simply be to ask when, or for how long, the castle itself existed. If so, then Fine's theory would state that a castle is connected to its keep via a series of operations of rigid embodiment. Castles, however, can change in size; they can grow and shrink. Thus, imagine the following scenario:

> We live in a castle on a hill. Initially, it has no surrounding structures. Owing to repeated barbarian attacks, we hurriedly construct a number of walls around it at time t_1. Eventually, the barbarians are driven from our lands. The walls fall into disrepair, and we demolish them entirely at t_2.

I think that it is natural to say that the walls are part of the castle between t_1 and t_2, but not before t_1 or after t_2. It is also natural to think that they are immediate parts of the castle; they are part of no proper part of the castle (except themselves). Hence, at first sight, we would want to say that the castle has an immediate timeless part – its keep – and an immediate temporary part – its walls. These verdicts, however, are incompatible with the claim that the castle is a rigid embodiment. Not only do rigid embodiments not have immediate temporary parts, they also exist only when all of their immediate parts exist and are appropriately arranged. However, the walls do not exist before t_1 or after t_2. Hence, if the castle were a rigid embodiment, it would only exist between t_1 and t_2. This is surely an unsatisfactory conclusion. Nor, though, can we say that the castle is a variable embodiment on Fine's system; this is inconsistent with the claim that it has its keep as a timeless part.

Fine's theory cannot easily account for objects, such as our castle, which have both immediate timeless and immediate temporary material parts. And I do not think that this is an isolated case. As I mentioned, we seem to be objects of this kind. And there seem to be many other cases of nucleated wholes. There is, for example, an established convention of identifying trains by reference to their locomotives, suggesting that a locomotive is a timeless part of its train. Entities of this type may also be very small; we might, for example, think that a helium atom can persist through ionisation despite losing its orbiting electrons, but not through the loss, or disintegration, of its nucleus. Examples of nucleated wholes thus stretch beyond the artifactual, or the customary, through biological systems, and even down to some of the entities recognised by chemistry. I want now to consider four ways in which the sting of this worry might be lessened.

The first two of these allow us to think of the castle as a variable embodiment while the third and the fourth represent different ways to preserve the claim that the castle is a rigid embodiment. I shall argue that we ought not to accept any of these responses.

7.4 Responses

Put in brief, the problem for Fine's theory is that if an object has timeless material parts, it must be a rigid embodiment. However, if it is a rigid embodiment, all of its immediate parts are timeless, and so it cannot change its immediate parts without ceasing to exist. Nucleated wholes, such as our castle, defy this pattern by having timeless material parts and yet losing and gaining temporary immediate parts. Thus, Fine's theory both requires and precludes that they are rigid embodiments. It will be helpful to organise responses to this argument around the following series of inconsistent claims:

1. The keep is a timeless (material) part of the castle.

2. Only rigid embodiments have timeless (material) parts.

3. Therefore, the castle is a rigid embodiment (from 1, 2).

4. The castle's walls are immediate parts of the castle.

5. A rigid embodiment only exists when all of its immediate parts exist.

6. Therefore, the walls must exist whenever the castle exists (from 3, 4, 5).

7. The castle's walls do not exist whenever the castle exists.

One could, of course, reject claim 4 by saying that the walls are never part of the castle, or claim 7 by saying that the castle only exists when it has walls (between t_1 and t_2 of our example). Both of these responses, however, are incredibly revisionary, and I do not think that the desire to preserve Fine's system unchanged gives us reason to accept them. Hence, I will set them aside in this context. Instead, I wish to explain four other ways to respond, each of which rejects exactly one of 1, 2, 4 and 7. I shall raise objections to each response, suggesting that we ought not to accept Fine's theory as it stands.

Chapter 7: De-Fining Material Things

7.4.1 Semipiternal Parthood

The first option I want to consider rejects the first claim, denying that the keep is a timeless part of the castle. Instead, it states that the castle is a variable embodiment whose manifestations all have the keep as a part. On this view, the keep is a 'semipiternal' part of the castle – a temporary part of the castle at every time at which it exists.[200] In this section, I wish to argue that we should not rescue Fine's theory by recourse to the notion of semipiternal parthood.

My worry is that acceptance of semipiternal parthood endangers the claim that material objects have timeless parts – something that Fine takes as a datum in developing his theory. To see this, first note that the keep can be contrasted with anything that just happens to be part of the castle throughout its existence – for example, a wooden door or an election. Such objects are only accidentally part of the castle whenever it exists. There is no temptation to count them as among its timeless parts. The keep, however, is metaphysically guaranteed to be part of the castle whenever it exists. This shows that we need some way to mark the distinction between those objects which are metaphysically guaranteed to be part of the castle whenever it exists and those which are not. I suggest that we should connect this distinction to that between timeless and temporary parthood. Of course, there is logical space to deny that we should do so; I accept that it is not immediately evident that one object is a temporary part of another only if it could cease to be part of that object without that object ceasing to exist. However, if we reject this claim, I struggle to see why we should accept that any of a material object's parts are timeless, rather than merely semipiternal.

Take Fine's example of a ham sandwich. According to Fine, it is inappropriate to ask when the ham became part of the ham sandwich (Fine 1999, p. 61). On the face of it, this seems a rather strange thing to say. The ham surely predated the sandwich's construction (and, if the sandwich were disassembled, could outlast it). If so, then there were times at which the ham existed without being part of the sandwich. Given this, it seems that it should make sense to ask when those times were, and hence to ask when (and for how long) the ham was part of the sandwich. What this demonstrates is that the notion of inappropriateness in question is not intended to capture a natural linguistic marker of the presence of timeless parthood. Further explanation is therefore needed if we

---

200. I mean the distinction between semipiternal and timeless parthood to mirror that drawn in Fine's 2005 between semipiternal and eternal truth.

are to accept that time-bound material objects have timeless parts; it is not just a datum that they have such parts.[201]

If the response under consideration is to succeed, a successful explanation of the circumstances under which a material object has timeless parts must distinguish between the ham sandwich and the castle, explaining why the former has timeless parts and the latter does not. To my eyes, however, the core feature of both cases is that they involve objects that are identity dependent upon some of their parts. Further, this feature seems to me to provide the best interpretation of what is meant by saying that it is 'inappropriate to ask' when some object has something as a part; if it is inappropriate to ask when the sandwich gained its parts, this looks to be because the sandwich could not exist without those parts. Unfortunately, however, this explanation applies equally to the castle, which depends upon its keep for its existence. Unless we can locate some other (plausible) differentiating factor, I would suggest that the cases should be described in the same way; the ham should be a sempiternal part of the ham sandwich, just as much as is the keep.[202] Given this, Fine's examples of timeless parthood, and with them the theory of rigid embodiment, are undermined by the response from sempiternal parthood.

There is an easy, and obvious, response to this challenge; if an object is a rigid embodiment, then its immediate parts are timeless, if it is a variable embodiment, its material parts are temporary. Hence, the difference between the castle and the ham sandwich results from their different generating operations. This response, however, is ineffective. Fine develops his theory of embodiments in response to the intuitive distinction between timeless and temporary part-

---

201. There is one possible exception; Fine presents the time-slices of a persisting object as an example of timeless parthood (Fine 1999, p. 61). I agree that these do appear to be part of their object in a different way than the ordinary 'spatial parts' of a persisting object. However, I do not quite know how to understand this example in the context of the theory of embodiments. Presumably, if some persisting objects have time-slices, then all do, rigid and variable embodiments alike. If so, then either the manifestations of a variable embodiment are both timeless and temporary parts of the embodiment, or a variable embodiment's time-slices are not identical to that variable embodiment's manifestations. But, then, this seems to suggest that two things account for an object's mereological structure at a time – its manifestation at that time, and the time-slice that it has at that time. I conclude that it is far from clear how this example is to be integrated with the theoretical apparatus of the theory of embodiments.

202. Koslicki makes a similar point more forcefully, concluding "even if there turn out to be [rigid embodiments], it is not clear that their analysis requires the introduction of an additional timeless notion of parthood into the domain of material objects" (Koslicki 2007, p. 157). My point is that if some objects are mereologically inflexible with respect to their temporary parts, there is even less reason to think that any material objects have timeless parts.

hood. It presupposes that there is such a distinction, and that it is clear to the reader that it applies to some material objects. I have suggested that we cannot draw this distinction if we accept that the keep is a semipiternal part of the castle. Once we accept this, we have little reason to think that any material objects have timeless, rather than semipiternal, parts. As such, it would be inappropriate to distinguish timeless and semipiternal parthood by reference to the theory of embodiments.

I cannot see an easy response to the challenges raised above, and detailed discussion of the options – claims about essences and the like – would take us far afield. Let us therefore accept that the keep is a timeless part of the castle, and instead consider another response.

7.4.2 Complex Principles of Embodiment

The second response maintains that variable embodiments might have timeless material parts, thereby rejecting the second premise of the above argument. The key idea is that a variable embodiment does already have a timeless part – its principle of embodiment.[203] This response claims that principles of embodiment can themselves have material objects as timeless parts. Assuming that timeless parthood is transitive, this response allows us to maintain that the keep is a timeless part of the castle without taking the castle to be a rigid embodiment.[204] This is a very interesting proposal. However, I am unsure whether we should accept it.

This proposal has two versions. Before considering them, I want briefly to note a feature common to both. It is significant that a principle of embodiment is, according to Fine, "not another physical object but something of a more abstract or conceptual nature" (Fine 1999, p. 69). This response therefore asks us to accept that abstract or conceptual entities can have physical parts. Although this result is in keeping with Fine's claim that "the divide between the concrete and abstract realms is not as great as it is commonly taken to be" (ibid., p. 73), it is certainly strange. Indeed, this faces us with a number of questions. Must we, for example, hold that a principle of embodiment is located where its physical parts

---

203. Fine does not explicitly say that a variable embodiment has its principle as a timeless part. However, given that rigid embodiments have their principles as timeless parts, it seems reasonable to assume that variable embodiments do also. See here footnote 21 of Koslicki 2007 (p. 146).

204. Although Fine does not explicitly commit to the transitivity of timeless parthood in his 1999, Koslicki indicates in footnote 20 of her 2007 (p.145) that Fine endorses this principle.

are? And, if so, when? Does it only have a physical location when the relevant variable embodiment exists, or does it have a physical location whenever its timeless part exists? Of course, these questions can be answered without great difficulty; they just require us to take a stand. However, the very fact that they arise should make us think twice about accepting the response at hand.

Now to develop the proposal. It has two variants, which differ as to whether one object can be both a timeless and a temporary part of another. First, let us say that this is impossible; an object can be either a temporary or a timeless part of a variable embodiment, but never both. This requires a slight modification to the theory of embodiments. At present, this theory postulates that a variable embodiment exactly shares its location with its manifestation. If, however, the keep is a timeless part of the castle, then we would wish to hold that the castle is (partly) located where its keep is. Yet we cannot say that the keep is part of the castle's manifestation (else, the keep would be both a timeless and a temporary part of the castle). Thus, we must propose that a variable embodiment is located at any point p if and only if either its manifestation, or a material part of its principle of embodiment is located at p.

It is worth noting how this proposal differs from the original details of Fine's theory. According to the original theory, a variable embodiment's manifestation entirely represented its physical presence. Now, however, the manifestation only contains the object's temporary parts (and their parts). The castle is therefore made up of one changeable, and one unchangeable element, both of which have material aspects. This constitutes a notable departure from more traditional hylomorphic accounts of material beings; such accounts ask us to distinguish a material object's physical aspect – its matter – from its conceptual component, or form. Further, I am suspicious of the claim that there is some part of the castle which comprises just its changeable aspects. This commitment is shared by the final response considered here – the response from intermediary parts. Setting this suspicion aside, I find the response from intermediary parts a little more natural than that considered here. Hence, I would submit that that view is preferable to this first way of developing the thought that material objects might have complex principles of embodiment.

Alternatively, one could hold that something can be both a timeless and a temporary part of a variable embodiment. Thus, the keep might be both part of the castle's principle of embodiment and part of its manifestation. This would allow us to preserve the theory of embodiments unchanged. It would also avoid

the need to postulate that the castle has a part which comprises all (and only) its changeable parts. I can think of no conclusive objection to this suggestion. It does, however, strike me as odd to claim that one must incorporate the keep twice into our account of what the castle is – once as part of its manifestation, and once as part of its principle of embodiment.[205] My preference would be for a theory of nucleated wholes which did not require this type of logical manoeuvre.

I have been unable to find any decisive reason to reject the response from complex principles of embodiment. I hope, however, to have suggested some considerations that mitigate against it. This proposal comes in two varieties, differing on whether the timeless parts of a nucleated whole are part of its principle alone or both part of its principle and part of its manifestation. Both varieties require one to accept that abstract objects can have material parts. Further, each has additional odd features. If one thinks that nothing can be both a timeless and a temporary part of another object, one must then think that there exists an object which comprises just the changeable parts of the castle. If, on the other hand, one rejects this claim, one must think that an object can feature twice in an account of another object's composition. Hence, although I do think that much remains to be explored here, my suggestion is that we look for a less exotic response.

7.4.3   Null Embodiments

I now want to consider two ways in which we might defend the claim that the castle is a rigid embodiment. The first response holds that the castle has two immediate timeless parts – its walls and its keep.[206] Both of these parts are variable embodiments, and, despite appearances, exist whenever the castle does. Thus, the response rejects the seventh claim of the above argument, holding that the castle's walls exist whenever it does. In order to hold that the walls exist at the

---

205. If one does not find this consequence unappealing, this opens up one more response – left undiscussed here for reason of space. This is the claim that the castle is a rigid embodiment which consists of the keep standing in the relation of 'being the core of' a variable embodiment which possesses all of the castle's parts whenever it exists. This object has the keep as a timeless part, and the walls as temporary parts. Additionally, both the keep and the walls can enter as immediate parts of the manifestations of the variable embodiment, accounting for our intuition that the walls are immediate parts of the castle. This is a promising solution. However, it strikes me as a little profligate to hold that three embodiments – a manifestation, a variable embodiment, and the castle itself – are required to account for the castle's compositional structure.

206. When I presented this chapter, this was Fine's preferred response. I hope here to suggest that we ought not to accept it.

beginning and the end of our example, this response relies upon the idea of a 'null' embodiment – a rigid embodiment which has no material component.[207] It states that the walls are manifested by such null embodiments before t_1 and after t_2. They therefore exist (and are part of the castle) throughout the castle's existence, but do not always have material parts. Although this response avoids the undesirable consequence that the castle ceases to exist when its walls are torn down, it is unsatisfactory for a number of reasons.

First of all, note that a null embodiment has no material components. As such, it has no location. A variable embodiment borrows its material parts, and its location, from its manifestation. Hence, this response asks us to believe that material objects, such as the walls, can exist without being located anywhere, or having any material parts. This is extremely strange. Even if we were willing to accept the existence of null embodiments, I am unsure whether we would want to accept these consequences.

The above is not the only reason to shy away from this response. First of all, note that the castle of our original example comes into existence without walls (or, more neutrally, without any physical parts which are its walls), and only gains walls in response to a later threat. Previous to this threat, no architect envisaged that this castle should have walls. Fine's theory thus asks us to state that the walls exist, and are part of the castle, long before any circumstance arises which could make sense of their coming to exist (or becoming part of the castle). Call this the 'problem of origins'. Two worries should ramp up the force of this problem.

First, take what I shall call the 'proliferation worry'. Having dismantled the walls, there is no reason why we should not later build walls in a different location. These walls are presumably not identical to the original set of walls. But if these are a new set of walls, and are also timeless immediate parts of the castle, then the castle always had two sets of walls as immediate timeless parts. And, if the castle were to exist for a long time, and undergo extensive modification throughout, then the number of its immediate parts could be even greater, pro-

---

207. I speak here of 'a' null embodiment, rather than of 'the' null embodiment because the fact that null embodiments do not have material parts (and so cannot be distinguished by reference to them) does not demonstrate that they could not have different principles of embodiment. Hence, in lieu of a discussion of the individuation of 0-place predicates, I leave it open that there may exist more than one rigid embodiment which lacks any material parts. A supplementary advantage of this route is that it does not commit one to saying that any two objects which are both manifested by null embodiments have a part (the null embodiment) in common.

liferating beyond even the wildest imaginations of any of its inhabitants. This is surely a consequence to be avoided.

Secondly, consider an additional 'novelty worry'. Imagine that the castle gains the first example of a novel kind of structure at time t_3 – an observatory. Imagine further that the technology for constructing observatories did not exist prior to t_3, and that no-one had even thought of such a thing prior to this moment in time. If the observatory is a timeless part of the castle, then we must say that the first observatory predated the invention of the observatory. This seems almost contradictory. Nor would it be desirable to say that the fact of novelty bars observatories from becoming part of the castle; this would be to unduly limit the alterations that the castle could undergo. I do not think this a happy result.

These worries add weight to the problem of origins; they reveal why it is unsatisfactory to say that events which occur late in the castle's career dictate what parts it has throughout. We do not ordinarily think that events have this kind of power over the past. And I hope to have shown that to think that they do would require us to say some rather strange things. I shall conclude our discussion of null embodiments with one more objection. Although this objection, 'the problem of the relation', is weaker than the problems raised above, it will further highlight the problematic nature of null embodiments.

A rigid embodiment, Fine says, exists whenever all of its immediate parts exist, and stand in the relation specified by its principle of rigid embodiment. If we say that the castle is a rigid embodiment, we therefore must specify a relation in which its walls stand to its keep both when they have physical manifestations, and when they are manifested instead by a null embodiment. It is, however, difficult to think of what this relation could be. Obviously, it cannot be something that requires that the walls have physical manifestations (such as the relationship of encirclement or of defending). Nor though can it be 'was designed for', or 'was made to accompany' since the walls were neither designed nor made to accompany the castle's keep until long after the castle's creation. And it would be flatly circular to say that our castle is generated from the operation of rigid embodiment upon the keep, the walls, and the operation of 'composing the same castle'. Hence, the introduction of null embodiments has a subsidiary cost; it becomes much harder to determine just what the castle's principle of embodiment is.

The idea of a null embodiment is far from intuitive. Further, their introduc-

tion is not without cost; this approach asks us to accept an unwieldy picture of our castle's compositional structure, and of the relation which is its principle of embodiment. On balance then, I conclude that we ought to reject this response.

7.4.4   Intermediary Parts

There is one more way in which we might defend the claim that the castle is a rigid embodiment. This involves relatively minor modification of what we take its compositional structure to be; instead of saying that the walls are immediate parts of the castle, we hold that they are part of some other variable embodiment which is an immediate part of the castle (and which exists whenever the castle does). This is to reject the fourth premise listed above. As a variable embodiment has temporary parts, the walls need not exist whenever this intermediary part exists. And, given this, we might say that the castle exists even at times when the walls do not. Of the responses considered in this section, this is the one I prefer. Nevertheless, it is not without problems of its own. I will present these difficulties, and conclude that they should motivate us to search for an alternative response.

To accommodate examples like that of the castle, we just need to find a part of the castle which exists whenever the castle does and of which it is plausible to say that the walls are temporarily part. For an example of such an object, we might appeal to the idea of a castle's 'grounds'—the area of land which surrounds it, and (at least some of) the structures which are found there. Let us say that the castle's grounds are a variable embodiment which has the surrounding land, and the structures upon it, as temporary parts. If we think that the grounds are part of the castle from the moment that it exists, and that the walls are part of the castle's grounds when they exist, then nothing stands in the way of taking the castle to be a rigid embodiment; we say that it consists in the grounds and the keep standing in an appropriate relation. And, as the walls are only temporary parts of the grounds, we avoid the worry that the castle does not exist when its walls do not.

While this response is promising, I do not think that it is entirely satisfactory. One way to get at the problem with this response takes the form of a dilemma. There are, I think, two distinct notions of a castle's grounds to which one might appeal. The first, fairly ordinary sense of the grounds takes these to be an area of land—the castle's gardens—and at most some of the structures built upon it (dividing walls for example). Using such a notion, it seems unlikely that all

of the structures which we would ordinarily take to be part of the castle will be parts of the grounds or of the keep; the castle's barracks, for example, may lie within the castle's grounds, but I am unsure whether we would ordinarily think of them as part of the castle's grounds. It may therefore be necessary to introduce supplementary objects (e.g. the castle's outbuildings) if we are to capture the compositional structure of the castle. And, it strikes me that we need not think that these supplementary objects are guaranteed to exist whenever the castle exists. Hence, I am unsure that this interpretation of the castle's grounds can, in the end, satisfactorily deal with cases such as the above without recourse to another suggestion, such as the appeal to null embodiments considered immediately above.

The second horn of the dilemma uses a slightly different interpretation of the castle's grounds – one more appropriate for a surveyor's report than for everyday speech. On this interpretation, the castle's grounds comprise the parts of the castle other than its keep – the parts which would 'come with it' were it sold. It strikes me that this notion of the castle's grounds is inappropriate for this context. When using this notion, we would not identify the castle's grounds independently of the castle; asked what they were, we would naturally begin with the castle, and then talk about which of its parts are to count as its grounds (or which of the castle's parts are not parts of its grounds). We would say, for example, that the grounds are the castle minus the keep, or the land belonging to the castle and the things built upon that land. We would, in other words, define the grounds by reference to the castle, and not the other way around. However, this suggests that the castle's grounds depend upon the castle for their existence, rather than the castle upon its grounds. If this is the case, then it would be odd to think that the castle is built from its grounds by a Finean operation of composition; this would require us to think that the grounds are ontologically prior to the castle whilst also dependent upon it for their existence (and their compositional unity). This is an unintuitive result, and one that I think should lead us to seek out an alternative.

So, it seems that the proponent of this response has no satisfactory way to understand the castle's grounds; either her account of the grounds does not guarantee that all of its temporary parts will be part of its grounds, or it generates the grounds by reference to the castle itself. Leaving the example of the castle aside, it is also worth noting that it may often be difficult to find such a natural intermediary object as the castle's grounds to rescue our claim that some entity

has timeless parts. When an object's temporary parts are highly heterogeneous (particularly when they are not spatially contiguous), we are unlikely to recognise any entity of which they are all parts. And, even if such an entity existed, I suspect that we would not wish to give it a place in an object's compositional structure simply in order to explain how that object could have timeless parts. Hence, although some cases can be dealt with by this type of move, I think that it does not generalise to all cases in which an object seems to have both timeless and temporary immediate parts.

The addition of intermediary parts to our account of an object's compositional structure is therefore unsatisfactory. This concludes our survey of responses on Fine's behalf. None of the responses we have considered are entirely satisfactory. I therefore submit that Fine's system is unable, as it stands, to account for nucleated wholes such as ourselves.

7.5 Conclusion

I have above introduced the notion of a nucleated whole – an object with both immediate timeless and immediate temporary parts. Such objects cannot straightforwardly be constructed by the Finean operations of rigid and variable embodiment. I have considered four ways in which one might try to accommodate nucleated wholes within the theory of embodiment, suggesting that each is unsatisfactory. Thus, we should conclude that Fine's theory of embodiments is unable to account for nucleated wholes.

Considered as a whole, this chapter does two things. First, it shows that we are not alone in being individuated by reference to a central part; there exist other examples of nucleated wholes. Second, it raises the general question of what resources we should use in providing an account of the nature of material objects. The variegated way in which nucleated wholes relate to their immediate parts requires explanation from any successful theory of the nature of material things. Such objects therefore provide an ideal testing ground for theories of composition.

Concluding Remarks

This thesis has developed a view according to which we have humanoid form, but are nevertheless capable of coming to be solely composed by the realisers of our mental states. This view is attractive. Nevertheless, I have not argued directly for it. I have simply argued that anyone attracted to the psychological view should also be attracted to it. Its merits have yet to be compared with views on which psychological connections are irrelevant to our continued existence. I want to step beyond those findings with two comments. First, this view has, at one time or other, appealed to other philosophers – some of whom have suggested arguments in its favour. Second, in further developing this view, it will have to come into contact with the view that we are animals.

First, though, I should summarise my findings. We began by evaluating the wide psychological view, on which no constraints are placed on the way in which psychological connections are realised between the different phases of a person's existence. This view says that we can be transferred from one place to another by exotic processes such as Teletransportation. It is easy to find this view attractive if one offhandedly reflects upon the scenarios presented by science fiction; it is easy to imagine the protagonist of one's story being beamed from one place to another to continue her adventures. However, upon close philosophical reflection, the view turns out to be hard to maintain.

I have investigated two attempts to motivate the wide psychological view. First, I considered the claim that the wide psychological view is supported by our intuitions. Unfortunately, though we certainly do have some intuitions that seem to count in its favour, we have others that seem to count against it. I concluded that we do not have a particularly robust set of beliefs about our persistence, by which to justify the wide psychological view.

I subsequently considered whether the wide psychological view is supported by functionalist accounts of mentality. If this claim were correct, then it would show that the wide psychological view connects our persistence to a

set of very significant causal interactions. Under these circumstances, the simplicity and naturalness of the view would strongly recommend that we accept it. Unfortunately, this presentation of the wide psychological view neglects the issue of branching psychological continuity. If it is possible to travel by teletransportation, all kinds of scenarios are possible in which one's psychological life branches. I argued that the best response is to adopt constraints upon the realisation of psychological connections between different phases of one's life. Though this does not entirely alleviate the problems that might be presented by instances of branching, it is nevertheless likely to limit them to cases in which we would be rightly unsure what to say. I concluded my argument against the wide psychological view by detailing why the absence of good motivation for the wide psychological view should count against it. In short, the metaphysically basic case of personal persistence requires the continuous realisation of one's psychological states. We cannot find reason to adopt more liberal criteria of personal identity than are required to account for the basic case. Thus, we should reject the wide psychological view.

The second half of the thesis considered what parts we have. I began by showing that facts about the way in which our mental states are realised in the brain do not count against the claim that we are material. I continued to show that these findings also need not unseat the claim that we are identical to a 'minimal' subject of our mental states – something which contains the realisers of our mental states, and as few parts as possible besides. Nevertheless, I argued that this view is at odds with the fact that we individuate many of our mental states by reference to the ways in which they are housed within larger entities. I concluded that it is therefore most reasonable to take us to be identical to these larger entities.

I demonstrated that the view which emerged – on which we are humanoid, but have a central individuating part – is incompatible with a picture of the metaphysics of material objects developed by Kit Fine. Since we are not the only example of objects of this kind, I conclude that we should reject Fine's theory of embodiments. With this said, I want to consider where we should turn next.

First, I should like to note that the views developed here are not original. At some time or other, a number of thinkers have believed similar things.[208] I want

---

208. C.f. Unger 1990; Agar 2003; Shoemaker 2004b; Shoemaker 2016; Wiggins 1967; Johnston

to highlight one remark as particularly interesting. Wiggins writes:

> The kind of individual we are to define is not made of anything other than flesh and bones, but, unlike the body with which it at some times shares its matter it has a characterization in functional terms which confer the rôle, as it were, of *individuating nucleus* on a particular brain which is the seat of a particular set of memory-capacities. (Wiggins 1967, p. 51)

Though Wiggins later rejects this account, it is very similar to the position that I have supported. I select this quotation for two reasons. First, Wiggins develops an account of our identity in conjunction with a neo-Aristotelian theory of the nature of objects.[209] We have seen how one, highly abstract, framework for dealing with the metaphysics of objects is unable to validate the existence of objects like us. One wonders what the upshot will be of pursuing and clarifying the idea of an entity's 'functional characterisation' in an account of the mereological structure of objects.[210]

The second reason I choose this quotation is that at some point, the account of our nature developed above will have to enter into conversation with another way of seeing us as functionally organised wholes – the thought that we are living organisms. I have stopped short of engaging in this conversation in this thesis, though what I have said strongly suggests that do we not yet have strong intuitions about the metaphysics of organisms, or the metaphysics of persons by which to find our way. Despite this, I think that we should be very impressed by the fact that, as Snowdon puts it "our basic, core, self-conception is [extremely] *animal-like*" (Snowdon 2014b, p. 85). Thus, if we are to develop this view further, it should be investigated whether, and how, it can be made to link up with an account of the nature and persistence of human organisms. Though this question has been asked before,[211] I think that there remains important work to be done before we will be in a position to endorse a definite account of our nature.

1987a; Johnston 2007; Johnston 2016.

209. See also (1980; 2001; 2012) for further developments on this framework.

210. See here Jones 2015 and Madden 2015b for attempts to draw out the ontological ramifications of adopting a framework of this kind. See also Madden 2016a for an attempt to bring this to bear upon questions of our persistence.

211. See Olson 1995; Olson 1997; Noonan 1998; Wiggins 1976; Wiggins 1987; Madden 2016a; Bailey 2015; Thornton 2016; Toner 2011.

Bibliography

Agar, N. 2003. 'Functionalism and Personal Identity'. *Noûs* 37: 52–70.

Anscombe, G. E. M. 1975. 'The First Person'. In S. Guttenplan, ed. *Mind and Language*. Oxford: Clarendon Press.

Aquinas. 1265–1274/2017. *Summa Theologiae*. http://www.newadvent.org/summa/index.html.

Árnadóttir, S. T. 2010. 'Functionalism and thinking animals'. *Philosophical Studies* 147: 347–354.

Ayer, A. J. 1963. 'The Concept of a Person'. In. *The Concept of a Person and other essays*. New York: St. Martin's Press Inc.

Ayers, M. 1974. 'Individuals without Sortals'. *Canadian Journal of Philosophy* 4: 113–148.

— 1991a. *Locke: Epistemology and Ontology*. London: Routledge.

— 1991b. 'Substance: Prolegomena to a Realist Theory of Identity'. *The Journal of Philosophy* 88: 69–90.

— 1997. 'Is Physical Object a Sortal Concept? A Reply to Xu'. *Mind & Language* 12: 393–405.

— 2005. 'Ordinary Objects, Ordinary Language, and Identity'. *The Monist* 88: 534–570.

Bailey, A. M. 2015. 'Animalism'. *Philosophy Compass* 10: 867–883.

— 2016. 'You Are an Animal'. *Res Philosophica* 93: 205–218.

Baker, L. R. 1997. 'Why Constitution Is Not Identity'. *The Journal of Philosophy* 94: 599–621.

— 2000. *Persons and Bodies*. Cambridge: Cambridge University Press.

— 2016. 'Animalism vs. Constitutionalism'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Barnes, E. 2010. 'Arguments Against Metaphysical Indeterminacy and Vagueness'. *Philosophy Compass* 5: 953–964.

Bibliography

Bennett, M. R. and Hacker, P. M. S. 2003. *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Berniūnas, R. and Dranseika, V. 2016. 'Folk concepts of person *and* identity: A response to Nichols and Bruno'. *Philosophical Psychology* 29: 96–122.

Blatti, S. 2016. 'Headhunters'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Block, N. 1978. 'Troubles with Functionalism'. *Minnesota Studies in the Philosophy of Science* 9: 261–325.

Bloom, P. 2004. *Descartes' Baby*. New York: Basic Books.

Bourget, D. and Chalmers, D. J. 2014. 'What do philosophers believe?' *Philosophical Studies* 170: 465–500.

Brewer, B. 2011. *Perception and Its Objects*. Oxford: Oxford University Press.

— 2015. 'Exclusive Individuals'. *Aristotelian Society Supplementary Volume* 89: 125–142.

— 2019. 'Basic Objects as Grounds: A Metaphysical Manifesto'. In J. Cumpa and B. Brewer, eds. *The Nature of Ordinary Objects*. Cambridge: Cambridge University Press.

Broad, C. D. 1925. *The Mind and Its Place in Nature*. New York: Harcourt, Brace and Co.

Brueckner, A. and Buford, C. T. 2009. 'Thinking Animals and Epistemology'. *The Philosophical Quarterly* 90: 310–314.

Burke, M. B. 1992. 'Copper Statues and Pieces of Copper: A Challenge to the Standard Account'. *Analysis* 52: 12–17.

— 2003. 'Is My Head a Person?': in. *On Human Persons*. Frankfurt: Ontos verlag.

— 2004. 'Dion, Theon, and the Many-Thinkers Problem'. *Analysis* 64: 242–250.

Butler, J. 1736. *The Analogy of Religion*. London: J. M. Dent and Sons Ltd.

Campbell, K. 1970. *Body and Mind*. London and Basingstoke: The MacMillan Press Ltd.

Campbell, S. 2005. 'Is Causation Necessary for What Matters in Surivval'. *Philosophical Studies* 126: 375–396.

Carnap, R. 1928/2003. *The Logical Structure of the World*. Chicago and La Salle, Illinois: Open Court.

Carter, W. R. 1989. 'How to Change Your Mind'. *Canadian Journal of Philosophy* 19: 1–14.

Cassam, Q. 1997. *Self and World*. Oxford: Oxford University Press.

Bibliography

Chisholm, R. M. 1948. 'The Problem of Empiricism'. *The Journal of Philosophy* 45: 412–417.

— 1976. *Person and Object*. London: George Allen & Unwin Ltd.

— 1978. 'Is there a Mind-Body Problem?' *Philosophic Exchange* 9: 25–34.

Clint Dowland, S. 2016. 'Embodied mind sparsism'. *Philosophical Studies* 173: 1853–1872.

Craver, C. F. and Bechtel, W. 2007. 'Top-down Cuasation Without Top-down Causes'. *Biology and Philosophy* 22: 547–563.

Dainton, B. 2004. 'The self and the phenomenal'. *Ratio* 17: 365–389.

— 2008. *The Phenomenal Self*. Oxford: Oxford University Press.

— 2012. 'Self-hood and the Flow of Experience'. *Grazer Philosophische Sudien* 84: 161–200.

— 2015. 'From Phenomenal Selves to Hyperselves'. *Royal Institute of Philosophy Supplement* 76: 161–197.

— 2016. 'The Sense of Self'. *Aristotelian Society Supplementary Volume* 90: 113–143.

Dainton, B. and Baine, T. 2005. 'Consciousness as a guide to personal persistence'. *Australasian Journal of Philosophy* 83: 549–571.

Davis, L. H. 1998. 'Functionalism and Personal Identity'. *Philosophy and Phenomenological Research* 58: 781–804.

— 2001. 'Functionalism, the Brain, and Personal Identity'. *Philosophical Studies* 102: 259–279.

Dennett, D. 2007. 'Philosophy as Naive Anthropology: Comment on Bennett and Hacker'. In M. Bennett et al., eds. *Neuroscience and Philosophy: Brain, Mind, and Language*. New York: Columbia University Press.

deRosset, L. 2011. 'What is the Grounding Problem?' *Philosophical Studies* 156: 173–197.

Descartes, R. 1641/1998. *Meditations on First Philosophy*. Indianapolis: Hackett Publishing Company, Inc.

Dickie, I. 2014. 'The Sortal Dependence of Demonstrative Reference'. *European Journal of Philosophy* 22: 34–60.

Dorr, C. and Hawthorne, J. 2013. 'Naturalness'. In D. W. Zimmerman, ed. *Oxford Studies in Metaphysics 8*. Oxford: Oxford University Press.

Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.

Elliot, R. 1991. 'Personal Identity and the Causal Continuity Requirement'. *The Philosophical Quarterly* 41: 55–75.

Bibliography

Evans, G. 1978. 'Can there be vague objects'. *Analysis* 38: 208.

— 1982. *The Varieties of Reference*. Ed. by J. McDowell. Oxford: Clarendon Press.

Fine, K. 1975. 'Vagueness, Truth, and Logic'. *Synthese* 30: 265–300.

— 1982. 'Acts, Events and Things'. In W. Leinfellner, E. Kraemer, and J. Schank, eds. *Language and Ontology*. Vienna: Holder-Pichler-Tempsky.

— 1999. 'Things and Their Parts'. *Midwest Studies in Philosophy* 23: 61–74.

— 2005. 'Necessity and Non-Existence'. In. *Modality and Tense: Philosophical Papers*. Oxford: Oxford University Press.

— 2007. 'Response to Kathrin Koslicki'. *dialectica* 61: 161–166.

— 2016. 'Identity criteria and ground'. *Philosophical Studies* 173: 1–19.

— Forthcoming. 'Acts and Embodiment'. In A. Geddes and A. Hinshelwood, eds. *Beings and Doings*. Oxford: Oxford University Press.

Foster, J. 1979. 'In *Self*-Defence'. In G. F. MacDonald, ed. *Perception and Identity*. Ithaca, New York: Cornell University Press.

— 1991. *The Immaterial Self*. London and New York: Routledge.

Fuller, G. 1992. 'Functionalism and Personal Identity'. *The Personalist Forum* 8: 133–143.

Gallois, A. 1998. *Occasions of Identity*. Oxford: Oxford University Press.

Garrett, B. 1990. 'Personal Identity and Extrinsicness'. *Philosophical Studies* 59: 177–194.

Gasser, G. and Stefan, M., eds. 2012. *Personal Identity: Complex or Simple?* Cambridge University Press.

Geddes, A. 2016. '*Animalism and the Human Perspective*'. PhD thesis. UCL.

Genone, J. 2016. 'Recent Work on Naïve Realism'. *American Philosophical Quarterly* 53: 1–26.

Gillett, C. 2014. 'Brains, Neuroscience, and Animalism: On the Implications of Thinking Brains'. *The Southern Journal of Philosophy* 52: 41–52.

Grice, H. P. 1941. 'Personal Identity'. *Mind* 50: 330–350.

Gustafsson, J. E. 2019. 'Non-branching personal persistence'. *Philosophical Studies* 176: 2307–2329.

Hawley, K. 2005. 'Fission, Fusion and Intrinsic Facts'. *Philosophy and Phenomenological Research* 71: 602–621.

Heller, M. 1987. 'The best candidate approach to diachronic identity'. *Australasian Journal of Philosophy* 65: 434–451.

Hershenov, D. 2005. 'Do Dead Bodies Pose a Problem for Biological Approaches to Personal Identity?' *Mind* 114: 31–59.

Bibliography

— 2009. 'Organisms and their Bodies: Response to LaPorte'. *Mind* 118: 803–809.

Hershenov, D. B. 2006. 'Shoemaker's Problem of Too Many Thinkers'. *Proceedings of the American Catholic Philosophical Society* 80: 225–236.

Hirsch, E. 1982. *The Concept of Identity*. Oxford: Oxford University Press.

Hudson, H. 2001. *A Materialist Metaphysics of the Human Person*. Ithaca, New York: Cornell University Press.

Hume, D. 1739/1985. *A Treatise of Human Nature*. London: Penguin Books.

Jacinto, B. and Cotnoir, A. J. 2019. 'Models for Hylomorphism'. *Journal of Philosophical Logic* Online Early Access: 1–47.

James, W. 1890. *The Principles of Psychology*. Vol. 1. New York: Dover Publications Inc.

Johansson, J. 2010. 'Parfit on fission'. *Philosophical Studies* 150: 21–35.

Johnston, M. 1984. '*Particulars and Persistence*'. PhD thesis. Princeton University.

— 1987. a. 'Human Beings'. *The Journal of Philosophy* 84: 59–83.

— 1987. b. 'Review of *Personal Identity (Great Debates in Philosophy* by Sydney Shoemaker and Richard Swinburne'. *The Philosophical Review* 96: 123–128.

— 1989. 'Relativisim and the Self'. In M. Krausz, ed. *Relativism: Interpretation and Confrontation*. Notre Dame, Indiana: University of Notre Dame Press.

— 2007. '"Human Beings" Revisited: My Body is Not An Animal'. In D. W. Zimmerman, ed. *Oxford Studies in Metaphysics 3*. Oxford: Oxford University Press.

— 2016. 'Remnant Persons: Animalism's Undoing'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Jones, N. 2015. 'Multiple Constitution'. In K. Bennett and D. W. Zimmerman, eds. *Oxford Studies in Metaphysics 9*. Oxford University Press.

Kant, I. 1781/1998. *The Critique of Pure Reason*. Ed. by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Kanwisher, N. 2010. 'Functional specificity in the human brain: A window into the functional architecture of the mind'. *Proceedings of the Narional Academy of Sciences in the United States of America* 107: 11163–11170.

Kenny, A. 1984. 'The Homunculus Fallacy'. In. *The Legacy of Wittgenstein*. Oxford: Basil Blackwell.

Kolak, D. and Martin, R. 1987. 'Personal Identity and Causality: Becoming Unglued'. *American Philosophical Quarterly* 24: 339–347.

Koslicki, K. 2007. 'Towards a Neo-Aristotelian Mereology'. *dialectica* 61: 127–159.

Bibliography

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

Kovacs, D. M. 2010. 'Is there a Conservative Solution to the Many Thinkers Problem?' *Ratio* 23: 275–290.

Kuhlmeier, V. A., Bloom, P., and Wynn, K. 2004. 'Do 5-month-old infants see humans as material objects?' *Cognition* 94: 95–103.

Langford, S. 2014. 'On What We are and How We Persist'. *Pacific Philosophical Quarterly* 95: 356–371.

LaPorte, J. 2009. 'On Two Reasons for Denying that Bodies Can Outlast Life'. *Mind* 118: 795–801.

Lavazza, A. and Robinson, H., eds. 2014. *Contemporary Dualism: A Defense*. New York: Routledge.

Lewis, C. I. 1946. *An Analysis of Knowledge and Valuation*. Chicago and La Salle, Illinois: Open Court.

Lewis, D. 1972. 'Psychophysical and Theoretical Identifications'. *Australasian Journal of Philosophy* 50: 249–258.

— 1976. 'Survival and identity'. In A. O. Rorty, ed. *the Identities of Persons*. Berkeley and Los Angeles, California: University of California Press.

— 1983. 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* 61: 343–377.

— 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Locke, J. 1689/1975. *An Essay Concerning Human Understanding*. Ed. by P. H. Nidditch. Oxford: Oxford University Press.

Lowe, E. J. 1996. *Subjects of Experience*. Cambridge: Cambridge University Press.

— 2006. 'Non-Cartesian Substance Dualism and the Problem of Mental Causation'. *Erkenntnis* 65: 5–23.

— 2012. 'The probable simplicity of personal identity'. In G. Gasser and M. Stefan, eds. *Personal Identity: Complex or Simple?* Cambridge: Cambridge University Press.

Mackie, D. 1999a. 'Animalism versus Lockeanism: No Contest'. *The Philosophical Quarterly* 49: 369–376.

— 1999. b. 'Personal Identity and Dead People'. *Philosophical Studies* 95.219–242.

Mackie, P. 2008. 'Coincidence and Identity'. *Philosophy* 83: 151–176.

Madden, R. 2011. 'Externalism and Brain Transplants'. In K. Bennett and D. W. Zimmerman, eds. *Oxford Studies in Metaphysics 6*. Oxford: Oxford University Press.

— 2015. a. 'The Naive Topology of the Conscious Subject'. *Noûs* 49: 55–70.

Bibliography

— 2015. b. 'The Place of The Self in Contemporary Metaphysics'. *Royal Institute of Philosophy Supplement* 76: 77–95.

— 2016. a. 'Human Persistence'. *Philosophers' Imprint* 16.

— 2016. b. 'Thinking Parts'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

— 2017. 'Animal Self-Awareness'. *Philosophy, Theory, and Practice in Biology* 9.

Martin, M. G. F. 2002. 'The Transparency of Experience'. *Mind & Language* 17: 376–425.

— 2006. 'On Being Alienated'. In T. S. Gendler and J. Hawthorne, eds. *Perceptual Experience*. Oxford: Oxford University Press.

Martin, R. and Barresi, J. 2006. *The Rise and Fall of Soul and Self*. New York: Columbia University Press.

McDowell, J. 1997. 'Reductionism and the first person'. In J. Dancy, ed. *Reading Parfit*. Oxford: Blackwell.

McMahan, J. 2002. *The Ethics of Killing*. Oxford: Oxford University Press.

McTaggart, J. M. E. 1927. *The Nature of Existence, vol II*. Cambridge: Cambridge University Press.

Merricks, T. 1998a. 'Against the Doctrine of Microphysical Supervenience'. *Mind* 107: 59–71.

— 1998. b. 'There Are No Criteria of Identity Over Time'. *Noûs* 32: 106–124.

— 2001. *Objects and Persons*. Oxford: Clarendon Press.

Moran, A. 2018. 'Kind-Dependent Grounding'. *Analytic Philosophy* 59: 359–390.

Nagel, J. 2013. 'Knowledge as a Mental State'. In T. S. Gendler and J. Hawthorne, eds. *Oxford Studies in Epistemology 4*. Oxford: Oxford University Press.

Nagel, T. 1986. *The View From Nowhere*. Oxford: Oxford University Press.

Nichols, S. and Bruno, M. 2010. 'Intuitions about personal identity: An empirical study'. *Philosophical Psychology* 23.293–312.

Noonan, H. W. 1985a. 'The closest continuer theory of identity'. *Inquiry* 28: 195–229.

— 1985. b. 'The Only X and Y Principle'. *Analysis* 45: 79–83.

— 1985. c. 'Wiggins, artefact identity and 'best candidate' theories'. *Analysis* 45: 4–8.

— 1989. *Personal Identity*. London: Routledge.

— 1998. 'Animalism versus Lockeanism: A Current Controversy'. *The Philosophical Quarterly* 48: 302–318.

Bibliography

Noonan, H. W. 2001. 'Animalism versus Lockeanism: Reply to Mackie'. *The Philosophical Quarterly* 51: 83–90.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Nudds, M. 2009. 'Recent Work in Perception: Naïve Realism and its Opponents'. *Analysis* 66: 334–346.

O'Connell, M. 2017. *To Be a Machine*. London: Granta Publications.

Olson, E. T. 1995. 'Human People or Human Animals?' *Philosophical Studies* 80: 159–181.

— 1997. *The Human Animal*. Oxford: Oxford University Press.

— 2002. 'What Does Functionalism Tell Us about Personal Identity'. *Noûs* 36: 682–698.

— 2003. a. 'An Argument for Animalism'. In R. Martin and J. Barresi, eds. *Personal Identity*. Malden, MA: Blackwell.

— 2003. b. 'Was Jekyll Hyde?' *Philosophy and Phenomenological Research* 66: 328–348.

— 2007. *What Are We?* Oxford: Oxford University Press.

— 2015. a. 'Animalism and the Remnant-Person Problem'. In J. Fonseca and J. Gonçalves, eds. *Philosophical Perspectives on the Self*. Bern, Switzerland: Peter Lang.

— 2015. b. 'On Parfit's View That We Are Not Human Beings'. *Royal Institute of Philosophy Supplement* 76: 39–56.

— 2016. 'The Remnant-Person Problem'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Parfit, D. 1971. 'Personal Identity'. *The Philosophical Review*: 3–27.

— 1984. *Reasons and Persons*. Oxford: Clarendon Press.

— 2012. 'We Are Not Human Beings'. *Philosophy* 87: 5–28.

Parsons, T. 2000. *Indeterminate identity*. Oxford: Oxford University Press.

Paul, L. A. 2010. 'The Puzzles of Material Constitution'. *Philosophy Compass* 5: 579–590.

Penelhum, T. 1970. *Survival and Disembodied Existence*. London: Routledge & Kegan Paul.

Perry, J. 1972. 'Can the Self Divide?' *The Journal of Philosophy* 69: 463–488.

— 2002. 'The Two Faces of Identity'. In. *Identity, Personal Identity, and the Self*. Indianapolis: Hackett Publishing Company, Inc.

Bibliography

Persson, I. 1999. 'Persons and Organisms'. *Dialogue* 38.519–533.

Polger, T. W. and Shapiro, L. A. 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.

Pryor, J. 2000. 'The Skeptic and the Dogmatist'. *Noûs* 34: 517–549.

Putnam, H. 1965. 'Brains and Behavior'. In R. J. Butler, ed. *Analytical Philosophy*. Oxford: Blackwell.

— 1967. 'Psychological Predicates'. In W. H. Capitan and D. D. Merrill, eds. *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

Quine, W. V. 1950. 'Identity, Ostension, and Hypostasis'. *The Journal of Philosophy* 47: 621–633.

Quinton, A. 1962. 'The Soul'. *The Journal of Philosophy* 59: 393–409.

Rea, M. C. 1995. 'The Problem of Material Constitution'. *The Philosophical Review* 104: 525–552.

— ed. 1997. *Material Constitution: A Reader*. Lanham, Maryland: Rowman and Littlefield Publishers Inc.

— 1998. 'Sameness without Identity: An Aristotelian Solution to the Problem of Material Constitution'. *Ratio* 11: 316–328.

— 2000. 'Constitution and Kind Membership'. *Philosophical Studies* 97: 169–193.

Reid, T. 1785/1941. *Essays on the Intellectual Powers of Man*. St. Martin's St, London: MacMillan and Co. Ltd.

Roache, R. 2006. 'A Defence of Quasi-Memory'. *Philosophy* 81: 323–355.

Robinson, D. 2016. 'Constitution and the Debate between Animalism and Psychological Views'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Russell, B. 1910/11. 'Knowledge by Acquaintance and Knowledge by Description'. *Proceedings of the Aristotelian Society* 11: 108–128.

Ryle, G. 1949. *The Concept of Mind*. Oxford: Routledge.

Schechtman, M. 1997. 'The brain/body problem'. *Philosophical Psychology* 10: 149–164.

— 2001. 'Empathic Access: The Missing Ingredient in Personal Identity'. *Philosophical Explorations* 4: 95–111.

Searle, J. 2007. 'Putting Consciousness Back in the Brain: Reply to Bennett and Hacker, *Philosophical Foundations of Neuroscience*'. In M. Bennett et al., eds. *Neuroscience and Philosophy: Brain, Mind, and Language*. New York: Columbia University Press.

Bibliography

Shammi, P. and Stuss, D. T. 1999. 'Humour appreciation: a role of the right frontal lobe'. *Brain* 122: 657–666.

Shoemaker, S. 1959. 'Personal Identity and Memory'. *The Journal of Philosophy* 56: 868–882.

— 1963. *Self-Knowledge and Self-Identity*. Ithaca, New York: Cornell University Press.

— 1970. 'Persons and Their Pasts'. *American Philosophical Quarterly* 7: 269–285.

— 1979. 'Identity, Properties, and Causality'. *Midwest Studies in Philosophy* 65: 321–342.

— 1980. 'Causality and properties'. In P. van Inwagen, ed. *Time and Cause*. Dordrecht: D. Reidel.

— 1981. 'Some Varieties of Functionalism'. *Philosophical Topics* 12: 93–119.

— 1984. 'Personal Identity: A Materialist's Account'. In S. Shoemaker and R. Swinburne, eds. *Personal Identity*. Oxford: Blackwell.

— 1985. 'Critical Notice on Reasons and Persons'. *Mind* 94: 443–453.

— 1997. 'Self and Substance'. *Philosophical Perspectives* 11: 283–304.

— 1999. 'Self, Body, and Coincidence'. *Aristotelian Society Supplementary Volume* 73: 287–332.

— 2004. a. 'Brown-Brownson Revisited'. *The Monist* 87: 573–593.

— 2004. b. 'Functionalism and Personal Identity: A Reply'. *Noûs* 38: 525–533.

— 2004. c. 'Reply to Wiggins'. *The Monist* 87: 610–613.

— 2007. *Physical Realization*. Oxford: Oxford University Press.

— 2008. 'Persons, animals, and identity'. *Synthese* 162: 313–324.

— 2012. 'Against Simplicity'. In G. Gasser and M. Stefan, eds. *Personal Identity: Complex or Simple?* Cambridge: Cambridge University Press.

— 2016. 'Thinking Animals Without Animalism'. In S. Blatti and P. F. Snowdon, eds. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.

Sidelle, A. 2011. 'Parfit on 'the Normal/a Reliable/any Cause' of Relation R'. *Mind* 120: 735–760.

Sider, T. 2001. 'Criteria of identity and the limits of conceptual analysis'. *Philosophical Perspectives* 15: 189–209.

— 2003. 'Maximality and Microphysical Supervenience'. *Philosophy and Phenomenological Research* 66: 139–149.

— 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Bibliography

Slors, M. 2001. 'Personal Identity, Memory, and Circularity: An Alternative for Q-Memory'. *The Journal of Philosophy* 98: 186–214.

Smit, H. and Hacker, P. M. S. 2014. 'Seven Misconceptions About the Mereological Fallacy: A Compilation for the Perplexed'. *Erkenntnis* 79: 1077–1097.

Snowdon, P. F. 1990a. 'Persons, Animals, and Ourselves'. In C. Gill, ed. *The Person and the Human Mind*. Oxford: Clarendon Press.

— 1990. b. 'The Objects of Perception'. *Proceedings of the Aristotelian Society Supplementary Volumes* 64: 121–166.

— 1991. 'Personal Identity and Brain Transplants'. *Royal Institute of Philosophy Supplement* 29: 109–126.

— 1996. 'Persons and Personal Identity'. In S. Lovibond and S. G. Williams, eds. *Identity Truth and Value: Essays for David Wiggins*. Oxford: Blackwell.

— 2005. 'Some Reflections on an Argument from Hallucination'. *Philosophical Topics* 33: 285–305.

— 2009. a. 'On the Sortal Dependency of Individuation Thesis'. In H. Dyke, ed. *From Truth to Reality: New Essays in Logic and Metaphysics*. London: Routledge.

— 2009. b. ''Persons' and Persons'. *Organon F* 16: 449–476.

— 2014. a. 'Animalism and the Lives of Human Animals'. *The Southern Journal of Philosophy* 52: 171–184.

— 2014. b. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.

— 2016. 'Dainton on Subjects of Experience'. *Aristotelian Society Supplementary Volume* 90: 145–159.

Spelke, E. S. 1990. 'Principles of Object Perception'. *Cognitive Science* 14: 29–56.

Stanley, J. 1998. 'Persons and their properites'. *Philosophical Quarterly* 48: 159–175.

Strawson, G. 2011. *Locke on Personal Identity*. Princeton and Oxford: Princeton University Press.

Strawson, P. F. 1959. *Individuals*. London: Methuen & Co Ltd.

Sutton, C. S. 2012. 'Colocated Objects, Tally-Ho: A Solution to the Grounding Problem'. *Mind* 121: 703–730.

Swinburne, R. 1984. 'Personal Identity: The Dualist Theory'. In S. Shoemaker and R. Swinburne, eds. *Personal Identity*. Oxford: Blackwell.

Tabb, K. 2018. 'Madness as method: on Locke's thought experiments about personal identity'. *British Journal for the History of Philosophy* 26: 871–889.

Thomson, J. J. 1997. 'People and their Bodies'. In J. Dancy, ed. *Reading Parfit*. Oxford: Blackwell.

Bibliography

Thornton, A. K. 2016. 'Varieties of Animalism'. *Philosophy Compass* 11: 515–526.

Toner, P. 2011. 'Hylemorphic animalism'. *Philosophical Studies* 155: 65–81.

— 2014. 'Hylemorphism, remnant persons and personhood'. *Canadian Journal of Philosophy* 44: 76–96.

Unger, P. 1990. *Identity, Consciousness and Value*. Oxford: Oxford University Press.

van Inwagen, P. 1990. *Material Beings*. Ithaca, New York: Cornell University Press.

— 1997. 'Materialism and the Psychological-Continuity Account of Personal Identity'. *Philosophical Perspectives* 11: 305–319.

Watson, J. L. 2016. 'Thinking Animals and the Thinking Parts Problem'. *The Philosophical Quarterly* 66: 323–340.

Wiggins, D. 1997. 'Sortal Concepts: A Reply to Xu'. *Mind & Language* 12: 413–421.

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

— 1974. 'Essentialism, Continuity, and Identity'. *Synthese* 23: 321–359.

— 1976. 'Locke, Butler and the Stream of Consciousness: and Men as Natural Kind'. In A. O. Rorty, ed. *The Identities of Persons*. Berkeley and Los Angeles, California: University of California Press.

— 1980. *Sameness and Substance*. Cambridge, MA: Harvard University Press.

— 1987. 'The Person as Object of Science, as Subject of Experience, and as Locus of Value'. In A. Peacocke and G. Gillet, eds. *Persons and Personality*. Oxford: Basil Blackwell.

— 1996. 'Reply to Paul Snowdon'. In S. Lovibond and S. G. Williams, eds. *Identity Truth and Value: Essays for David Wiggins*. Oxford: Blackwell.

— 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

— 2004. a. 'Reply to Shoemaker'. *The Monist* 87: 594–609.

— 2004. b. 'Reply to Shoemaker's Reply'. *The Monist* 87: 614–615.

— 2012. 'Identity, Individuation and Substance'. *European Journal of Philosophy* 20: 1–25.

Wilkes, K. V. 1988. *Real People: Personal Identity without Thought Experiments*. Oxford: Oxford University Press.

Williams, B. A. O. 1956–1957. 'Personal Identity and Individuation'. *Proceedings of the Aristotelian Society* 57: 229–252.

— 1960. 'Bodily Continuity and Personal Identity: A Reply'. *Analysis* 21: 43–48.

Bibliography

— 1970. 'The Self and the Future'. *The Philosophical Review* 79: 161–180.

— 1973. 'Are persons bodies?': in. *Problems of the Self*. Cambridge: Cambridge University Press.

Williamson, T. 1994. *Vagueness*. Oxford: Routledge.

— 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

— 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Wilson, J. 2016. 'Internal and External Validity in Thought Experiments'. *Proceedings of the Aristotelian Society* 116: 127–152.

Wollheim, R. 1979. 'Memory, Experiential Memory and Personal Identity'. In G. F. MacDonald, ed. *Perception and Identity*. Ithaca, New York: Cornell University Press.

— 1984. *The Thread of Life*. Cambridge, MA: Harvard University Press.

Word and Object. 1960. The MIT Press.

Xu, F. 1997. 'From Lot's wife to a pillar of salt: Evidence that physical object is a sortal concept'. *Mind & Language* 12: 365–392.

Yang, E. 2013. 'Thinking animals, disagreement, and skepticism'. *Philosophical Studies* 166: 109–121.

— 2019. 'Resisting the Remnant-Person Problem'. *Acta Analytica* Online Early Access:

Zimmerman, D. W. 1997. 'Immanent Causation'. *Noûs* 31: 433–471.

— 1999. 'The Compatibility of Materialism and Survival: The "Falling Elevator" Model'. *Faith and Philosophy* 16: 194–212.

— 2002. 'The Constitution of Persons by Bodies'. *Philosophical Topics* 30: 295–338.

— 2009. 'Properties, Minds and Bodies: An Examination of Sydney Shoemaker's Metaphysics'. *Philosophy and Phenomenological Research* 78: 673–738.

— 2010. 'Bodily Resurrection: The Falling Elevator Model Revisited'. In G. Gasser, ed. *Personal Identity and Resurrection: How Do We Survive Our Death?* Surrey, UK: Ashgate Publishing Ltd.