Embodied cognition, the extended mind thesis, and personal identity

Guy Shennan

BA in Mathematics, MA in Applied Social Studies, MA in Solution-Focused Brief Therapy

B2732093

Submitted for the MA in Philosophy, The Open University

September 2014

ABSTRACT

There is a fast-growing body of work within cognitive science that has become known as ‘embodied cognition’, which is posing a challenge to more traditional cognitivist models. It comprises a variety of research programmes, which share the common ground of viewing cognitive processes as not solely caused or constituted by the brain, but as being deeply dependent in some way upon the ‘beyond-the-brain’ body. At the more radical end of these ideas can be found the extended mind thesis, which views cognition as extending not just beyond the brain into the body, but into the body’s environment too.

This dissertation seeks to tease out a number of implications of embodied and extended cognition for accounts of personal identity over time, which it will do in the following way. After a brief introduction, chapter 1 will summarise some of the main positions on personal identity, focusing on psychological continuity, bodily continuity and human being accounts. One particular human being account, that of David Wiggins, will be mentioned but not referred to in detail until chapter 2. This is because, following a general survey of embodied cognition and its possible implications for personal identity, an argument will unfold from the second part of chapter 2 onwards, first showing that one particular embodied cognition programme (O’Regan and Noë, 2001) supports Wiggins in his argument against the psychological continuity position and for his human being account (an account which challenges the notion that psychological and bodily continuity can be prised apart).

Chapter 3 will introduce the extended mind thesis and consider whether extended minds entail extended selves. We show that this cannot be ruled out, which leads to the second part of our argument, that the extended mind thesis requires an expansion of our conception of human beings. In this, we draw on the

work of Andy Clark in particular. The nature of this expansion in turn suggests that personal identity might require not only individualistic continuity conditions, but also continuity of elements of the person's environment. This will be the subject of chapter 4.

In summary, we hope to show that embodied and extended cognition might offer some support for a human being account of personal identity, though with an extended concept of human being, requiring environmental and social continuity alongside both psychological and bodily continuity.

Embodied cognition, the extended mind thesis, and personal identity

Guy Shennan

BA in Mathematics, MA in Applied Social Studies, MA in Solution-Focused Brief Therapy

B2732093

Submitted for the MA in Philosophy, The Open University

September 2014

No part of this dissertation has been submitted for a degree or other qualification at The Open University or at any other university or institution

The entire work has been prepared by myself alone

TABLE OF CONTENTS

INTRODUCTION

The historian, Charles Weiner, seeing a number of original notes and sketches by Nobel Prize-winning physicist, Richard Feynman, commented that these materials represented ‘a record of (Feynman’s) day-to-day work’. Feynman reacted sharply:

> “I actually did the work on the paper”, he said.
>
> “Well,” Weiner said, “the work was done in your head, but the record of it is still here”.
>
> “No, it’s not a record, not really. It’s working. You have to work on paper and this is the paper, ok?” (Gleick, 1993, p409, cited in Clark, 2008, pxxv)

Feynman’s suggestion was that the ‘loop into the external medium’ (in Clark’s words) was integral to his intellectual activity, his ‘working’. Andy Clark goes further, by suggesting ‘that Feynman was actually *thinking* on the paper’ (Clark, 2008, pxxv[1]).

As I begin to write this dissertation, I am surrounded by a plethora of sticky post-it notes, stuck on the desk in front of me and the walls to either side. Or am ‘I’ so surrounded, or just the biological organism that is only a part of me? Clark might go further still, and claim that I am actually spread into the world around me (Clark and Chalmers, 1998, p18); that not only are the post-it notes going to be indispensable to the cognitive processing that will lead to this dissertation being written, and as such to being part of my ‘extended mind’, but that they will actually, together with the biological part of me contained within my skin and skull, make up my extended self. And if I am extended in this way, then what bearing might this have on my personal identity over time? The persistence conditions of personal identity are wrapped up in the nature of personhood, so if a person might be

[1] Emphases in quotations are in the original, unless otherwise stated.

spread beyond the brain and body of a human organism, then this might appear to have implications for accounts of personal identity.

So, let the right hand that is part of me pick up these post-it notes that may be part of me, and, together with the rest of me that is going to form the requisite cognitive system, decide how to proceed.

Embodied cognition, extended minds and personal identity

The developing interest in the possibility that cognition and mind might extend not only beyond the brain into the rest of the body (usually given the label of *embodied cognition*), but also beyond the body into the world around it (usually termed the *extended mind thesis*), has begun to lead philosophers to wonder about the implications of such extensions for agency, personhood and personal identity (Clark and Chalmers, 1998; Clark, 2003, 2005, 2008; Wilson, 2004, 2005; Wilson and Foglia, 2011). A few actual investigations of what such implications might consist of have been set in train (Baker, 2009; Olson, 2011; Mandrigin, 2007; Rhodes, 2011; Wilson and Lenart, forthcoming), though the wondering is largely outweighing the investigating as yet. In this dissertation I therefore intend to try to shed a little light on what such investigations have turned up or might do in the future.

I shall do this by first clarifying what is meant by personal identity and exploring some of the main accounts of this, restricting myself to those on which I believe embodied and extended cognition might have something to say. It will be seen that there are two questions tied up in these accounts, the first about the kind of things persons are, and the second about the persistence conditions of these kinds of things over time. I shall then turn my attention, in chapters 2 and 3, to embodied cognition and the extended mind thesis respectively. Given the wide-ranging nature of these fields of study, I shall say just enough about them in order

to be able to consider some of their possible implications for personhood and its persistence over time.

My conclusions

Let me summarise what I will be saying in terms of these implications. Having surveyed psychological continuity, bodily continuity and human being or organism accounts of personal identity (chapter 1), I shall focus in particular on the human being account of 'later Wiggins' (2001), and show that at least some elements of embodied cognition support his argument against the neo-Lockean psychological accounts and for his own (chapter 2). They do this in particular by challenging received wisdom concerning an influential brain transplant thought experiment (Shoemaker, 1963). However, I will then go on to show that the extended mind thesis challenges Wiggins in turn (chapter 3), by suggesting that our traditional human being concept needs to be expanded. In chapter 4 I will argue that this expansion might call for other, less individualistic criteria to be added to bodily and psychological continuity conditions for personal identity, namely environmental and social continuity. A thread of 'inclusivity' will run through the dissertation, first in the argument that aspects of embodied cognition support the idea that *both* bodily *and* psychological continuity are required for personal identity; and second in the argument that extended cognition implies that *both* factors external to a person *as well as* internal features are important for continuing personhood and identity.

CHAPTER 1

PERSONAL IDENTITY

Two questions

Parfit considers 'it will help to distinguish' two questions tied up in accounts of personal identity, which he states as being:

> 1) What is the nature of a person?
>
> 2) What makes a person at two different times one and the same person? What is necessarily involved in the continued existence of each person over time? (1984, p202)

As Wiggins notes, the first question 'leads into a thicket of philosophy, (which) anyone who has been through once… may wonder whether it can be skirted' (2001, p193). However, as Locke made clear, identity has to be 'suited to the idea' (1694, p37), and the different 'ideas' we might have, whether of substance, or man, or person, to use Locke's language, will lead to different identity conditions. So, its difficulty notwithstanding, our considerations of personal identity will necessarily include some responses to this question about the kind of thing a person is. In particular, the extended mind thesis may also affect our ideas about what persons are, as we shall see in chapter 3.

There is not universal agreement about the framing of the second question above, which is about persistence. Note how Parfit asks what makes a *person* at one time the same as a *person* at a different time. According to Eric Olson, a major proponent of the animalist account of personal identity (1997), this wording precludes the possibility that '*you* were ever an embryo or a foetus, or whether *you* could survive in an irreversible vegetative state or as a corpse' (Olson, 2010[2]; emphases added, to highlight the ambiguity of the 'idea', to use Locke's word, in

[2] Where no page number is given for a quotation, this is because it is from an online article without page numbers.

Olson's use of the indexical *you*). Therefore, Olson words his persistence question as follows:

> 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? (1997, p25)

This appears to suggest that Olson is not actually asking about *personal* identity at all. I believe this runs like a fault line through his whole account, to which I shall return below.

Some important concepts

There are some important terms and concepts in relation to identity, some of which I will now clarify, as they will be relevant in what is to follow. First, time is an important factor in discussions of personal identity. The distinction to be made here is between synchronic and diachronic identity, the former referring to identity at a particular point in time, the latter to identity across different times. It is the latter we are interested in with regard to personal identity, at least, with regard to the persistence question. When we consider identity over time, the persistence question becomes fascinating and apparently paradoxical. As the logician, Irving Copi, put it:

> 1. If a changing thing really changes, there can't literally be one and the same thing before and after the change.
> 2. However, if there isn't literally one and the same thing before and after the change, then no thing has really undergone any change. (cited in Gallois, 2012)

The mention of 'literally one and the same thing' leads us next to the distinction between qualitative and numerical identity. A white billiard ball may be qualitatively identical to another white billiard ball, made out of the same stuff at the same factory to the same specifications. The identity relationship here is one of

exact similarity, rather than the two balls being 'literally one and the same thing'. However, the application of a coat of green paint to one of these white billiard balls would probably not cause it to be said that it was no longer literally one and the same thing as the resulting green ball. Here the relationship is one of numerical identity, and it is the conditions for the holding of this relationship over time that we are interested in when considering personal identity, or at least that is my contention.

Let me also underline that, in the philosophical study of personal identity, identity is to be understood as a *relationship* between persons - for example, the writer of this dissertation today is identical to the person who at the same desk was writing his project a year ago (so I might claim) - rather than as a *characteristic, role* or *property* of a person - for example, this writer is concerned that he is about to lose his identity as a philosophy student. And finally, identity is an equivalence relation, which means it has the properties of transitivity: if a is identical to b and b is identical to c, then a is identical to c; symmetry: if a is identical to b, then b is identical to a; and reflexivity: a is identical to itself.

Personal identity over time

Let us now turn to some of the main accounts of personal identity over time, or in other words, answers to the persistence question as posed by Parfit above. The currently dominant view is based on seeing persons as minds (Johnston, 1987, p61), and is defended largely by what Mark Johnston calls 'the method of cases' (1987, p59seq). He is referring to the use of thought experiments which aim to draw on our intuitions about what happens to the person when mind and body are seemingly prised apart, by, for example, brain transplants or teletransportation. There are a number of difficulties in using such cases, and the theses of embodied cognition and the extended mind have raised further questions about them. A

problem with intuitions is that they might be based on outdated received wisdom, and what is believed to be the case following brain swaps might be called into question if, for example, the mind is not fully contained in the brain after all. These are issues to which we shall return, especially in relation to a well-known brain transfer case (Shoemaker, 1963) and, in chapters 2 and 4, to the use of teletransportation (Parfit, 1984).

So, the dominant view is based on the idea that persons are defined by their minds and this has led to psychological continuity accounts of personal identity. These have traditionally been contrasted with bodily continuity accounts, with the main opposition to the dominant view now being clustered around a number of variants of a naturalistic view that equates persons with human beings. At one end of these variants lies the animalism of Eric Olson (1997), which has dropped psychology out of the mix altogether, while for Johnston (1987) and Wiggins (2001), mental functioning is an essential aspect of continuing personhood. As noted in my introduction, I shall be arguing that embodied cognition offers support for Wiggins' view, and that the extended mind thesis suggests ways in which his account could in its turn be extended, to fit with Andy Clark's wider conception of human beings. Let us begin our brief survey here though with the dominant view.

Psychological continuity

This view dates back to John Locke, and can be seen as a refinement of his suggestion that continuity of consciousness was the essential criterion[3] of personal identity over time. By this suggestion, Locke was breaking with the Platonic and Cartesian view that saw the soul account for personal identity (Nimbalkar, 2011),

---

[3] One more clarification of terms is required. I am using the ambiguous word 'criterion' here not to refer to *evidence* that one person is the same as another, but to '*what this identity necessarily involves, or consists in*' (Parfit, 1984, p202).

which lingers on as the Simple View (Swinburne, 1973-4). This view, more popular now with the person on the street[4], especially if religiously inclined, than with philosophers, is not one that I will be addressing here. This is partly for reasons of space, though I also think it unlikely that new developments in cognitive science about how and where the mind is constituted could have much bearing on an idea which holds that a person is not reducible to psychological or physical states.

In paving the way for the modern study of personal identity, Locke used the method of cases himself, including famously when he imagined the soul of a prince replacing that of a cobbler in the cobbler's body, and claimed that 'everyone sees he would be the same person with the prince' (1694, p44). His claim was not based on the presence in the cobbler's body of the prince's soul per se, but because this was the bearer of the consciousness of the prince's past life. So for Locke, experiential memory was the essential criterion of identity. A is the same person as B at an earlier time if A can remember the doings of B. Unsurprisingly, this criterion soon attracted its critics, with Reid objecting that A might remember something B did 20 years ago, and B might remember something C did 20 years before that, but A would in all likelihood have forgotten C's action, ruling out the memory criterion by *reductio*. This has been quite easily dealt with by the refiners of Locke, who defined the transitive relation of psychological continuity as overlapping chains of psychological connectedness, where the latter relation was expanded to include other psychological states in addition to memory, such as beliefs, desires and intentions. This expansion also guarded against another objection to Locke, which was that his memory criterion appeared to entail the counter-intuitive consequence that memory loss precluded persistence of the person.

---

[4] The term *person* on the street being more my attempt at updated non-sexist language than an allusion to the philosophical idea being interrogated in this essay.

A more enduring objection is the one made by Butler, who said that Locke had made a ‘wonderful mistake’ (1736, p100), by not recognising the circularity of the memory criterion, in that experiential memory presupposes identity and so cannot constitute it. So, rather than A remembering having B’s experiences making A the same person as B, it is because A is the same person as B that A remembers B’s experiences. This powerful objection might have been the end of a psychological account, but the idea that we are defined by our mental states and psychological character has been strong enough for modern philosophers to look for ways to accommodate Butler’s point. This has been done most rigorously by Parfit (1984), building in particular on the work of Shoemaker (1970). A crucial step was to find an impersonal alternative to memory, which could form the basis of a non-circular psychological relation. This led to the creation of quasi-memory, defined as follows:

> I have an accurate quasi-memory of an experience if
>
> 1) I seem to remember having an experience
>
> 2) someone did have this experience, and
>
> 3) my apparent memory is causally dependent, in the right kind of way, on the past experience. (Parfit, 1984, p220)

The purpose of the third part of this definition is to tie a memory (the same causal dependency condition applies to other mental states) to an experience in such a way that ensures psychological connectedness as opposed to a random psychological similarity. Parfit suggests that the ‘right kind of cause’ can be the normal cause, such as what has caused me to remember being at my desk last year writing my project; any reliable cause, such as a brain scanner that would fully and infallibly record the contents of my brain to be uploaded into another; or any cause.

Parfit is now close to having established his criterion of personal identity, by having defined the Relation R, which is ‘psychological connectedness and/or continuity with the right kind of cause’ (1984, p262). However, he adds a uniqueness condition to ensure that this is an identity relation, and we can understand his reason for doing so if we consider the possibilities arising from fission. Parfit relies heavily on the method of cases (see above) to argue for his psychological account, a typical case being one in which A’s brain and body is scanned and then destroyed, before being replicated elsewhere a moment later as B. It is not hard to imagine that another person C could be replicated at the same moment from the same blueprint of A. It would appear that R holds between A and B and between A and C (or so Parfit and the neo-Lockeans believe), but not identity in either case. For if A was identical to B, then it would have to follow that A is identical to C, which due to the transitivity of identity would mean that B was identical to C, which would be absurd, as they are two separate entities. So, following Parfit’s elegant abbreviation (with PI standing for personal identity and U for uniqueness) of his account, which most thoroughly represents the dominant psychological view of personal identity, we have

PI = R + U. (1984, p263)

Bodily continuity

Other views are often defined by their opposition to the psychological continuity account, for example the bodily continuity account most frequently associated with the work of Bernard Williams (1956-7, 1970a, 1970b). Simply stated, this holds that the criterion of personal identity is ‘the spatiotemporal continuity of a functioning human body’ (Korfmacher, 2006). It is arguable whether this is an equivalent account to those which state the criterion as being the continuity of the same human being or human organism. Olson does not seem to think so, as he

states that he has never been able to work out what the bodily criterion is meant to be (2006, p242). For his part, Wiggins (1998, p309) alludes to Williams's 'notorious and surprising suggestion' (in 1970b) that persons exist as long as their bodies exist. Wiggins' comment strikes me as a little odd, for a couple of reasons. First, it is not clear to me that Williams actually says this in the paper referred to[5]. Second, Williams' views on personal identity are actually more nuanced than this, and it is clear elsewhere that he does not actually identify person and body. In his earliest paper on the topic (1956-7), he defends the necessity of bodily continuity for personal identity while stating clearly that it is not a sufficient condition (1956-7, p1). Williams also anticipates Wiggins' later doubts (2001) about the very possibility of prising personality away from bodily features (1956-7, p11-12), thereby offering support to my overall argument, as we shall see in chapter 3.

In a later paper, Williams (1970a) challenges the psychological account of personal identity by presenting a thought experiment in two versions, which seem to elicit conflicting intuitions. Each version involves prospective torture, in the first case to two persons who undergo a brain state transfer, and in the second case to a person who first undergoes massive psychological changes. Williams concludes with regard to the latter case that it is reasonable to fear future pain whatever psychological changes precede it. Therefore, in the first case, given the choice, one should decide that the torture should be inflicted on the other body, following an operation whereby the information in one's brain was swapped with the information in the other. In other words, Williams is coming down in favour of a bodily continuity account, though cautiously, as he suggests the decision would be risky (1970a, p63). He is concerned again here to draw our attention to the difficulty in the very notion of 'changing bodies'.

---

[5] In any case, I believe Williams' argument to be based on a misreading of Peter Strawson's concept of a person (see Williams, 1970b, fn2), a concept I will be describing in chapter 3.

Human being accounts

I will turn finally in this chapter to accounts of personal identity which see persistence conditions as being connected in some way with the concept of a person as a human being or organism. Given the role that it plays in Mark Johnston's account (1987), this exposition will provide me with the opportunity to describe Shoemaker's brain transfer case mentioned above, which will play an important part within the arguments of the next chapter. These arguments will concern Wiggins's human being account (2001), so I will only mention this in passing here. Let me begin with Olson, as we will be looking closely at his views on implications of the extended mind thesis in chapter 3.

Olson states that his animalism, or 'biological approach', makes two claims (1997, p17). First, 'you and I are animals: members of the species *homo sapiens*, to be precise', and second, 'psychological continuity is neither necessary nor sufficient for a human animal to persist through time'. By making the first claim, Olson is not ruling out that there may be non-human persons[6]; his claim is rather that all human persons are animals. The word 'are' is ambiguous here, though Olson promptly tries to clear this up by clarifying that he does not mean that 'your *body* is a human animal, or that you are "constituted by" an animal', but that 'you' are numerically identical with an animal, 'that living primate sitting in your chair right now' (1997, p17). This is of course trivially the case if the pronoun 'you' is standing in for the human animal (it is interesting that Olson does not use the term 'human being'; see below with reference to Mark Johnston), so I assume he intends to mean: 'the person that is you is numerically identical with the living animal in your chair'.

---

[6] Olson uses 'people' rather than 'persons' throughout, apparently to use ordinary, non-philosophical language (1997, p6). It strikes me that this is more likely to be associated with Olson's lack of precision in his use of indexicals such as 'you' and 'I' and in his *idea* (to use Locke's term) of 'person'.

Olson states that his second claim is more controversial, though I am not so sure. It might be controversial to say that psychological continuity is not necessary for a *person* to persist through time. However, Olson is arguing here against a persistence condition for the 'human animal'. He believes his claim is evident when we consider what it means to be a human organism. To support this, he refers to the case of a person who enters a persistent vegetative state[7]. An adherent of a psychological approach to personal identity would have to consider the possibility that the person does not survive this state, although, of course, a living human animal persists. So in fact, I agree with Olson that the second claim of his biological approach is evident, or rather I would say that it is self-evident and so in fact says very little. It is not the persistence conditions of the human organism that are at stake in questions of personal identity, but those of persons.

Let us compare another account, similar but different to Olson's, that set out by Mark Johnston (1987). Johnston takes Olson's first claim as his starting point, that we are essentially the organism *homo sapiens*, but then immediately departs from this because of what he sees as the implications of Shoemaker's puzzle case, which I will now sketch it out before examining its effects on Johnston (and in the next chapter, on Wiggins).

The case involves two men, Brown and Robinson, who have their brains temporarily extracted in order to ease the removal of brain tumours (this is in a somewhat more advanced National Health Service than at present). The operations are successful, but the brains are mistakenly replaced in the wrong bodies. One of the men immediately dies, but the other, the one now consisting of Robinson's body and Brown's brain, survives and eventually regains consciousness. Shoemaker names the latter Brownson, and says that when asked

---

[7] Although he does not actually use the term, *person*, but *you*, which continues to add to the lack of precision, or clarity in intended meaning, in my view.

his name, Brownson automatically replies ‘Brown’, that he recognises Brown’s wife and family, is able to describe in detail events in Brown’s life, while evidencing no knowledge of Robinson’s past life (1963, p23-4).

Johnston reports that the ‘predominant reaction’ to this case is that Brown has survived as Brownson, and considers that this intuition would remain even if Brown’s body - without Brown’s or Robinson’s brain - was given enough brain-stem tissue to be kept alive indefinitely. Johnston gives this surviving human organism the name Brownless and suggests that, as an organism, Brown has survived as Brownless rather than Brownson, since for the survival of organisms, ‘metabolism is more important than mentation’ (1987, p76). He goes on to distinguish between human beings and human organisms, by defining human being as being one of a kind ‘such that its members survive if their mental life continues on as a result of the survival of their organ of mentation.’ (1987, p79). This is how Johnston explains the supposition in cases such as Shoemaker’s that the person goes where the brain goes.

Note that this is different from saying that a person is of the kind ‘human brain’, which actually is the view of a small number of philosophers (Mackie, 1980; Nagel, 1984). Johnston suggests that what motivates this view is the observation that ‘the survival of one's brain can be sufficient for one's survival and may well be necessary for it’ (1987, p79). This is what prompted Parfit to suggest that a ‘physical criterion’ of personal identity required, ‘not the continued existence of the whole body but the continued existence of *enough* of the brain to be the brain of a living person’ (1984, p204). It is worth noting these ideas within personal identity accounts, as doubts about the sufficiency of the brain are central to the challenges posed to traditional cognitive science by embodied and extended cognition, as we shall see.

CHAPTER 2

EMBODIED COGNITION

Introduction

In this chapter I will be considering some implications of embodied cognition for personal identity. This is potentially an unwieldy task, given the multifariousness of both embodied cognition and accounts of personal identity. I shall address this by dividing the chapter into two parts. First I will make some fairly general comments about implications for personal identity during a brief overview of the embodied field. In the second part I will focus on what I have termed 'later Wiggins' (2001), where Wiggins challenges neo-Lockean psychological continuity arguments, to support his own human being account. At the centre of both his challenge and his positive argument is a reconsideration of Shoemaker's Brown-Brownson case (1963), a case which has been central to the growing dominance of the psychological continuity view. I shall attempt to show that if certain views within embodied cognition are accepted, then the intuitions that have Brown surviving as Brownson are mistaken, which will in turn support Wiggins's argument. In his argument, Wiggins makes use of Strawson's 'primitive concept' of a person, as a being necessarily having both physical and psychological properties (Strawson, 1959). I shall therefore also summarise this philosophical concept, and suggest its congruence with the more scientifically-based embodied cognition. The problems that Wiggins will then run into if we accept a further view within embodied cognition - the extended mind thesis - we will save for the next chapter. Let us begin this one with an overview of embodied cognition.

Embodied cognition

Embodied cognition is commonly understood as being an alternative or a challenge to standard cognitive science (Shapiro, 2011, p1). To simplify matters,

we can show the contrast between the two by characterising the latter, also known as cognitivism, as situating cognition entirely within the head, or, more accurately perhaps, within the brain, while the former sees the beyond-the-brain body as being integrally involved. As Wilson and Foglia put it in their 'embodiment thesis':

> Many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing. (2011)

There are a variety of strands under the overall rubric of embodied cognition, as we shall see, and this means that one cannot simply discuss its implications for personal identity as if these strands form a single unified set. There might seem to be a *prima facie* case for the embodiment thesis entailing a bodily continuity requirement for personal identity. However, I shall argue that most of the work done within the embodiment field does not suggest such a requirement, and could sit quite easily with a psychological continuity account. The strands of embodied cognition are grouped together by Shapiro into three categories, which he terms the conceptualisation, replacement and constitution hypotheses respectively. I will briefly sketch this categorisation, to give a flavour of the work being done within embodied cognition. As I do this for the first two categories I will also share some considerations about implications for personal identity. It is in the last of the three where we will find accounts of cognition that do present clear issues for accounts of personal identity. Among these is the sensorimotor account of visual experience (O'Regan and Noë, 2001), which appears to present a challenge to the current neo-Lockean view, and so I will look at this in some detail. Extended cognition, which will be the subject of the next chapter, also comes

under the constitution hypothesis. But let us consider the conceptualisation and replacement hypotheses first.

Shapiro's use of the term 'hypothesis' fits with his view that embodied cognition is best seen at this early stage in its development[8] as a diverse research programme. Work that is testing out the *conceptualisation* hypothesis seeks to show that an organism's understanding of the world is somehow determined by its bodily properties (Shapiro, 2011, p68). An example of this comes from the work of Lakoff and Johnson on metaphors. Their central claim is that 'most of our normal conceptual system is metaphorically structured… most concepts are partially understood in terms of other concepts' (Lakoff and Johnson, 1980, p56). They argue that this is shown by the language we use to describe concepts. For example, *love* is often understood as a *journey*, as we might ask *where is our relationship going*, consider *how far we have come*, or perhaps agree to *go our separate ways*.

The question arises of how such a system is grounded, as not all concepts can be explained by means of other concepts. There has to be a starting point, and some 'basic concepts' that are understood directly. Lakoff and Johnson claim that these are concepts that arise from our bodily experience. For example, the spatial concept *up* - which structures the concept *happy*, for example - I'm feeling *up*, his spirits *rose* - arises from the fact that our bodies stand erect and 'almost every movement we make involves a motor program that either changes our up-down orientation, maintains it, presupposes it, or takes it into account in some way' (1980, p56). Other examples of such basic concepts are *front* and *back*, and the claim is that these concepts, derived from the kinds of bodies we have and how they act in our environment, are the building blocks for our entire conceptual

[8] Although embodiment has philosophical roots, stretching back to Dewey (1925) and Merleau-Ponty (1962), most of the scientific work on embodied cognition has been carried out within the past 20 years.

system. The conclusion that Lakoff and Johnson draw is that if we had different kinds of bodies, then our understanding of the world would necessarily be different. To support this, they posit a spherical being living outside any gravitational field, which has no way of knowing or imagining any other kind of experience, and suggest that this being would not be able to develop the basic concepts, of *up* and *down* and so on, that are fundamental to our understanding (1980, p57).

Let us suppose that Lakoff and Johnson are correct in their view that our understanding of the world is ultimately determined by the kinds of bodies we have, and consider the implications this might have for personal identity. It might seem at first glance that it supports a bodily continuity account, given the importance it ascribes to our bodies for our understanding of the world. However, according to Lakoff and Johnson, this understanding seems to depend only on having the particular *kind of body* we have, and not on continuing to have the same *actual* body. So the most that could be claimed here regarding bodily continuity is that to continue as the same person might require the continuation of the same kind of body, which is something that a psychological continuity adherent can accept (see, for example, Parfit, 1984, p285). However, accepting Lakoff and Johnson does not require even this. Note the condition, for their imaginary spherical being, that it has no knowledge of another kind of experience. This suggests that as long as some beings have existed with the kinds of bodies that we have, then any being can develop the same understanding of the world that we have, simply by gaining knowledge of our understanding. It does not have to continue in this kind of body, or even to have had this kind of body itself at any point. It is beyond the scope of this short dissertation to consider the implications of each specific contributor to ideas of embodied cognition, but I think it likely that

the same will apply to other ideas coming under the conceptualisation hypothesis. Our understanding of the world is unlikely to depend sufficiently strongly on our actual bodies as opposed to the kind of bodies we have.

I believe that a similar point will apply to ideas coming under the *replacement* hypothesis. This groups together work which aims to replace the typical tools of cognitivism and to find a *better* way to do cognitive science (Shapiro, 2011, p114). The cognitivist sees the mind as operating entirely within the brain, which acts as a computer, receiving input from the senses, transforming this into symbolic representations on which it then performs computations, before finally producing the behavioural output. The disembodied nature of this model is clear:

> ...(n)ote that because the cognitive system traffics only in symbolic representations, the human body and the physical environment can be dropped from consideration; it is possible to study the cognitive system as an autonomous, bodiless, and worldless system whose function is to transform input representations into output representations. (Van Gelder, 1995, p373)

Van Gelder, amongst others, makes a case for *replacing* this computational framework with a dynamical systems theory approach, which sees cognition 'emerging from continuous interactions between a body, a brain and a world' (Shapiro, 2011, p156). Advocates of replacement share the view that embodiment - the brain being in *a* body - and situatedness - 'how the world's structure imposes constraints and opportunities relative to *the type of body* an organism has' (2011, p156, emphasis added) - are the core explanatory concepts within cognitive science. But being in *a* body, of a certain *type*, appears to be

sufficient for these explanatory concepts and so, again, it does not seem that bodily continuity accounts are favoured here over psychological continuity[9].

In the *constitution* hypothesis the main point of contention with cognitivism concerns the *constituents* of the mind. Where standard cognitive scientists situate the processes constituting the mind entirely within the brain, advocates of constitution assert that the body is, literally, part of the mind, or even that the world beyond the body can be such a constituent - the idea of the extended mind, referred to above. It is the ideas subsumed in the constitution hypothesis of embodiment that have the clearest implications for personal identity, given the attention they pay to how the mind is constituted and hence, perhaps[10], to how a person is constituted. This is especially clear in the case of the extended mind, for the question naturally arises, does an extended mind entail an extended self? That question can wait until the next chapter. Let me now turn to a constitution-based account of sensory experience, and its implications for personal identity.

A sensorimotor account of sensory experience

O'Regan and Noë (2001) focus on visual experience, though their theory applies to sensory experience more broadly. They reject the standard view, which holds that vision results from activations of internal representations in the brain, which implies that 'an envatted brain, stimulated in the right way, should have the same visual experiences as a normally embodied brain' (Shapiro, 2011, p164). They propose an alternative, embodied framework, by emphasising the importance of bodily movement to perception, thereby extending 'the constituents of perception into the bodies of perceivers' (Shapiro, 2011, p164). According to this framework,

---

[9] The conclusions with regard to conceptualisation and replacement types of embodied cognition are not that bodily continuity accounts are *unsupported* by them, but that their implications for personal identity are neutral. My conclusions have been framed as the embodied cognition under review here *not* supporting bodily continuity in response to the prima facie suggestion that they *might* mentioned earlier.

[10] The 'perhaps' because there is great debate about the entailment here, as we shall see.

‘vision is a mode of exploration of the world that is mediated by knowledge, on the part of the perceiver, of what we call sensorimotor contingencies’ (O’Regan and Noë, 2001, p940). By sensorimotor contingencies they are referring to ‘the *structure of the rules* governing the sensory changes produced by various motor actions’ (2001, p941).

There are sensorimotor contingencies that relate to the visual apparatus and another set determined by the visual attributes of the objects that are presented to that visual apparatus. Examples of the rules in the first group include those concerning the effects on retinal stimulation of making certain eye movements. For instance, shifting an eye’s focus from the midpoint of a horizontal line to a point above it would lead to a line represented on a flattened-out retina as straight becoming curved: ‘in general, straight lines on the retina distort dramatically as the eyes move’ (O’Regan and Noë, 2001, p941). Another rule relates to forward body movement leading to an expanding flow pattern on the retina, and backward movement to a contracting flow.

The second group of sensorimotor contingencies relate to the visual attributes of objects. A good illustration of those concerning the attribute of shape is found in the records of patients whose vision had been restored after having been born blind with congenital cataracts. ‘One such patient… is surprised that a coin, which is round, should so drastically change its shape when it is rotated (becoming elliptical in projection)’ (O’Regan and Noë, 2001, p942). Similarly, a teenage boy who was shown his father’s picture in his mother’s locket, post-treatment, was startled ‘that a large face could be express’d in so little room’ (ibid.). These examples point to the routine knowledge of perspective and how movement effects this that is held by people with normal vision. O’Regan and Noë claim that ‘the visual quality of shape is precisely the set of all potential

distortions that the shape undergoes when it is moved relative to us, or when we move relative to it' (2001, p942). It is the knowledge of the laws abstracted from this set of distortions that enables the perception of shape.

An important point to make about such knowledge of sensorimotor contingencies is that it is practical rather than propositional: know-how, not know-that. Visual experience arises from the mastery of a set of skills based on the knowledge of laws that is largely tacit - most perceivers would not be able to describe what actually happens on the retina when moving their eyes this way or that. And, crucially for how we are about to apply these ideas to Shoemaker's brain transplant case, this knowledge and the visual experience that arises from it are, according to O'Regan and Noë, determined by the particular features of the particular visual apparatus of the perceiver. They clarify this in their reply to doubts expressed in a response to their article, about whether 'every small difference in the low-level details of sensing and acting will make a difference to the conscious visual experience' (Clark and Toribio, 2001, p980). By low-level, the respondents were referring, for example, to the emphasis on the small details of the effect of eye movements on a retina. We do not need, for our purpose of considering implications of these embodied accounts of experience, to take sides in these disputes. However, airing this objection enables me to report the responses made by the article's authors, which will make some of these implications all the clearer. First:

> … where there are physical differences, there are also qualitative differences. To this extent, our proposal deviates from classical functionalism. That is, we are rejecting a certain way of thinking about the multiple-realizability of functional systems. In particular, we hold, as Clark & Toribio are right to point out, that differences in our bodies (and thus in sensorimotor contingencies)

will make a difference to our experiential states.' (O'Regan and Noë, 2001, p1013)

And later:

> For two systems to have the same knowledge of sensorimotor contingencies *all the way down*, they will have to have bodies that are identical *all the way down* (at least in relevant respects). For only bodies that are alike in low-level detail can be functionally alike in the relevant ways.' (2001, p1015)

To summarise this embodied cognition account, specifically of visual perception and experience but which is generalisable to other sensory experiences, 'perceptual experience *just is* a mode of skilful exploration of the world' (Noë, 2004, p194), where the specific nature of this experience depends on the specific skills of the perceiver, which depend in turn on the perceiver's *particular* visual (or other sensory) apparatus.

We are now in a position to apply this account to the question of personal identity arising from Shoemaker's brain transplant case, and I think the conclusion to draw is quite clear. To recall what was said about the case in the previous chapter, the predominant reaction is that Brown survives as Brownson, and this was taken, more or less, to establish that Brown is Brownson. In his later musings on these reactions, Wiggins concluded that they had two distinct origins. Most people who took this view believed that, assuming the brain transfer was executed sufficiently well, 'Brownson's *experience* would be a subjectively seamless continuation of Brown's' (Wiggins, 2001, p207, emphasis added). A second group (smaller, Wiggins thought, and including himself) believed Brownson was Brown because, by virtue of receiving his brain, 'the seat of memory and *consciousness*', he was 'the functional inheritor and continuator of all of Brown's vital faculties' (2001, p207, emphasis added). The reasoning of the first group led them

to a psychological continuity account, whereas it will be seen that the second group, including Wiggins, were thinking along the same lines as Mark Johnston, twenty years later (see chapter 1).

It will be clear that if our sensory experience arises in the way that O'Regan and Noë propose, then neither of these reasons for believing that Brown continues as Brownson hold water. Assuming the bodies of Brown and Robinson not to be identical, then there will be differences in their visual apparatus and in the apparatus relating to their other sensory modalities. This will imply in turn that there will be differences in their sensorimotor contingencies and hence they would have different experiences. O'Regan and Noë see consciousness too as governed by a person's sensorimotor contingencies (2001, p963). They also state that the brain 'must be "tuned to" these laws of sensorimotor contingencies' (2001, p943), and hence it will be integrally involved in the active exploring that leads to sensory experience and consciousness (and this much should be obvious). So, on their account, it is hard for us to make sense of what it would be like to be Brownson, or if 'he' could have survived at all. If a person did persist, it is not clear what their experience would be like or what sense they would be able to make of the world, and there is certainly no reason to believe that the person would be Brown, any more than it would be Robinson.

However, this does not in itself rule out the psychological continuity account, as the same implications of the sensorimotor contingency theory do not follow for Parfit's teletransportation thought experiments. These typically involve a person entering a teletransporter and losing consciousness, when a scanner then destroys the person's brain and body, having first recorded the exact state of all its cells. This information is transmitted to a 'replicator' machine on Mars, which creates a new body exactly like that of the person who walked into the

teletransporter on Earth, out of new matter. Having the same brain states and the same (qualitatively identical) visual apparatus as Parfit on Earth, the replica should also have the same mastery of the same set of sensorimotor contingencies, and in this respect the replica's sensory experience and consciousness will be the same as the person's who stepped into the teletransporter on Earth. Therefore, psychological continuity is not ruled out by the sensorimotor account in this case.

Later Wiggins

At this point let us turn to Wiggins's account, in which he rules out the possibility of identity in the case of teletransportation, as this does not obey a principle he sees as central in determining identity through change. He attempts to use the same principle to show that Brown does not survive as Brownson, and it is here that he receives support from O'Regan and Noë and other embodied cognition theorists, as we shall see below. Wiggins's aim is to demonstrate the validity of his human being account of personal identity, over psychological continuity accounts. His argument is involved and difficult, and I will need to summarise it in places here, but it should be borne in mind that my main objective is to show the support he receives from embodied cognition.

In *Sameness and Substance Renewed* (2001), Wiggins replaces his previous accounts (1967, 1976, 1980) in order to strengthen his human being conception of personal identity. He begins by stressing the importance, when presented with the question of whether A is the same person as B, of being clear about what kind of thing a person is. He fundamentally identifies *person* with *human being*, though this still leaves open the question of what kind of thing a human being is, and this question will be addressed in what follows. To summarise in advance, Wiggins claims that to be human is to have both mental and physical

properties that cannot simply be separated, at least, without doing serious damage to our concept of a human being. But I am getting a little ahead of myself.

Central to Wiggins's argument is his reconsideration of the Brown-Brownson case. As noted above, Wiggins had originally shared the view that Brownson, having Brown's brain and because this was 'the seat' of Brown's 'memory and consciousness', was the same person as Brown. He was therefore 'deflected… towards the neo-Lockean conception of personal identity' (2001, pxiii-xiv). The shift in his position involves the acceptance of Butler's circularity objection to Locke's memory criterion (see chapter 1), which he had previously rejected (see, for example, 1976). Butler's objection in itself did not affect the acceptance of the idea that Brown continued as Brownson: '(n)othing appears to prevent us from thinking of Brownson as having full cognitive responsibility for the claims that he makes from direct personal memory' (2001, p226).

Where Butler's objection does cause a problem is in the variant of Shoemaker's case where Brown's brain is split in two and the halves placed into twin Robinsons, resulting in Brownson (1) and Brownson (2). Wiggins argues that the most that can be claimed here is that it is 'as if' Brownsons (1) and (2) remember, as his interrogation of Parfit's concept of quasi-memory in relation to this outcome leads him to conclude it is ill-defined: 'neither Brownson (1) and Brownson (2) is the same as Brown, and there is no newly minted, properly defined *remembering-of-experiences* relation… in which Brownsons (1) and (2) can stand to Brown' (2001, p226). Wiggins then considers a further variant, in which the transfers of the halves of Brown's brain result in only one surviving person, Brownson Sole, with the other dying. Wiggins invokes the *Only a and b* rule from logic to exclude the possibility of Brown and Brownson Sole being

identical: judgements of the identity of a and b cannot depend on what happens to a third thing c (2001, p231).

He then imagines a challenge being made that this rule, while plausible as an abstract logical requirement, is unworkable in practice, which he responds to by invoking a principle that he then places at the centre of his whole argument. Let me call this principle **P**, for the sake of clarity. **P** holds that positive judgements of identity after a change has taken place require this change to preserve 'the principle of activity'[11] of the kind of thing in question, and equally that a verdict of identity should be withheld 'in any case where the change that is in question would, in *other* cases that exemplified the same process, *fail* to preserve that principle of activity' (2001, p232). Wiggins points out that this principle is not satisfied in the Brownson Sole case, as Brownson Sole came into being by the same process that led to Brownsons (1) and (2), and that this process had not been sufficient to preserve the principle of activity of human beings as cognitive beings, which Wiggins takes to have been established by the problems he found with the quasi-remembering concept.

Applying **P** to the Brown-Brownson case, Wiggins asks how much difference there is in the processes that lead to Brownson, Brownsons (1) and (2), and Brownson Sole coming into being respectively, and 'if the second and third (cases) cannot yield identity, can the first?' (2001, p233). I should note here that the sensorimotor account provides a straightforward 'no' in response to this question. Wiggins begins his own answer by reminding the reader that when confronted with a question of personal identity one needs to be clear about the kind of thing a person is, and so he turns to unpack the concept of person as human being. He draws on Locke's idea of *person* as a 'forensic term', and points

[11] This is a technical term introduced by Wiggins which I will leave undefined, as the defining would take up too much space, in the hope that its meaning is, roughly at least, clear.

out that for a term to be forensic it must have a public use, within the ‘interpersonal sphere’ (2001, p234). He cites Strawson’s work on ‘reactive attitudes’ (1962) and asks what this requires of the concept of person, so that human beings are able to react to each other in ways that are essential to their very humanness. Wiggins believes he finds an answer to this question in another idea of Strawson’s, that of the concept of a person as a bearer of both M- and P-predicates (Strawson, 1959, p104). M-predicates are those which can be properly applied to material bodies whether or not they are capable of consciousness, for example, ‘is in the study’ or ‘weighs 13 stone’. P-predicates refer to all the other predicates that can be applied to persons. Each of these predicates imply the possession of consciousness and they comprise a wide-ranging set, which includes, for example, ‘is smiling’, ‘looks out of the window’, ‘is in pain’, ‘believes in God’, and so on. Crucial to Wiggins’s purpose in using this concept is the conclusion drawn by Strawson that:

> … when we self-refer we refer to an entity which has two sides or aspects, the physical and the mental, and not to a thing which possesses only the mental sort of feature, something else having the physical features… Persons are things with two aspects – bodily and mental. (Snowdon, 2009)

What Wiggins goes on to describe draws on this concept of a person in which the bodily aspects and those aspects that might be deemed psychological, or suggestive of the person’s character, cannot be prised apart. ‘How a human being stands or walks or frowns or smiles or laughs or sulks or earnestly entreats, or how he fries an egg, this is one part of what he is’, just as much as those actions that might be more commonly thought to express a person’s psychology, such as what he or she might choose to say, or how musically he or she might play the violin (2001, p234). What Wiggins is getting at here is not so much that the ways a person physically appears and behaves express aspects of that person’s

character and so are as much an essential part of the person as their character, as even this implies a separateness, but that the two aspects are one and the same. The mental does not supervene on the physical so much as they necessarily co-exist together. (The difficulty finding the language to express this might be one of the enduring and unhelpful legacies left us by the Cartesian dualist hegemony). And it is the physical aspect of this co-existence that provides the human presence that can be responded to by others. Wiggins sums up his point as being that a person's character is not independent of his or her physiognomy or body. In this he is supported by 'some remarks on bodily interchange' made by Bernard Williams (1956-7, p11-12), who sees it as a puzzle that 'when we are asked to distinguish a man's personality from his body, we do not really know what to distinguish from what' and concludes 'I take it that this was part of what Wittgenstein meant when he said that the best picture of the human soul was the human body[12]' (p12). So, if Brownson is Brown then the bodily and other aspects of Brown have been separated from each other in a way that is hard to conceive.

Wiggins wonders why these considerations, of the 'properly forensic view of human being-hood' (2001, p236) and of the associated importance of physiognomy and so on, did not prompt doubts about the Brown-Brownson case when it was first presented. I might point out here that this was a considerable time before ideas of embodied cognition gained currency, and venture the tentative suggestion that the case might have received a different reaction if it had been presented for the first time in more recent years. In any case, given the principle **P**, Wiggins is able to conclude that the nature of the case against identity in the Brownson Sole and Brownson twins cases mean that these doubts now have to

---

[12] *Philosophical Investigations*, II, iv (Wittgenstein, 1953).

be taken more seriously. I believe that adding the ideas being developed within embodied cognition forces us to consider the doubts more seriously still.

Finally, it is via the principle **P** that we can see - in summary form at least, for space precludes a fuller elucidation - how Wiggins rules out cases such as Parfit's teletransportation. Remember that **P** concerns the process by which a change takes place, after which a judgement of identity is to be made, and one way to think of this principle is that this process must preserve the 'principle of activity' of the things whose identity is in question. In considering what kind of things we are as human beings, Wiggins includes the ways in which change takes place for us, how our 'dispositions and capacities are gradually but constantly shaped and reshaped' (2001, p235), as well as how we are governed by the overarching human process of birth, maturing and eventual death. Wiggins then sets up a continuum of changes which might affect a human being, from minor interventions at one end, of orthodox medicine and dentistry and so on, to the far extreme 'where it seems a human being is simply treated as a template for the production of copies' (2001, p241). He sees no problem for identity in the minor interventions, as they leave the person's 'organic independence… undiminished', whereas if those at the far end 'are taken as amounting to the perpetuation of the person… we have lost hold altogether of the notions we began with of what (the person) is' (p241).

But here's the rub. Wiggins concludes this most recently articulated account of his of personal identity, with the conception of the person as a human being at its heart, by voicing his concerns about what we might be doing to this conception. He accepts that in the intermediate cases on his continuum, where large changes might result from transplants, other surgical interventions and gene therapy, the human being can still persist as an autonomous being subject to recognisable,

natural processes of change. But he fears that as we move further towards the extreme end, the conception of a person will shift away from the purely biologically human to the artefactual, with corresponding shifts in the problem of identity which will serve to bewilder us. And this is an interesting prospect, because the extended mind thesis is moving some of its advocates in this very direction, of the person becoming an assembly of the biological and the artefactual, as we shall be seeing in the next chapter.

CHAPTER 3

THE EXTENDED MIND THESIS

Introduction

I am not aware that Wiggins has commented on the extended mind, but I would hazard a guess that he might be concerned by it or at least by some of its implications. Yet I believe that one of its main advocates would surely want to put his mind at rest. Just as Wiggins is at pains to point to conditions of personal identity which necessarily contain both bodily and psychological continuity, so does Andy Clark's approach rest on a similarly 'both-and' inclusivity. While his conception of person is fluid and allows for the incorporation of elements of the environment, in more or less lasting ways, he believes that this very ability is an essential part of our humanity - we are 'natural-born' as well as being 'cyborgs' (2003). In this chapter, I will introduce the extended mind thesis (EMT) and clear the way for the consideration of some of its implications for personhood and personal identity. This will come in part via a discussion of objections to EMT, notably those of Gertler (2007) and Rupert (2004). The main burden of the chapter will be to show that it is at least possible that extended minds entail extended selves, contra Olson (2011). This leaves open Clark's expanded conception of human beings, and a resultant account of personal identity which embraces Wiggins's and goes beyond it, the details of which I shall save for the next and final chapter.

The extended mind thesis

The EMT holds that our cognitive processes and mental states 'ain't (all) in the head' (Clark and Chalmers, 1998). The demotic language here is an allusion to Putnam's externalism about meaning (1975, see p144), though Clark and Chalmers are citing this more in order to distance their *active* externalism from

what they consider the passive externalism of Putnam, and in so doing to clarify the EMT. To support his externalist claim, Putnam used the Twin Earth thought experiment in which an individual on Earth and a twin individual on Twin Earth would be in identical psychological states when thinking about a substance, which appeared identical and which they both called 'water', and yet was in fact H_2O on Earth and XYZ on Twin Earth. Putnam had thereby shown that the meanings of the twins' thoughts could not be found in their heads - the differences between the meaningful contents of their thoughts would not show up there. Clark and Chalmers call the external features of the twins' thoughts *passive*, because 'they play no role in driving the cognitive process in the here-and-now' (1998, p9). That there would be no difference in the twins' actions in relation to the different 'water's would be a reflection of this.

In contrast, Clark and Chalmers call the externalism of EMT 'active' because it posits a causal role being played by an external entity when coupled with a human organism in a cognitive system. Simple examples of this include the use of pen and paper to aid the performance of long multiplication, or the rearranging of letter tiles to aid word recall in Scrabble. Storing intermediate parts of the solution to a long multiplication sum on paper, before adding them together to reach a final answer, is an alternative to storing them in the brain, and there is no reason to see the former as any less a part of the cognitive process than the latter, or so Clark and Chalmers claim. They generalise this as the 'Parity Principle', the central claim of the EMT:

> If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. (1998, p8)

This in itself is likely to have implications for personal identity, if the cognitive systems, which in part we are, comprise more than is contained within the boundary of our skin (and not all agree that they do, as we shall see below when considering Rupert). That EMT has such implications might appear more likely when Clark and Chalmers apply it to *mental states* as well as to *cognitive processing*. They produce Otto and his notebook to show that minds extend into the world (and so Otto too, they believe, as we shall see). Otto's fellow Manhattanite, Inga, hears of an exhibition at the Museum of Modern Art and decides she wants to go. She thinks for a moment, recalls it is on 53rd Street, and sets off. Later, Otto, who has Alzheimer's disease, hears the same and decides he wants to go too. Unable to retain such information in his head, Otto consults the notebook he carries with him everywhere, in which he writes down new information as he learns it. He sees that the museum is on 53rd Street, and sets off. According to EMT, the notebook plays the same role for Otto as Inga's biological memory does for her, with, crucially, both Inga and Otto having the standing (non-occurrent) belief that the museum is on 53rd Street. Clark and Chalmers canvass a number of doubts that might be raised about whether what is in Otto's notebook should count as beliefs. As we are concerned with implications of EMT rather than its truth, we do not need to detail these, suffice to say that Clark and Chalmers conclude that there are only 'shallow' differences (1998, p16) between Inga's beliefs and Otto's notebook entries, which, crucially, play the same role as beliefs in causing Otto's actions. From a functionalist perspective, EMT appears to have a strong case in its favour.

Extended mind, extended self?

Let us now begin to consider what EMT might say about personal identity. An important question to consider is whether an extended mind entails an extended

self, on which there are different views. Clark and Chalmers believe that it does. To clear the way to showing this entailment, they first discount that the self could be restricted within the boundaries of consciousness (1998, p18). They had addressed earlier a potential worry about EMT, arising from the combination of a frequent identification of the cognitive with the conscious and the implausibility of consciousness extending outside the head (1998, p10). They pointed out that not all cognitive processes are conscious, and neither are all mental states, our standing beliefs for example, such as those in Otto's notebook, which they assert to be an important part of who we are. To reduce the mental only to conscious states would be 'to shrink the mind and self beyond recognition' (Clark, 2008, p160). Shapiro (2011, p198) is in agreement, saying that 'this view calls into question ordinary conceptions of psychological unity', which depend on the persistence of some non-occurrent states, otherwise, 'every morning one would awake as a new person' (2011, p199). There is an assumption here that some level of psychological continuity is a necessary condition for personal identity, which fits with the the thesis being developed in this dissertation, of supporting though extending Wiggins's human being account, given that psychological continuity is an essential part of the normal process of being and staying human. However, Brie Gertler (2007) claims to have shown that such shrinking of the mind must follow from accepting EMT. If this is the case, then EMT is not consistent with this account of personal identity. I shall first set out her argument, and then show where I believe it to be flawed.

Gertler suggests that if non-occurrent beliefs can be recorded externally, then so too can non-occurrent desires. She imagines that, rather than being in a notebook, non-occurrent beliefs and desires of Otto's are recorded onto an external computing device, for example, 'the desire to make banana bread on

Tuesday; the belief that banana bread requires bananas; the belief that the corner grocery store is a good source for bananas' (2007, p196). The device is also programmed with the cognitive abilities to devise action plans based on the stored beliefs and desires, and is inserted into an appropriately designed robot. Otto (or, at least, his organic part) goes to bed early on Monday, after which the robot heads off to the corner shop, and awakes late on Tuesday to the pleasant smell of freshly baked banana bread. Gertler asks who it was that made the bread, and argues that Clark and Chalmers must say it was Otto. For if his going to the museum is adequately explained by the non-occurrent belief in his notebook, then the bread-making can be adequately explained by his non-occurrent beliefs and desires, which have been programmed into another part of the extended system which makes up Otto. And if some of Otto's actions can be performed only by his organic part, which is surely the case, then by the parity principle, some could be performed by his non-organic part only. Passing over that this seems to extend the parity principle further than it was intended to go, let's see where extending her thought experiment a little further took Gertler. It should be possible to construct many robots in the same way, programme them all with Otto's non-occurrent beliefs and desires, and enable them to communicate with each other. Then Otto could do many things all at once - take a slow boat to China, ring all the doorbells in a Texas neighbourhood, perform in a karaoke contest in Tokyo - and he is not simply responsible for these actions but literally performing them, as long as his actions are 'the product of his standing states alone' (2007, p197). As Gertler sees this absurd conclusion arising from the EMT, and as she believes this thesis rests on a valid argument, she investigates its premises to discover where the problem lies. She believes she finds this in the premise that Otto's standing beliefs are part

of his mind, and so she rejects this, which in turn leads to the difficulty for psychological continuity already mentioned.

The flaw in Gertler's argument appears to be in her reading Clark and Chalmers to say that Otto's action in going to the museum was caused only by his non-occurrent states, in particular his belief about where the museum was. Yet they began the account of Otto's museum visit with 'Today, Otto hears about the exhibition at the Museum of Modern Art, and decides to go see it' (1998, p12), which clearly seems to show an occurrent desire. It was the combination of this occurrent state with the non-occurrent belief that led to Otto walking to 53rd Street. It is in fact hard to see that the EMT can be taken as saying that occurrent states are not needed to determine an agent's actions. The robot might have been programmed exactly as Gertler suggested, but if it did not have the ability on waking up that Tuesday to override the non-occurrent desire to make banana bread on Tuesdays with the occurrent desire to not bother that particular day, which Otto, however extended, surely would have, then it is hard to see that the action of baking can be ascribed to Otto. I think Gertler's objection can thus be discounted, and we still have the possibility that the extended mind, unconscious aspects and all, entails an extended self.

Olson (2011) does not believe that an extended mind entails an extended self, unless thinking things were literally bundles of mental states, which of course given his support for the animalist account of personal identity he does not believe to be the case. He makes his argument by responding to the questions of whether an extended self really does follow from the EMT and what it would mean if what he terms the extended-self thesis (EST) - the subjects of mental states which extend beyond their skin also extend beyond their skin - were true (2011, p481). He responds to the first question not by showing that EMT does not lead to EST,

but by suggesting one argument that it does and then knocking this argument down. Given the flimsiness of the argument he knocks down, and that he devotes most of his attention to his second question, I will concern myself here just with the latter.

He draws what he sees as a number of absurd conclusions from the supposition that EST is true, such as that none of us is an organism and, leading on from that, the 'too many thinkers' problem (Olson, 1997). But these conclusions are only absurd to someone, for example an animalist such as Olson, who equates person and organism (see chapter 2). To someone such as Baker (2000), who sees the organism as not identical to but constituting the person, they are not absurd at all. However, Olson claims that a further unpalatable consequence follows from combining our not being organisms with EST, which is that no organism could ever have mental properties. Olson labels this 'mind-life dualism'. He further concludes from this that 'psychological beings are located where, and only where, their mental states are', which he terms 'thinking-subject minimalism' (2011, p492), from which he concludes that either we are identical just to a part of our brains or nervous systems or that we are immaterial bundles of mental states. I find it impossible to follow his argument into 'thinking-subject minimalism', which thankfully will not matter here, as I am going to challenge him further upstream.

I will not challenge his reasoning which leads to his concluding that neither Otto nor any of us in fact are organisms (2011, p486), not least because I agree with the conclusion. I might just comment that it seems odd that Olson claims Otto not being a biological organism to be a 'momentous' *implication* of the extended-self thesis. Rather than this, it just seems to be a *statement of* this thesis. After all, Clark and Chalmers say themselves that Otto is 'best regarded as an extended

system, a coupling of biological organism and external resources' (1998, p18). Let us see then what Olson does in combining the conclusion that we are not organisms with the EST.

Olson calls the biological organism in Otto's story *O*. Note that I do not say 'the biological organism part of Otto' as this does not seem to be how Olson sees it. In fact, what Olson goes on to do is to talk about *O* as if it (he?) is a cognitive system in its (his?) own right, and he compares this cognitive system in a number of ways to the cognitive system that is Otto. So there appears to be a mereological confusion at the heart of Olson's argument. This confusion can be seen as Olson begins this section of his argument by stating that the beliefs in the notebook are Otto's but not *O*'s. In addition, *O* shares Otto's 'internal' mental states, as these all lie within *O*'s boundaries. Therefore 'Otto and *O* are psychologically identical apart from the beliefs in the notebook' (2011, p487). Olson then finds a contradiction to this in the fact that *O* uses the notebook just in the same way that Otto does and it has the same effect on *O*'s actions as it does on Otto's, suggesting that *O* comes to have the same beliefs as Otto when he consults the notebook. Olson concludes that the only way to avoid this contradiction and to save the EST is to see *O* as having no mental life at all, which implies that no human organism, or any biological organism at all, could have any mental properties.

It is hard to pick on one aspect of Olson's argument (either this section of it or the overall one in the paper) as the confusion I have alluded to runs throughout it. A clear manifestation of it in the part I have summarised is in Olson's statement that *O* uses the notebook exactly as Otto does. Let us follow Olson in separating out the parts of the cognitive system that is Otto (we can do so without abusing Olson's argument given that he is assuming the EST at this point), in which case we can replace 'Otto' with '*O and the notebook*'. Then we can see that Olson is

saying '*O* uses the notebook exactly as *O and the notebook* does'. I hope this shows the confusion clearly (it is hard to make sense of *O* and the notebook using the notebook) and my conclusion is that Olson has not shown that the extended self does not follow from the extended mind.

On the other hand, I am not sure that Clark and Chalmers have shown that it does follow - they are a little tentative in saying that the self '*may* also fall beyond the skin' (1998, p18, my emphasis). They may be following a 'natural default' position that self-follows-mind, arising from the dominant view that 'much of what matters to the identity of one's self is cognitive in nature' (Wilson and Foglia, 2011). Perhaps it is not possible to *show* whether or not the EMT entails an extended self, and the best way of addressing the question is to consider possible implications of accepting that the self is extended and make a judgement based on these. This sort of approach is taken by both Wilson (2004) and Clark (2008) though they reach differing conclusions. Wilson, while an extended mind sympathiser whose article on 'wide computationalism' (1994) was a forerunner of the thesis, takes a more conservative view about the extension of the self. His concern centres on notions of freedom and responsibility. For if a cognitive agent is spread into the world, then why, for example, should we punish just one part of the agent, the part within the body, when it commits a crime? He is also uneasy about the reconceptualisation of personal identity that the recasting of agency within an extended framework would require (Wilson and Foglia, 2011). Wiggins would surely join with this uneasiness, as might be seen from the note at the end of the previous chapter, in that he would worry about what extending the self would do to our human being concept. It may be that Clark's current position would reassure Wiggins (though I have my doubts), so let us turn to this now for the final part of this chapter. In the next chapter I will outline a reconceptualisation of personal

identity arising from an extension of the self about which I believe we can be positive rather than uneasy.

Clark and extending the human conception

Clark has developed his thinking on extended cognition considerably since his 1998 paper with David Chalmers, and this development is characterised in this comment:

> It is our biological nature... to be open to many forms of physical and cognitive hybridization. Some of these… may be so intimate as to effectively extend the thinking agent. All of them are crucial parts of the nested, iterated and ongoing process of cognitive self-re-creation that is the characteristic mark of human intelligence. (2005, p9)

I have chosen to quote this partly to offer some immediate reassurance to Wiggins. Clark believes it is in 'our biological nature' to extend ourselves by recruiting parts of the world to join our biological organisms in our cognizing. He suggests this recruitment is governed by the 'Principle of Ecological Assembly', according to which 'the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort' (2008, p13). These resources might include 'memory glasses' to help Alzheimer's patients recognise faces, silicon substitutes to replace an impaired neural resource (suggested by the example of a Californian spiny lobster) (Clark, 2005), or, more prosaically, pen and paper, notebooks, calculators and phones. Crucially for Clark's position, as we shall see shortly, the brain plays the lead role in the recruitment drive, in which it is 'cognitively impartial' (2008, p118-22). For example, it will store information externally as well as internally, and use gestures and movement as well as introspection to do its cognizing. It 'simply uses whatever it can… to get the job done' (p122).

We can see how Clark's views on the implications of an extended thinking agent differ to Wilson's by considering his responses to certain objections to extended cognition (Rupert, 2004). Renaming the extended mind thesis as the Hypothesis of Extended Cognition (HEC), Rupert sets against it the Hypothesis of Embedded Cognition (HEMC):

> Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place. (2004, p393)

Rupert challenges supporters of extended cognition as to why they would adopt HEC when HEMC does the same work of recognising the importance of the environment in cognition, while appealing more to common sense and being the less radical of the two, as accepting HEC would 'significantly change our conception of persons' (2004, p390). This slightly begs the question here, for Clark's view is that we do need to alter this conception so that it fits more with how our minds actually work. Nonetheless, Rupert believes that he has thus placed the burden of proof onto HEC. Secondly, Rupert is concerned about the scientific costs of the acceptance of HEC, believing it would rob cognitive science of its 'traditional target of… a suite of integrated, persisting, organismically grounded capacities' (Clark, 2008, p113), presenting it only with couplings that are too ad hoc and temporary for serious scientific study.

Clark's response to the first challenge leads him to suggest another hypothesis, HOC, the Hypothesis of Organism-Centred Cognition:

> Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism the brain/CNS) remains the core and currently the most active

element. Cognition is organism centered even when it is not organism bound. (2008, p139)

So, while Otto's notebook is a part of the extended system which comprises Otto, it is his organism (or *O*, to use Olson's term), or perhaps his brain, which is the 'persisting core' (2008, pp116-8). Clark offers this hypothesis to guard against the idea that the extension of cognition beyond the human organism means that it is no longer 'centered' in the organism. This does suggest that if extending cognition requires a new conception of persons, then the difference might be more modest than first thought. However, that a different conception *is* needed is suggested by Clark's otherwise robust defence of HEC.

In particular, Clark argues that it is HEMC which has the greater scientific costs, and that HEC holds out the greater promise of a science attuned to how organisms and the environment are actually coupled. Taking, for example, the cognitive system comprising 'pen, paper, graphics programs, and a trained mathematical brain' (2008, p116), Clark questions how far the behaviour of this system can be understood by examining its component parts separately before putting them back together. While accepting that the scientific study of such hybrid systems is in its infancy, Clark makes the case for the need for such study, and points to its possibility in the work of Wayne Gray and his colleagues (Clark, 2008, pp118-21): 'the first stirrings of a science… targeting genuinely hybrid ensembles: soft-assembled coalitions comprising biostorage, motoric and perceptual modes of access and bioexternal storage' (p121). So seeing persons not simply as *using* the environment around them in ways not previously considered by traditional cognitive science, but as *being actually spread into* the environment, could encourage the type of scientific study required to fully understand the extended systems that, on Clark's view at least, we are.

Finally in this chapter, let me return to Wilson, who views cognition as extended but not the self, and consider a suggestion he makes for how to understand the subjects of extended cognition, and how Clark counters this. Wilson suggests that the agent can be identified as 'the locus of control' for a given, possibly extended, cognitive system, the locus of control being 'housed in the agent's body' (Wilson, 2004, cited by Wilson and Foglia, 2011). This might sound initially to be saying the same thing as Clark when he talks of the brain as the 'persisting core'. However, the special role that Clark was ascribing to the brain was in *recruiting* the resources to be used by the extended cognitive system. Once they are in place, 'it is the flow and transformation of information in… an extended, distributed system that provide the machinery of ongoing thought and reason' (Clark, 2008, p122). And the idea that there is a locus of control within the body, presumably in the brain, of the agent would set us on an inward journey for which it would be hard to find the eventual destination. Clark points out that the brain itself is not 'scientifically unified' and asks what, therefore, would happen if we applied 'the locus of control criterion *inside the head*. Do we now count as *not part of my mind or myself* any neural subsystems that are not the ultimate arbiters of action and choice?' (2008, p160). The conclusion would be a similar shrinking of the mind and self apparent in Gertler's ideas noted above, or even, if no final locus of control were discovered, that the self would disappear altogether.

However, Clark argues that even if there *were* such a point of final control, there would be no more reason to identify the agent with it than with 'the whole body of memories, skills and dispositional beliefs that guide, shape and characterize my behaviors' (2008, p160). And if these memories and beliefs can be found outside the human body, then perhaps that is where to look too for the person, rather than only inwards. And if this is the case, then perhaps we should

consider whether the conditions of personal identity need also to be extended. I believe that Wiggins is right about the need to consider both psychological and bodily continuity, given a conception of human being as necessarily having both mental and physical properties. If Clark is also right, then, in order to do full justice to a conception of humanness in which parts of our environment are combined with the biological parts of ourselves to form extended selves, perhaps we need to consider forms of environmental continuity too.

CHAPTER 4

ENVIRONMENTAL AND SOCIAL CONTINUITY

Introduction

Let us take stock about the point at which we have now arrived. Having considered some of the main accounts of personal identity (chapter 1), we then introduced the idea of embodied cognition (EC), and looked at some implications it might hold for these accounts (chapter 2). There we saw that, while most embodied cognition research programmes would only go as far as suggesting that a person would need the same *kind of body* to persist, a condition which might in any case be conceded by psychological continuity advocates, some programmes coming under the 'constitution hypothesis' (Shapiro, 2011) appeared to offer support for the human being account favoured by David Wiggins (2001). Ideas of embodied cognition support the necessity of the physical and the mental continuing together, as an integrated whole, reflecting our concept of human being. In particular, this supported Wiggins's revisionism concerning the Brown-Brownson thought experiment, which made up part of his case against the neo-Lockeans (Parfit, 1984; Shoemaker, 1984) and for his own account. We then introduced the extended mind thesis (chapter 3), also seen (by Shapiro) as in the constitution group of embodied cognition ideas. A case was made for the extended mind entailing an extended self, or, at least, cases were made *against* challenges to Clark's assertion of such an entailment.

This has brought us to the point of tentatively suggesting that the extended mind thesis, affecting our conception of persons as it appears it might, may also require an augmenting of accounts of personal identity. Having suggested that embodied cognition supports the necessity of psychological *and* bodily continuity, we are going to suggest in this final chapter that extended cognition adds the need

for some level of environmental continuity[13], and that this might include some form of social continuity. One effect of this will be to challenge the ratio-centricity and individualism found within most accounts of personal identity (Wilson and Lenart, forthcoming). Locke transposed into his account the ratio-centricity of the Aristotelian tradition, in which humans were distinguished from other animals and life-forms by uniquely having a *rational soul*, and Locke's emphasis on a memory criterion added to the account's individualist bias. This was because the 'rational cognitive capacities' required for memory were assumed to depend solely on internal aspects of the person (ibid.). These biases continue not only in neo-Lockean accounts, but are implicit too in biological accounts, given the conception of humans as rational beings. Such biases can 'imply that individuals with certain cognitive limitations cannot claim the right-conferring status of personhood' (ibid.), so there may be an ethical bonus for favouring accounts which allow for the enhancement of cognitive capacity via external resources.

However, we do not need to think in terms of separate continuities, of mind, body and environment. Indeed, Wiggins himself does not talk of psychological continuity and bodily continuity, but of the person continuing as a human being, which necessarily has both mental and physical properties joined together. In the previous chapter we talked of *systems*, in which biological organisms are *coupled* with aspects of their environment to form extended selves. Gibbs has recently articulated a conception of personhood, 'as an emergent property of interactions of the brain, body and world', according to which 'our sense of who we are as individual persons… require(s) special attention to these couplings, not just to brain, bodies or world as separate entities' (2006, p41). Given that an individual person's world includes other people, we could say that Wiggins, in his

---

[13] Actually, as will be seen, at least one other form of embodied cognition might also add such a need, but the implications in this chapter arise mainly from the extended mind.

appropriation of Locke's concept of person as a forensic term (see chapter 2), paid attention to one such coupling, which we will return to when talking about social continuity.

The *interactions* of human beings with their environment are also central in one account of how personhood first emerged, which I am going to present next. If the environment was an essential constituent of personhood from the outset, then this adds to the case for its continuing presence being necessary for personal identity through time.

Becoming persons

Clark's reasons for favouring suitably augmented accounts of personal identity are probably less ethically based than simply reflecting the conception of the person that he believes we should hold. That is, as 'natural-born cyborgs' (2003), we do not simply use the environment, it is a part of us. The case for this is strengthened, it seems to me, by a compelling story told by the archeologist, Steven Mithen, in his 2004 review of Clark's 2003 book. If not all animals are persons, then there must have been a time in their evolutionary history before which humans were not persons. So how was the transition into personhood made? Daniel Dennett cautioned not to ask: 'it is a fool's errand to try to identify a first or most simple instance of the "real" thing' (1998, p362). However, Mithen's account of the role played by a particular coupling between humans and their environment seems both believable and important in what it says about the place of the environment in personhood, so let me recount it here.

Mithen begins by crediting a significant change in his thinking about cognitive evolution to his reading of Clark's and Chalmers' early work on the extended mind (Clark, 1997; Clark and Chalmers, 1998), which led him to the appreciation that 'the mind was as much a product of human culture as culture

was a product of the mind' (Mithen, 2004, p164-5). He believes that this 'remarkably simple idea' might help to explain the period of rapid 'cognitive inflation' (Baker, 2009, p652) which took place in human culture about 60,000 to 30,000 years ago, the 'Upper Palaeolithic' revolution that saw the first appearance of art and ritual, and significant technological developments. Clark and Dennett have speculated that it was the development of language that enabled the emergence of personhood (Clark, 2002). However, Mithen makes a different suggestion. First he gently takes Clark to task for conflating speech and text (Mithen, 2004, p166), pointing out that spoken language probably originated 500,000 years ago, whereas writing is a relative newcomer at around 3000 BC. So neither of these timescales can explain what happened during the Upper Palaeolithic period. What did begin to appear closer to that time was art, as pictures on cave walls or as carvings. According to Dennett, '(t)he aspirant to a high order of self-control must have the capacity to represent his current beliefs, desires, intentions and policies *in a detached way, as objects for evaluation*' (1984, p86, my emphasis). Speech alone, while enabling thoughts to be shared with others, is too fleeting to have enabled such evaluation, whereas the material existence of art would have served to capture and to freeze thoughts and hence to enable a 'second-order cognitive dynamics' (Clark, 2003, p79). This is the dynamics of an active externalism (see chapter 3), where thoughts, captured and represented as art, can become part of extended cognitive systems, thereby contributing to the creation of further thoughts, capturable in the same way, and so on in continuous cognitive loops. So, Mithen concludes, '(i)t was with the origin of art rather than speech or written text that ''the floodgates of self-reflec- tive reason'' were opened' (2004, p167).

Environmental continuity

If Mithen is right, then, as Clark and Chalmers (1998, p11) say, 'it may be that the biological brain has in fact evolved and matured in ways which factor in the reliable presence of a manipulable external environment'. It is the idea of a 'reliable presence' of an environment that I now wish to consider, as this suggests that environmental continuity may need to be considered for personal identity.

Before focusing on the extended mind, let us turn back first to the embodied account of sensory experience encountered in chapter 2 (O'Regan and Noë, 2001). It will be remembered that this was based on the idea that, for example, visual experience resulted from the knowledge of 'sensorimotor contingencies' and that these contingencies related to both the perceiver's visual apparatus and to the external objects of perception. O'Regan and Noë (2001, p943) suggest that vision requires not only knowledge of the laws of sensorimotor contingencies, but that the perceiver is 'tuned' to these laws and hence able to actively exercise its mastery of them. For example, a missile guidance system (that works) is tuned to the sensorimotor contingencies that govern airplane tracking: as a missile zigzags to evade enemy fire, the image of the target plane moves to the right or left; as it slows down, the image appears larger, and the system must be able to interpret and adapt to these changes to continue tracking the plane. Now, O'Regan and Noë note that this idea of being tuned 'only makes sense within the context of the behavior and purpose of the system or individual *in its habitual setting*' (2001, p943, my emphasis). If the same missile guidance system were to be used as a funfair attraction, it would probably be adjusted for thrills and spills purposes, to suddenly lunge, accelerate and decelerate, behaviours that would be avoided when used for its proper purpose. The adjustment would require mastery of new sensorimotor contingency laws. In chapter 2 it was claimed that this account

suggested that personal identity required continuity of an actual body, given that a person is in part constituted by their sensory experiences and that these depend on mastery of the sensorimotor contingencies of the person's sensory apparatus. It can now be seen that a complementary environmental continuity might also be implied, given the need for mastery of the sensorimotor contingencies of those aspects of the person's environment being sensed.

To move now to implications of the extended mind for environmental continuity, let us consider a variation of Parfit's teletransportation thought experiment. Imagine that Otto steps in to the teletransporter, on his (or at least, his replica's) way to Mars, clutching his notebook (it goes everywhere with him). The addition of the notebook should not cause any difficulty given the obvious sophistication of the machine. If it can record every cell of the biological part of Otto, then recording his notebook will be a breeze. So when the replica of Otto arrives on Mars, he (it?) will have an exact replica of Otto's notebook with him. Though it will not be of any immediate use, he will be able to write new information in it, and also advise any of Otto's friends making contact from Earth to ask for the whereabouts of that restaurant they both liked and so on (thanks to Mandrigin, 2007, for this appealing idea).

Earlier in this chapter I mentioned the possibility of an individualist bias within neo-Lockean (among other) accounts of personal identity. Parfit's use of thought experiments such as teletransportation might seem to confirm this bias, as Parfit is only concerned about the continuity of internal psychological states in the replica of the person who stepped into the teletransporter. Yet, if he were confronted with the possibility of someone like Otto, who carried his non-occurrent beliefs externally, Parfit could point to the above scenario to show that this could be accommodated quite easily. But Otto and his notebook was simply one thought

experiment designed to demonstrate at least the logical possibility of an extended mind. If there is reason to suppose that this is possible then there is no reason to suppose that the external storage and apparatus that is part of an extended cognitive system would have to be limited in size to a notebook. And if we consider Clark's ideas of cognitive systems being temporary hybrids including whichever external resources the brain recruits to best undertake a cognitive task, then perhaps we are led to the need for a teletransporter which is able to transmit any sized and any indefinite number of potential external parts of a cognitive agent. As it would not be known in advance what these parts might be, perhaps the only way to be sure that psychological continuity would be maintained would be to transport every aspect of the person's environment, that is, to transport Earth onto Mars.

The absurdity of the idea, even within a thought experiment that is fantastical at the outset, might be seen to reflect the lack of attention given to the environment in thought experiments used to bolster traditional accounts of personal identity, and to suggest the profound part that one's environment can play in maintaining one's sense of self. By the same token, could it not be the case that at least some of the oddness in the thought experiment whereby present-day Charles wakes up one day, apparently with the memories and character of Guy Fawkes (Williams, 1956-7), might also arise out of the environmental disjunction between the two? Williams uses this case to support his arguments for the necessity of bodily continuity and that there are difficulties inherent in the idea of separating bodily and psychological features, which are all of a piece with the thesis of this dissertation. I would just add that there is a difficulty in separating out environmental features too, and, at the risk of being overly literal, suggest that perhaps another difficulty in Charles being Guy Fawkes was that their environments were 350 years apart and thus too different.

Turning from outlandish puzzle cases to more real life ideas, Wilson and Lenart (forthcoming) have written recently on extended mind and personal identity for a professional audience in the field of neuroethics. They focus on the role that the environment can play in the continuing memory and cognitive functioning of people with Alzheimer's disease, and give an example of Dennett's (1996, p134-9), who describes elderly people becoming incapable of recalling simple daily routines when admitted to institutions. However, on returning to the familiar environments of their homes where they had 'off-loaded' routine schedules, for example concerning taking medication, onto certain objects or places, many signs of dementia disappeared. Similarly, Lindemann suggests that for dementia sufferers, '(f)amiliar places and things, beloved objects, pets, cherished rituals, one's own bed or favorite shirt, can and do help... to maintain (a) sense of self' (2010, p162).

Consistent with psychological continuity accounts is the idea that personhood diminishes, perhaps to zero, as a human being enters a persistent vegetative state. The argument I have been putting forward in this section could be summarised in this question: Could personhood also diminish, perhaps to zero, as a human being is gradually shorn of his or her environment? And in the case of a person whose cognitive capacity is becoming progressively impaired, could personhood be retained by appropriate attention being given to that person's environment? Much of what Wilson and Lenart focus on is that part of a person's environment that consists of other people, and that there is a possible requirement for social continuity for personal identity will be the final part of my argument.

Social continuity

Impairment of cognitive capacity can also be compensated through reliance on other people, as trusted others can 'pick up some of the ensuing cognitive

slack' (Wilson, 2014, p23). However, it is not only in cases of cognitive dysfunction where other people play an important role. Lindemann (2010) argues that a person's identity is preserved by the complex interactions between families and other groups, and the interconnected narrative memories transmitted within them. At this point one might start to think of collective memory and group minds, and Wilson and Lenart make the interesting suggestion that what might have been thought of in those terms might be 'more plausibly viewed as cases in which the extended cognition of the individual involves a social environment involving other people' (forthcoming, p?). They invoke the *social manifestation thesis*: 'the idea that individuals engage in some forms of cognition only insofar as they constitute part of a social group' (Wilson, 2005, p229), and conclude that this offers 'an expanded role for the extended mind thesis', and one that 'contains implications for personal identity' (Wilson and Lenart, forthcoming, p?).

What is being suggested here is that personhood requires the presence of other people, both for its development perhaps - would the apes who brought up Tarzan have been able to induct him into personhood? - and in its persistence across time. As Wilson and Lenart say, '(s)elves are a product of both individual and communal processes, and thus personhood cannot be defined in solely individualistic terms' (ibid.). The implications for personal identity which Wilson and Lenart draw differ somewhat from those favoured here under the influence of Andy Clark in particular, and I shall return to this shortly. Before this, let us make a link back to Wiggins, and his reconsideration of the Brown-Brownson case.

It will be recalled from chapter 2 that in this reconsideration Wiggins made use of ('misappropriated' to use Wiggins's own expression) Locke's idea that *person* is a 'forensic term' (Wiggins, 2001, p234). The use that Wiggins made of this was to suggest that *person* (or *human being*, given that he identifies the two)

must have a public use, and he alluded to its uses in the 'interpersonal sphere' that Strawson (1962) drew our attention to, 'when he showed how intimately our ideas of agency and responsibility depend on human beings' reactive and participative attitudes towards other human beings' (Wiggins, 2001, p234). The conclusion that Wiggins draws from this concerns the importance of the person's 'human presence', which brings physical attributes to the fore, and which elicits others' reactive attitudes. But 'reactive attitudes' do not only require the human presence of the person whose persistence as a person is in question, but the presence of those others who are reacting to that human presence. And this returns us to Gibbs's stipulation mentioned towards the beginning of this chapter, that we need to pay attention to the couplings, in this case between ourselves and others, in order to gain and retain 'our sense of who we are as individual persons' (2006, p41).

Wilson and Lenart agree about the need to attend to these couplings, between the cognitive agent and its environment, though they resist the idea that the environment, including other people or their mental states, become part of that agent. The idea they suggest is that it is identity that is extended and not the person. In suggesting this, they describe an individual's personal identity as being an important *property* of that person, and that this is extended in the sense of not being determined solely by the person's internal features. This accounts for their belief that their account of 'extended personal identity' can 'loosen' the ratio-centricity of other accounts. As we saw in chapter 1, however, identity is a *relation* rather than a *property*, so there might be a little confusion here. It should be said that many people would side with Wilson and Lenart over Clark, in that it can sound far-fetched to suggest that a person could be constituted in part by another person's mental states. It may be, however, that one does not need to come down

on one side or the other, to be convinced that some form of social continuity - in the sense of having the continuing existence of other people in one's life - might be required for the persistence of personhood and to be the same person at one time as another.

Perhaps the tentativeness of the suggestion here might be reflected by ending this section, as the previous one, with a question. If, in a persistent vegetative state, the organism continues and the person does not, what about in a persistent isolated state? Could personhood also diminish, perhaps to zero, as a function of isolation from other people? Do we need other people to enable us to become and to persist as persons?

Concluding remarks

We have argued that embodied cognition, and the extended mind thesis in particular, suggest that accounts of personal identity have been too reliant on individualistic notions of personhood. Both psychological and physical accounts have focused us too much on what happens within the brain and body of the person, and so have led us to miss the essential contributions made to our becoming persons by social and environmental factors. And if these factors have been essential to our becoming persons, then by the same token they are essential in our remaining the same persons over time. So, to psychological and somatic, bodily continuity criteria of personal identity we have added social and environmental continuity. And here I use the word 'added' advisedly. Thinking in this area has been predominantly either-or, where theorists, in a form of Cartesian hangover and supported by those neuroscientists who see cognition as a disembodied process taking place entirely in the brain, have been able to imagine a person's psychological aspects coming away from the actual body in which they developed. However, it appears to make sense to take a both-and approach, one

probably always favoured by Williams, later to be joined by Wiggins, which sees a person's body not just as a vessel within which the essential psychology can be developed, but as an integral part of that psychology, not just showing it and communicating it but in a very real sense constituting and being it.

This approach fits with Strawson's concept of a person as *primitively* being a thing with two aspects, bodily and mental, and is supported by some of the work being carried out within the embodied cognition programme, notably that by O'Regan and Noë on sensory experience and consciousness. This suggests we are necessarily made up of these particular bodies, with our particular mental states and ways of being and acting. These will grow and change and eventually come to an end, according to processes that unfold in ways that are particular to our being human. And being human, we have made connections with our environment that have helped us to become the persons we are, and as we continue to make more use of the environment around us, are forever becoming.

REFERENCES

Baker, L. R. (2000), *Persons and Bodies: A constitution view*. Cambridge: Cambridge University Press.

Baker, L. R. (2009). Persons and the extended-mind thesis, *Zygon*, 44, 3, 642-658.

Butler, J. (1736). Of personal identity, in *The Analogy of Religion*; reprinted in J. Perry (2008), pp99–105; page references are to the reprinted version.

Clark, A. (1997). *Being There: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, A, (2002). That special something: Dennett on the making of minds and selves, in A. Brook and D. Ross (eds.). *Daniel Dennett*. New York, NY: Cambridge University Press, (pp187-205).

Clark, A. (2003). *Natural-Born Cyborgs: Minds, technologies and the future of human intelligence*. New York, NY: Oxford University Press.

Clark, A. (2005). Intrinsic content, active memory and the extended mind, *Analysis*, 65, 1, 1-11.

Clark, A. (2008). *Supersizing The Mind: Embodiment, action and cognitive extension*. Oxford: Oxford University Press.

Clark, A. and Chalmers, D. (1998). The extended mind, *Analysis*, 58, 1, 7-19.

Dennett, D. (1984). *Elbow Room*. Cambridge, MA: MIT Press.

Dennett, D. (1996). *Kinds of Minds: Toward an understanding of consciousness*. New York, NY: Basic Books.

Dennett, D. (1998). *Brainchildren: Essays on Designing Minds*. Cambridge, MA: MIT Press.

Dewey, J. (1925). *Experience and Nature*. New York, NY: Dover.

Gallois, A. (2012). Identity over time, in E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*. URL = http://plato.stanford.edu/archives/win2010/entries/identity-time/

Gertler, B. (2007). Overextending the mind?, in B. Gertler and L. Shapiro (eds.). *Arguing about the mind*. New York, NY: Routledge, (pp.192-206).

Gibbs, R. (2006). *Embodiment and Cognitive Science*, Cambridge: Cambridge University Press.

Gleick, J. (1993). *Genius: The life and times of Richard Feynman*. New York, NY: Vintage.

Johnston, M. (1987). Human beings, *Journal of Philosophy*, 84, 2, 59-83.

Korfmacher, C. (2006). Personal identity, in *Internet Encyclopedia of Philosophy: A peer-reviewed academic resource*. Accessed 7.08.14. http://www.iep.utm.edu/person-i/#SH2e

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lindemann, H. (2010). Holding one another (well, wrongly, clumsily) in a time of dementia, in E. F. Kittay and L. Carlson (eds.). *Cognitive Disability and its Challenge to Moral Philosophy*. Malden, MA: Wiley-Blackwell, (pp161-168).

Locke, J. (1694). Of identity and diversity, in *Essay Concerning Human Understanding (2nd edn.)*, ch27; reprinted in J. Perry (2008), pp.33-52; page references are to the reprinted version.

Mackie, J. (1980). The transcendental 'I', in Z. van Straaten (ed.). *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Clarendon Press, (pp48-61).

Mandrigin, A. (2007). *Personal Identity, Psychological Continuity and Externalism*, unpublished masters dissertation, University of Edinburgh.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*, trans. C. Smith (first published 1945). London: Routledge.

Mithen, S. (2004). Review Symposium: We have always been… cyborgs; on Andy Clark's Natural-Born Cyborgs: Reviews by Terry Dartnall, Adrian Mackenzie, Alicia Juarrero, and Steven Mithen; Reply by Andy Clark, *Metascience*, 13, 139–81.

Nagel, T. (1984). Are you your brain?, paper delivered to Princeton Philosophy Colloquium and APA Pacific Division, cited in Johnston, M (1987) op.cit.

Nimbalkar N. (2011). John Locke on personal identity, in A.R. Singh and S.A. Singh (eds.). *Brain, Mind and Consciousness: An International, Interdisciplinary Perspective*, *MSM*, 9, 1, pp268-275.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Olson, E. (1997). *The Human Animal: Personal identity without psychology*. Oxford: Oxford University Press.

Olson, E. (2006). Is there a bodily criterion of personal identity? in F. MacBride (ed.). *Identity and Modality*. Oxford: Oxford University Press, (pp.242-259).

Olson, E. (2010). Personal identity, in E. N. Zalta (ed.). T*he Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*.
URL = http://plato.stanford.edu/archives/win2010/entries/identity-personal/

Olson, E. (2011). The extended self, *Minds and Machines*, 21, 481–495.

O'Regan, J.K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences*, 24, 939–1031.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Perry, J. (ed.). (2008). *Personal Identity (2nd edn.)*. Berkeley, CA: University of California Press.

Putnam, H. (1975). The meaning of 'meaning', in K. Gunderson (ed.). *Language, Mind and Knowledge*. Minneapolis, MN: University of Minnesota Press, (pp131-193).

Rhodes, K.M. (2011). *Personal Identity and the Extended Mind*. Unpublished doctoral dissertation. Irvine, CA: University of California, Irvine.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition, *Journal of Philosophy*, 101, 8, 389-428.

Shapiro, L. (2011). *Embodied Cognition*. London: Routledge.

Shoemaker, S. (1963). *Self-knowledge and Self-identity*. Ithaca, NY: Cornell University Press.

Shoemaker, S. (1970). Persons and their pasts, *American Philosophical Quarterly*, 7, 269-285.

Shoemaker, S. (1984). Personal identity: A materialist's account, in S. Shoemaker and R. Swinburne (eds.). *Personal identity*. Oxford: Basil Blackwell.

Snowdon, P. (2009). Peter Frederick Strawson, in E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*.
URL = http://plato.stanford.edu/archives/fall2009/entries/strawson/

Strawson, P. (1959). *Individuals: An essay in descriptive metaphysics*. London: Methuen.

Strawson, P. (1962). Freedom and resentment, *Proceedings of the British Academy*, 48, 1–25; reprinted in P. Strawson (1974). *Freedom and Resentment and Other Essays*. London: Methuen.

Swinburne, R. (1973-4). Personal identity, *Proceedings of the Aristotelian Society*, 74, 231-247.

Van Gelder, T. (1995). What might cognition be, if not computation, *Journal of Philosophy*, 92, 345-81.

Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.

Wiggins, D. (1976). Locke, Butler and the stream of consciousness: And men as a natural kind, *Philosophy*, 51, 196, 131-158.

Wiggins, D. (1980). *Sameness and Substance*. New York, NY: Oxford University Press.

Wiggins, D. (1998). *Needs, Values, Truth (3rd edn.)*. Oxford: Oxford University Press.

Wiggins, D. (2001). *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Williams, B. (1956-7). Personal identity and individuation, *Proceedings of the Aristotelian Society*, New Series, 57, 229-252; reprinted in B. Williams (1973), pp1-18; page references are to the reprinted version.

Williams, B. (1970a). The self and the future, *Philosophical Review*, 79; reprinted in B. Williams (1973), pp46-63; page references are to the reprinted version.

Williams, B. (1970b). Are persons bodies? in Spicker (ed.). *The Philosophy of the Body*. Chicago, Il: Quadrant Books; reprinted in B. Williams (1973), pp64-81; page references are to the reprinted version.

Williams, B. (1973). *Problems of the Self*. Cambridge: Cambridge University Press.

Wilson, R. (2004). *Boundaries of the Mind: The individual in the fragile sciences*. New York, NY: Cambridge University Press.

Wilson, R. (2005). Collective memory, group minds, and the extended mind thesis, *Cognitive Processing*, 6, 227-236.

Wilson, R. (2014). Ten questions concerning extended cognition, *Philosophical Psychology*, 27, 1,19-33.

Wilson, R. and Foglia, L. (2011). Embodied cognition, in E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*. URL =http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition/

Wilson, R. and Lenart, B. (forthcoming). Extended mind and identity, in J. Clausen and N. Levy (eds.). *Handbook of Neuroethics*. New York, NY: Springer.

Wittgenstein, L. (1953). *Philosophical Investigations*. trans. G.E.M. Anscombe. Oxford: Blackwell.