Available online at www.sciencedirect.com
ScienceDirect
Consciousness and Cognition 16 (2007) 533–550
Consciousness and Cognition
www.elsevier.com/locate/concog

A model of the synchronic self

Glenn Carruthers

Discipline of Philosophy, Adelaide University, Adelaide, SA 5005, Australia

Received 21 March 2006
Available online 17 July 2006

Abstract

The phenomenology of the self includes the sense of control over one's body and mind, of being bounded in body and mind, of having perspective from within one's body and mind and of being extended in time. I argue that this phenomenology is to be accounted for by a set of five dissociable cognitive capacities that compose the self. The focus of this paper is on the four capacities that compose the synchronic self: the agentive$_B$ self, which underlies the sense of control over one's body; the boundary$_B$ self, which underlies the sense of being bounded within one's body; the agentive$_M$ self which underlies the sense of control over one's thoughts; and the boundary$_M$ self which underlies the sense of being bounded within one's mind. I model the agentive$_B$ and agentive$_M$ selves as parts of the motor control system and the boundary$_B$ self as the capacity to form and integrated map of the body. I point to the delusion of thought broadcast as a possible source of evidence for future research on the boundary$_M$ self.

Keywords: Self; Phenomenology of self; Agency; Boundedness; Ownership; Delusions; Body integrity identity disorder; Motor control; Integrated body map

1. Introduction

Who am I? Who are you? Why do I think that I am not you? Perhaps a more fundamental question is: why do I think that I am anyone at all? The natural answer to these questions is that I have a self of which I am aware. Because I am aware of this self it seems to me that I am a person. My self is not your self so I feel that I am me and not you. Essentially I have a sense of being me. The problem is that this natural answer just raises further questions. What am I aware of when I am aware of my self? How do I come to be aware of it?

Currently it is not clear what, if anything, the self is. Recently Metzinger (2003) proposed that there is no such thing as the self, merely phenomenal models of the self. I am not, yet, inclined to agree. I share with Metzinger the idea that there is a rich phenomenology of the self. Our minds must somehow generate this

E-mail address: glenn.carruthers@adelaide.edu.au.

1053-8100/$ - see front matter
doi:10.1016/j.concog.2006.06.002

phenomenology. It seems sensible to suppose that the self is what ever underlies this phenomenology.[1] It is the substantive claim of this paper that it is a set of cognitive capacities that underlie this phenomenology. To model these capacities is to model the self. It may turn out that once we have explained he phenomenology of the self in terms of these capacities that there is nothing left over to call the self. If this is the case, then it would seem we have explained away the self. Before I can answer this question must first model the self.

In order to model these capacities I will consider recent empirical work on the self. I will begin by examining the phenomenology of the self. That is, the sense of self we all share. Whilst none of the phenomenology I describe is new it is important to be detailed at this point as there is variance in how certain terms (e.g. ownership) are used in the literature. Alternative uses of the terms describing the phenomenology of the self are in footnotes. The remainder of this paper will be an attempt to explain how this phenomenology arises. First, by considering a study conducted by Daniel Povinelli, I will show that there are two dissociable capacities that underlie the phenomenology of the self. I will call these the diachronic self and the synchronic self. By considering phenomena such as the delusion of thought insertion I will show, in turn, that the synchronic self is composed of four dissociable capacities. Two of these make up what I will call the boundary self; the remaining two compose what will call the agentive self. Having made these distinctions I will go on to identify what computational systems compose these capacities. I will argue that the agentive self as it relates to the body is the capacity to represent a similarity between the predicted and actual sensory consequences of a movement. Following this I will argue that the boundary self as it relates to the body is nothing more than the capacity to form an integrated map of the body. Next I will argue that the agentive self as it relates to the mind is also a part of the motor control system. Namely, it is the capacity to compare the goal state of a movement to the predicted sensory consequences of that movement. As for the final aspect of the synchronic self, the boundary self as it relates to the mind, I do not yet have a proposal as to what computational systems compose this capacity. However, I point to the delusion of thought broadcast as a probable source of evidence. Please note that this model is designed to account for certain feelings, which will be discussed below. It is not an account of our knowledge of ourselves.

2. The phenomenology of self

We all share a rich phenomenology of self. Look around you now. If your reading environment is anything like mine, you will be aware of a chair, a desk and the humming of a computer. You will no doubt be aware that these objects are not you. Again if your reading environment is anything like mine you are about to be interrupted. When the next person enters the room I'm sure it will not surprise you that they are a distinct person from you. You experience objects and other persons differently than how you experience yourself. This is perhaps the most obvious aspect of the phenomenology of the self: the experience we all (we healthy adult human beings) share of being distinct entities from other objects and persons (Eilan, Marcel, & Bermudez, 1995, p. 7; Legerstee, 1998, p. 625). You are able to represent this distinctness in two ways. First you represent things as either part of your body or not. You consciously track certain bodily events such as the sense that certain visceral events are inside of you (Dennett, 1991, p. 428). Pay attention to the movement of your stomach. Where is it? Inside the boundary of your body, of course! Second you can represent things as part of or occurring in your mind or not.[2] Think of your favourite meal. Where does this thought occur if not in your mind? These experiences constitute a sense of *boundedness*. These experiences manifest in many ways. For

[1] Metzinger may respond to this that of course it seems that way, but that this merely due to the transparency of the self model. If there is a real difference between my proposal and Metzinger's (and not merely a terminological difference i.e. an argument about what deserves the name 'the self' [note that I am not committed to the view that there is a deeper difference than this]) then ultimately there will be testable differences between my proposal and his. Before we have any hope of discovering what the difference between the two proposals is I must provide a model. That is the aim of this paper.

[2] Stephens and Graham approach a similar point. They describe the sense of subjectivity, this being the sense that "I am the *subject* in whom, or in whose psychological history, the thought occurs" (Stephens & Graham, 2000, p. 7). Later we find a clearer description of the sense of subjectivity as the sense that a "thought occurs in me" (Stephens & Graham, 2000, p. 119). However, I wish to depart from their description somewhat as the phrase "psychological history" distracts from the point that this is a part of the phenomenology arising from the synchronic self (see below).

example, they allow you to distinguish your body and your mind from the rest of the world. I will call these the 'in my body/not in my body' and the 'in my mind/not in my mind' distinctions, respectively.

Focus on this page and I mean really focus. You will notice something odd about the page. It is a solid object in three-dimensional space, and yet you cannot observe it all simultaneously. The explanation for this is obvious. You can only observe the page from one perspective. Which perspective? Yours! Being a distinct entity you have your own perspective of experience (Damasio, 1994, p. 238). There are many aspects to this experience of perspective. Your experience of your body for example includes knowledge that you are inside your body (Martin, 1995, p. 267). Furthermore, you are well aware that external influences impact on your body at a particular place (Damasio, 1994, p. 233; Martin, 1995, p. 267). If you were to drop a brick on your foot it would not hurt the top of your head, nor would it be a diffuse experience of pain. The pain would be in your foot.

Underline this sentence. Now, who was in control whilst you did that? You were. Beyond simply being bounded within the body, it also feels as if you *control* your body (de Vignemont & Fourneret, 2004, p. 2; Dennett, 1989, p. 1; Dennett, 1991, p. 419). However, this is not just the feeling that the body is being controlled; it is also the feeling that *you* are in control of your body (Strawson, 1997, p. 408). We might call these feelings of control the sense of *agency*.[3] This feeling is the sense that you are an agent working with and against the world (Eilan et al., 1995, p. 10). This sense of agency extends beyond your experiences of your body into two further realms: these are the experience of being in control of your own thoughts (Campbell, 2002; Gallagher, 2000b, p. 203; Proust, 2003; Stephens & Graham, 2000, p. 9) and the experience of being in control of your social actions (Legerstee, 1998, p. 633).

The feeling of perspective is greater than suggested above. It is not just that there is experience from *a* perspective. It is experience from *your* perspective; from the perspective of your self (Dennett, 1991, p. 412). This is part of our sense of ownership. Think of your mother's face. Now, in line with what we discovered above, you will be aware that this memory occurs within your mind and is generated by you (Campbell, 2002; Gallagher, 2000b, pp. 203–204). Low and behold the memory is yours! In addition to the sense of boundedness and of agency there is also the sense of *ownership*.[4] For instance, as a distinct mental entity you possess your own thoughts that are not observable by others (Strawson, 1997, p. 407). Your boundaries and the thoughts and experiences within those boundaries belong to you. At least it seems that way.

So far all the feelings discussed have been experiences of the self as it exists at a particular time. These experiences do not exhaust the phenomenology of the self. You also experience yourself as existing as *extended in time* (Gallagher, 2000a). Think back to when you picked up this text. Who did this happen to? Again, you won't be surprised that the answer is you. Many writers have suggested that the phenomenology of self has a diachronic character. That is, it feels as though the self is extended in time (Dennett, 1991, pp. 419 and 423; Strawson, 1997, p. 408). Some describe the experience as the feeling that one is a persisting object (Eilan et al., 1995, p. 3), whilst others place less emphasis on these feelings (Strawson, 1997, pp. 419–421 & 423). It is reasonable to suggest that even those who do not strongly experience themselves as the same object over time do have some sense of internal causal connectedness to their past self, some sense that they came from that self (Campbell, 1995, p. 36). Furthermore, you experience yourself as having a particular autobiography (Dennett, 1992, p. 7; Gallagher, 2000a). You do not just experience the fact that you existed in the past, you experience, for example, that you went to the zoo, you laughed at the monkeys, you enjoyed an ice cream et cetera. That is you remember being a self (in the synchronic sense) and doing things as that self. The diachronic feelings of self extend further to the feeling that you will exist in the future (Damasio, 1994, p. 239). *You* will wake up tomorrow morning. Furthermore, it appears that you can plan behaviour based on these feelings. You make plans for *you* to go to work tomorrow, not someone else.

[3] Metzinger (2003) uses the term agency somewhat differently to what I and others [e.g. (Campbell, 2002; Eilan et al., 1995; Stephens & Graham, 2000)] do. I hold that the sense of agency is the sense of being in control of one's thoughts or actions. In contrast Metzinger has a more exclusive notion of agency as involving a "sense of effort" (Metzinger, 2003, p. 391). This more exclusive notion creates problems for understanding thought insertion see Footnote 4.

[4] Note that I am using the term 'ownership' slightly differently to how others use it. For example Stephens and Graham (2000, p. 125) take the view that what I am calling the sense of boundedness is sufficient for the sense of ownership to arise. Following Campbell (2002) I hold that sense of both boundedness and agency are required for the sense of ownership to arise. See also Footnote 6.

By answering a set of simple questions we are able to show that there are several robust senses of self present in experience. These are the senses of boundedness, ownership, agency and of being extended in time. It is the substantive claim of this paper that the self is the set of cognitive capacities that generates the different aspects of the phenomenology discussed above.

In what follows I will use the terms 'the self' and 'the sense of self'. To avoid confusion consider the phenomenology discussed above as the *sense of self*. On the other hand *the self* refers to whatever underlies these senses of self. I propose that it is a set of cognitive capacities that does so. In the following section I will draw some robust distinctions within the phenomena of the self. These distinctions will allow me to home in on more precise notions of the self, which I will model below. It is my aim that this model will explain the phenomenology discussed here.

3. A conceptual framework for understanding the self

The word 'self' has been used to describe many distinct phenomena and concepts. Rather than examine all of these uses (there are too many) it will be helpful to look for some principled distinctions that will assist in clarifying the phenomena of interest. The first distinction will be between the diachronic and synchronic selves. Next I will distinguish between two aspects of the synchronic self, namely the boundary and the agentive self. I will also distinguish between the mental and bodily aspects of the agentive and boundary selves.

3.1. The diachronic vs. the synchronic self

In the literature it is common to distinguish between the diachronic and synchronic aspects of self. This is usually done merely by introspection. As we saw above the feeling that you are you now seems distinct from the feeling that you are the same person you were ten minutes, a day, a week or ten years ago (Gallagher, 2000a; Hume, 1739, p. 3; James, 1890, p. 4; Povinelli, 2001, p. 77; Strawson, 1997, p. 408). However, introspection is not enough to support the claim that these different feelings arise from dissociable capacities. Povinelli (2001) has recently provided evidence which supports these introspective claims. His series of studies suggests that the diachronic and the synchronic aspects of the self arise at different stages in development. If it is the case that these two aspects of self are developmentally dissociated then they are distinct.

Povinelli (2001) investigated how children aged 2, 3 and 4 years reacted to delayed videos of themselves. The children are filmed playing a game with a familiar adult. At one point in the game the children are rewarded for their efforts with a friendly touch. During this gesture the adult secretly places a sticker on the forehead of the child. Three minutes after this the child is shown the video of the game. On the video it is clear that a sticker is being placed on the child's head. When asked who is in the video all the children recognise themselves. Importantly on seeing the tape 75% of the 4 year olds immediately reach up to remove the sticker. On the other hand, only 25% of the 3 year olds and none of the 2 year olds do so. However, all of the children, regardless of age, reach up to remove the sticker when they observe themselves in a mirror. The difference between each scenario (tape vs. mirror) is that a mirror gives immediate feedback as to one's physical appearance, whereas the video gives delayed feedback.

To explain this difference in behaviour we must invoke two different selves. It is not enough to suppose that older children have better capacity to remember the past alone. They must recognise that the self they are now is the same self they were in the past (Proust, 2003). To be motivated to remove the sticker when looking in a mirror subjects need only have a *sense* of the synchronic self. That is, they need only the synchronic self that underlies this sense. To be motivated to remove the sticker when observing the video, subjects must not only have the synchronic self, they must also recognise themselves *as extended in time*. Thus, they must also have a diachronic[5] self that underlies this sense. The capacity to generate the sense of oneself as extended in time develops after the capacity to generate the sense of oneself at a particular moment. Two year olds have the

[5] Note that in this sense the diachronic self is a cognitive capacity that (ultimately) explains how feelings of being extended in time arise. This does not have any direct bearing on the issue of personal identity overtime.

synchronic self, where as most four year olds have both the synchronic and diachronic selves. The synchronic and diachronic selves have distinct ontogenetic trajectories, thus they are two distinct capacities.

Having drawn this distinction we are now able to see links to the phenomenology discussed above. The sense of the synchronic self is the moment-by-moment feeling we all have of being a distinct entity. Beyond this, our sense of control over our own thoughts and actions also arises from the synchronic self. I will have more to say on what aspects of our phenomenology are generated by the synchronic self after I have examined this aspect of the self in more detail. Diachronic self, on the other hand, underlies feelings of temporal extension, the feeling that we have of being the same person over time. For those who do not strongly experience these feelings, diachronic self seems most strongly related to the feeling of some weak causal connection to one's past self. The remainder of this paper will focus on the synchronic self.

3.2. Boundedness, agency and the sense of ownership

Within the synchronic self we can distinguish between the *boundary* and the *agentive selves*. The boundary self is the capacity to recognise oneself as a distinct entity. Important to this is the sense that one's mind is a distinct part of the world. The agentive self, on the other hand, is the capacity to recognise one's actions and thoughts as under one's own control.

Like the diachronic and the synchronic selves, the boundary and the agentive selves are empirically dissociable. Consider the phenomenon of the ownership of thoughts. Campbell (2002) has proposed that there are two 'strands' to the ownership of thoughts.[6] The first strand being that the thought must appear to occur within my boundaries.[7] The thought must appear to be 'in me'; specifically it must be in my mind. The second strand is that it must feel as though I, as an agent, generated the thought (Campbell, 2002, p. 36). If I generated a thought in my mind then that thought is mine. If a sense of ownership of thoughts is lost it seems likely that this is due to a loss of either the sense of boundedness or of agency.

We see a loss in the sense of ownership of thoughts in some cases of schizophrenia. This loss occurs in those patients who suffer from the delusion of thought insertion. The delusion involves variations on the belief that someone or something is placing thoughts in one's mind that are not one's own (Mellor, 1970; O'Brien & Opie, 2000, p. 1). The patient is well aware that the apparently inserted thoughts are occurring in their own mind. Thus, it seems they have a *sense* of their mind being a distinct part of the world. Therefore, their boundary self, which underlies this sense, is in place (O'Brien & Opie, 2000, p. 1). What the patient lacks is a sense that they generated the thought. Their *sense* of control is absent[8] (Campbell, 1999; Campbell, 2002). Therefore their agentive self, that underlies these feelings, is absent (O'Brien & Opie, 2000, p. 2). Thus, it appears possible to lose the capacity to generate the sense of agency, whilst maintaining the capacity to generate the sense that one's mind is a distinct part of the world. This selective loss suggests that we should treat the boundary and the agentive selves as distinct.

Similar disorders show us that there are also two aspects of the agentive self that are dissociable. In particular these disorders suggest that the sense of control over one's *bodily actions* can be lost independently of the sense of control over one's *mind*. Consider the delusion of alien control. Subjects suffering the delusion of alien control believe that someone or something other than themselves is causing them to move (Frith,

[6] Gallagher (2000b) also distinguishes agency and ownership. However, the idea that there are two strands of ownership is not part of his analysis. We will see that the idea that both the sense of boundedness and of agency contribute to the sense of ownership is useful for analysing the delusion of thought insertion. Thus, I will use Campbell's analysis as the canonical version. Furthermore, even though the terms the diachronic self and the synchronic self correspond very closely to Gallagher's 'extended self' and 'narrative self' (Gallagher, 2000a) I have chosen to depart from his terminology so as not to imply his distinction between agency and ownership (which is a distinction within the minimal self).

[7] Obviously the boundaries of the mind do not seem to the subject to be spacial boundaries.

[8] We can now see why Metzinger's more exclusive notion of agency as involving a sense of effort is problematic. If we used this notion of agency it would be difficult to account for thought insertion. Think of yourself sitting on a bus riding to work. You are bored so you let your thoughts wonder. Suddenly you realise that you are thinking about someone you met in your first year of university. Certainly you did not go to any *effort* to think of this person. But such experiences do not give rise to delusions of thought insertion. Thus if the sense of agency required a feeling of effort, it could not be a loss of this sense that accounted for delusions of thought insertion. Thus, I prefer the more liberal notion of the sense of agency.

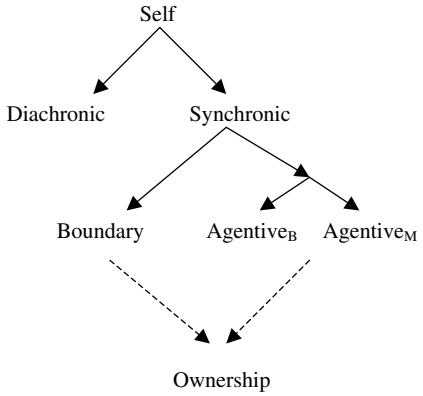

Fig. 1. Distinctions so far.

Blakemore, & Wolpert, 2000, p. 1784). Again, they know that the movement is being performed by their body, hence their boundary self is in place. However, they do not *feel* that they have any control over the movement. Thus there is a deficit in their agentive self.

Importantly subjects can present with delusions of control without delusions of thought insertion. Thus, the sense of control over one's body can be lost independently of the sense of control over one's mind. To explain this we must postulate two independent capacities that give rise to these feelings. I will call these the *agentive$_B$ self* and the *agentive$_M$ self*. The agentive$_B$ self underlies the sense of control over one's body, whereas the agentive$_M$ self underlies the sense of control over one's thoughts. Below we will see that there are good reasons to suppose that a similar distinction holds for the boundary self. In later sections I will identify two aspects of the motor control system that may compose the agentive$_M$ and the agentive$_B$ selves.

We can now see how the phenomenology above relates to the different aspects of the synchronic self. Boundary self underlies the sense we share of being distinct entities. Furthermore it gives rise to the sense that we are bounded in both body and mind. The boundary self also gives rise to the feeling that we have a perspective from inside these boundaries. The agentive self underlies feelings of control over one's behaviour (agentive$_B$) and one's thoughts (agentive$_M$). The sense of ownership arises from both the boundary and the agentive self. All of these experiences are of the self as it exists at a single moment. Hence, they fall under the title of the synchronic self in this framework.

So far I have argued that the self is composed of a number of dissociable capacities. All of the distinctions I have argued for are based on the fact that the different senses of self are empirically dissociable. Underlying each of these *senses* of self is a distinct cognitive capacity—a *distinct self*. These distinctions are presented in Fig. 1.

3.3. The self and the mirror

By considering what is required to pass the mirror self recognition test we will see that the distinction between bodily and mental aspects also holds for the boundary self. The mirror self recognition test is the most common test for the self applied to animals.[9] In this test, first proposed by Gordon Gallup (1970), a caged subject is presented with a mirror. Members of most species initially react to the image as if it is another animal (Gallup, 1970, p. 86). Over a few days of exposure the animal's social responses rapidly decline. Members

[9] Some, for example an anomous reviewer of this paper, may argue that mirror self recognition is a highly specific task that is not analogous to other forms of bodily recognition. The motivation for this claim would be that the body in the mirror is recognised visually whereas the body is normally recognised via other means, for example by proprioceptive and kinaesthetic feedback. Povinelli (2001) argues that it is by recognising the equivalence between these kinds of representations of the body and the visual representations of the body that one is able to recognise oneself in the mirror. As I argue below this task actually involves two different selves. However, we will see that proprioceptive and kinaesthetic representations of the body are central to the construction of these selves. Thus, we see that whilst mirror self recognition is a specific task it is absolutely dependent on the kinds of self recognition we are interested in here.

of some species replace these responses with self directed behaviour, such as picking food from between teeth and exploration of parts of the body that are not visible without the mirror. On the basis of these observations Gallup proposed a more rigorous test for self recognition. After some initial exposure to the mirror, the subject is anesthetised and marked with a red dye on their forehead. Once the animal is awake it is exposed to a mirror again. An animal is deemed to pass the mirror self recognition test if there is a significant increase in mark directed behaviour coincident with the animal observing itself in the mirror (Barth, Jochen, Povinelli, & Cant, 2005, p. 4; Gallup, 1970, p. 87).[10]

What exactly is the mirror self recognition test a test for? As Gallup pointed out, most animals initially act toward the mirror as though it were another animal, suggesting that they do not initially know what the mirror image is (Gallup, 1970, p. 86). To pass the test the subject must recognise that the mirror image is a representation of their body. Povinelli proposes that to do this the subject needs to recognise an equivalence relationship. It is on the basis of recognising that "what is true of my body is true of the image. Therefore they must be equivalent" that the subject is able to recognise themselves in the mirror (Povinelli, 2001, p. 84). We need not suppose that this reasoning occurs in language, but as a metaphor this is a clear enough description.

To discover what is required to recognise this equivalence relationship we must first draw a distinction within the boundary self. The image in the mirror needs to be recognised as an image of the body. This is the same notion of the body the subject uses in the 'in my body/not in my body' distinction. That is, the image must be recognised as an image of the body in which one is bounded. This does not require there to be any representation of the mind. To pass the mirror self recognition test one *does not* need to use the 'in my mind/not in my mind' distinction. So it seems that the senses of boundedness are independent. Thus the boundary self divides into two aspects, just as the agentive self did. Only one of these aspects is involved in recognising oneself in the mirror. Call these aspects of the self the boundary self as it relates to the body, or the *boundary*$_B$ *self* and the boundary self as it relates to the mind, or the *boundary*$_M$*self*. The boundary$_{\mathrm{M}}$ self is the capacity that gives rise to the sense of being bounded within one's mind, to recognise that one's mind is distinct from the rest of the world and that one has perspective from within it. Whereas the boundary$_{\mathrm{B}}$ self gives rise to the sense of being a bounded body and that one's body is distinct from other objects in world.[11]

I propose that the subject must make use of both the boundary$_{\mathrm{B}}$ and the agentive$_{\mathrm{B}}$ selves in order to recognise the equivalence between their body and the mirror image. The image of the body in the mirror can then be recognised as an image of one's own body by comparing the action of the body to that of the image (Povinelli, 2001, p. 85). For this to be possible the actions of one's body must first be recognised as one's own. As we saw above this requires *both* the boundary and the agentive selves. In particular it requires a sense of control over one's actions, so it requires the agentive$_{\mathrm{B}}$ self. The subject must also have a sense that the movement is a movement of their body. As we have just seen this requires the boundary$_{\mathrm{B}}$ self. To recognise oneself in a mirror one must generate a motion of the body and recognise that the image makes the same movement. The movement of the body is recognised as belonging to oneself as it is a motion of the body and it is generated by oneself. That is, a sense of ownership emerges for the action due to the boundary$_{\mathrm{B}}$ and the agentive$_{\mathrm{B}}$ selves. This motion is compared to the motion of the image in the mirror. On recognising that the motions are equivalent the subject is able to recognise that the image is an image of the body in which they are bounded. Thus, we can conclude that the mirror self recognition test shows the presence of *both* the boundary$_{\mathrm{B}}$ and the agentive$_{\mathrm{B}}$ selves. The final framework for conceptualising the self is depicted in Fig. 2.

4. The self in the brain

In this section I will explore the computational capacities that give rise to the various senses of self identified above. In other words, I will attempt to discover the cognitive capacities that compose the self. In

[10] 'Mark directed behaviour' refers to behaviours such as rubbing the dyed mark with the hands or otherwise exploring the mark.

[11] Povinelli and colleagues have also argued that the ability to 'conceptualise' the body evolved prior to the ability to 'conceptualise' the mind (Barth et al., 2005; Povinelli & Cant, 1995). With some adaptation, this constitutes stronger evidence than that provided here for the dissociation of the boundary$_{\mathrm{B}}$ and the boundary$_{\mathrm{M}}$ selves. To keep this paper to a readable size I have omitted this argument. However, this argument based on the functions of these selves is a sufficient (if superficial) way to make the point.

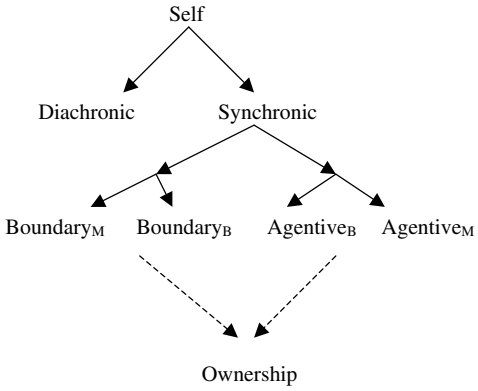

Fig. 2. A final framework. (Some readers will find the idea that the self is a set of dissociable cognitive capacities as at odds with the idea of the unity of the self. (Thanks to an anonymous reviewer for pointing this out.) This is not so. Unfortunately arguing this point is way beyond the scope of this paper (and may indeed be beyond the scope of a single paper). Here I can only give a brief discussion of this point. Whilst the self certainly appears in consciousness as a unified thing, we need not suppose that this means there is a single mechanism underlying the sense of self. The reason for this has to do with the multi-track nature of consciousness in general (O'Brien & Opie, 1999, 2000). The multi-track theory of consciousness provides a model of how experience arising in different parts of the brain or due to different cognitive mechanisms can be unified into a single experience. I cannot provide a detailed account of this model here, so a simple metaphor will have to suffice. When recording two or more instruments we can unify the recordings into a single piece of music in one of two ways. One can record the instruments together onto a single track. Alternatively the instruments can be recorded individually onto separate tracks, which are then played back simultaneously. (O'Brien & Opie, 2000). The multi-track theory of consciousness holds that consciousness is more analogous to the second approach than the first. On this approach to explain the apparent unity of consciousness and the apparent unity of the self one needs an account of why occurring at the same time is sufficient for two conscious experiences arising somewhat independently nevertheless form a unified experience.)

contemporary cognitive science it is orthodox to hold that cognitive capacities are computational capacities. Like cognitive capacities generally, those that compose the self are computational capacities. Vital to understanding any computational capacity is an understanding of the representations it operates over (Clark, 2001, p. 17; O'Brien & Opie, forthcoming, p. 4). Here, I will examine what representations and computational systems compose the self. Recall that the target of explanation for this paper is the phenomenology of the self. The following model is designed to account for certain feelings, such as the feeling of control. It is by no means an account of what is known. The delusion of alien control can be understood as a deficit in the agentive$_B$ self. By considering this disorder in greater detail I will argue that the agentive$_B$ self is a part of the motor control system. I will argue that the boundary$_B$ self is a very different capacity, namely the capacity to form an integrated representation of the body. By considering the delusion of thought insertion I argue that the agentive$_M$ self, like the agentive$_B$ self, is a part of the motor control system. I will conclude by considering the boundary$_M$ self. Currently I have no proposal as to what computational system composes the boundary$_M$ self. However, I point to the delusion of the broadcast of thoughts as a possible source of evidence for the future development of a model.

4.1. The agentive$_B$ self

The agentive$_B$ self gives rise to the sense of control over one's actions. Above we saw that people suffering from delusions of alien control can lose this sense of control. Patients suffering this kind of delusion believe that someone or something else is in control of their actions:

> When I reach my hand for the comb it is my hand and arm which move, and my fingers pick up the pen, but I don't control them... I sit there watching them move, and they are quite independent, what they do is nothing to do with me... I am just a puppet who is manipulated by cosmic strings. When the strings are pulled my body moves and I cannot prevent it (Mellor, 1970, p. 18).

This patient attributes their action to an external agent, one who manipulates them via "cosmic strings". Importantly, there is no denial of *ownership* of the body. The patient knows full well that it is their hand, arm and fingers that are moving. What they do deny is they initiated their action. It is the phenomenology associated with agentive$_B$ self, not boundary$_B$ self, that is lost. What is damaged in the case of delusions of alien control?

Christopher Frith has proposed that the loss of the sense of control over one's actions can be understood as a deficit in the motor control system (Frith, 1987; Frith & Done, 1989; Frith, Rees, & Friston, 1998, 2000). He presents a model of the computational function of the motor control system "based on established models of normal motor learning and control" (Frith et al., 2000, p. 1771). Frith uses this model to explain a variety of motor disorders including the delusion of alien control.

Any cognitive system that navigates the world faces the problem of generating movement appropriate to its environment. To do this there must be some co-ordination between sensory information and motor commands. The motor control system solves the problem of monitoring and controlling the relationship between motor commands and the sensory information resulting from movement (Frith et al., 2000, p. 1771).

The coordination between sensory information and motor commands is achieved using a system of five different representations. The system begins with a representation of a particular target state for the body; the goal of the movement, if you like (1 in Fig. 3). The system uses a second type of representation to specify how the movement is to be made. These are the motor commands. Frith proposes that these motor commands arise from a specialised computational system he calls the 'controller'. Such commands cause the muscles to contract so as to move the body in the desired way (Frith et al., 2000, p. 1773). Two copies of every motor command are made. One of these copies is sent to the periphery and causes the muscles to contract in the desired way (2a in Fig. 3). The second copy (also called the efferent copy) remains in the brain (2b in Fig. 3). Using this copy of the motor commands, and a variety of other information (such as proprioceptive information), the motor control system is able to predict the state of the body after the movement has been performed. This, the third type of representation, is about the predicted sensory consequences of the movement (3 in Fig. 3). Frith proposes that this computation is performed by a second specialised computational system he calls the 'predictor'. The predictor makes use of the second copy of the motor commands (Frith

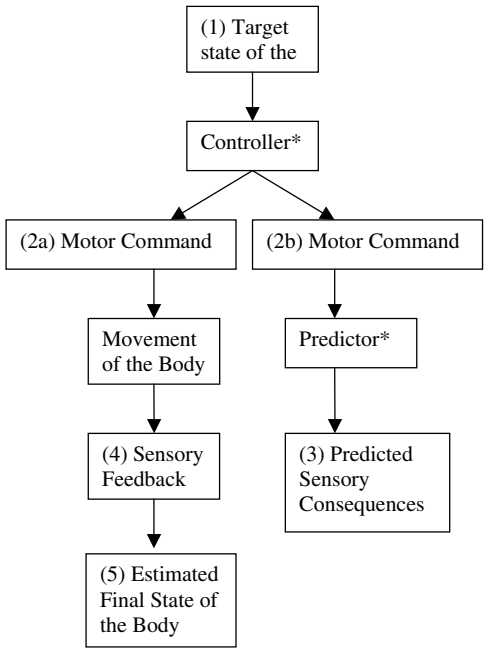

Fig. 3. The series of events involved in a simple movement as described by Frith. The numbered boxes indicate particular representations described above. The boxes marked with an * indicate particular computational systems proposed by Frith.

et al., 2000, p. 1772). The fourth type of representation used by the motor control system is about any actual sensory feed back that arises due to the movement (4 in Fig. 3). Such sensory feedback may be visual, tactile, auditory or proprioceptive (Frith et al., 2000, p. 1773; Sherwood, 2001, pp. 140–141 & 175). Using this sensory feedback the motor control system can represent the final state of the body (Frith et al., 2000, p. 1773). This is the fifth type of representation used by the system (5 in Fig. 3). Fig. 3 depicts the series of events that make up a simple movement.

On Frith's model, the motor control system is able to compare some of these representations to one another, toward a variety of ends. Important for my purposes is the comparison between the predicted sensory consequences and any actual sensory consequences of a movement. One purpose of this comparison is to allow the creature to correct mistakes in movement (Frith et al., 1998, p. 173). I will argue that it is the capacity to represent the similarity between the actual and predicted sensory consequences of a movement that underlies the sense of control over one's body.

Frith and Done (1989) provide evidence that those with delusions of alien control have a deficit somewhere in the motor control system. They compared the performance of subjects suffering delusions of control to control subjects on a computer game. In this game subjects can choose to fire a gun either left or right as a target appears. To fire the gun left subjects move the joystick left, to fire right they move the joystick right. With the knowledge of the subject this situation is altered, such that the gun will fire in the opposite direction to the direction the joystick is moved. This ensures a high rate of error as subjects easily loose track of which kind of trial they are up to. From the time it is fired it takes 2.8 s for the bullet to reach the target. Should the subject fire the bullet in the wrong direction they can correct this error at any time during this period by moving the joystick in the opposite direction (Frith & Done, 1989, pp. 360–361). In the test of most relevance to the agentive$_B$ self, the path of the bullet is obscured for the first 2.0 s of its flight. This prevents subjects from using *visual* feedback to correct any errors during this time. If they are to correct errors in this time they must know what trial they are up to and which movement they performed. This information must be obtained via non-visual (e.g. proprioceptive) feedback. The subject must then use this knowledge to determine whether they performed the wrong movement (Frith & Done, 1989, p. 361). Frith and Done found that there was no difference between control subjects and those suffering delusions of control in terms of the number of errors made and the number of errors corrected. However, they did find that subjects suffering delusions of control only corrected their errors in the last 0.8 s of the bullet's flight. That is, they only corrected errors when the direction the bullet was moving became visible. Control subjects, on the other hand, corrected their errors within the first 2.0 s of the bullet's flight, before the trajectory was visible (Frith & Done, 1989, p. 361). This suggests that those suffering delusions of control have difficulty representing which action they have performed in the absence of visual feedback (Frith & Done, 1989, p. 361).

Frith and Done's interpretation of these results is controversial. Kopp and Rist criticise this kind of interpretation based on their finding that whilst schizophrenic patients were slower at correcting errors they did not do so less often than controls (Kopp & Rist, 1994). However, as discussed above, Frith and Done also found that there was no difference in the number of corrections made between the alien control and the control groups. What matters is the basis on which the corrections are made. Furthermore, even if Kopp and Rist's data contradicted Frith and Done's we could not be certain that this finding impacts on Frith and Done's interpretation as they did not limit their schizophrenic group to those with delusions of control. Furthermore, it is not clear whether or not those in the schizophrenic group were experiencing delusions at the time of the test or not. If the subjects were not suffering delusions of alien control at the time of the test we would have no reason to suppose that the purported deficit in the motor control system is present.

Here, it is important to emphasise that it is only in contrived experimental procedures that those with delusions of control have problems with correcting errors of movement. In day-to-day life this problem is not pronounced. Patients with delusions of control can, for example, drive or get food to their mouth without a problem.[12] What would account for this difference? Frith and Done's study suggests an answer. In day to day situations the patients can use *visual* feedback as a basis to correct errors of movement. However, in Frith and Done's study patients cannot use visual feedback to correct errors. Perhaps those with delusions of

[12] Thanks to Philip Gerrans for these examples.

control lack access to *non-visual* types of sensory feedback. It would appear that they especially lack access to proprioceptive and kinaesthetic feedback. This conclusion is supported by studies that suggest that schizophrenic patients are worse at detecting that visual feedback of their action has been distorted than controls (Jeannerod et al., 2003).

We now have two deficits associated with delusions of control. First there is the lack of a sense of agency as it relates to movement. Second there is the inability to know what movement has been performed in the absence of visual feedback. Following Frith and colleagues (2000), I propose that both these deficits can be explained by the same underlying disorder: a failure to properly represent the comparison between predicted and actual sensory consequences of a movement.

Frith and colleagues discuss evidence that the representation of the predicted sensory consequences contributes significantly to awareness for healthy people. They cite a study by Libet and colleagues that suggests that the awareness of initiating a movement *precedes* the movement by 50–80 ms (Frith et al., 2000, Libet, Gleason, Wright, & Pearl, 1776; Libet et al., 1983). In the model of the motor control system above there are three types of representation that are present *before* movement occurs. These are the target state, the motor commands and the predicted sensory consequences of the movement. It appears that *at least one* of these must contribute significantly to our awareness in order to give rise to the sense of initiating a movement. Merely having a target state should not give rise to the sense of actually starting the movement, as one can know where one wishes to move before one actually does so. Thus, the choice is between the motor commands and predicted sensory consequences of the movement.

Frith and colleagues discuss evidence that the motor commands themselves do not make up a significant part of our awareness. Consider the Ebbinghaus circles illusion (Fig. 4). In this illusion two circles of equal size appear to be different sizes due to contextual information. When asked to grasp the circles as though to lift them, subjects move toward the same target state for either circle. That is the motor commands that are issued to grasp either circle are toward a similar target state, even though subjects maintain that the circles look to be different sizes (Aglioti, DeSouza, & Goodale, 1995; Frith et al., 2000, p. 1774). Frith argues that this evidence suggests that the motor commands do not contribute significantly to our awareness. If the results of such experiments generalise to all situations, and the motor commands do not contribute significantly to awareness, then it must be the case that it is the predicted sensory consequences of a movement that contribute significantly to awareness (Frith et al., 2000, p. 1774).

Unfortunately there is a flaw in this reasoning. We saw above that the predicted sensory consequences of a movement are generated based on motor commands. As the motor commands to grasp either circle are similar, the predicted sensory consequences are also similar. In this case the relevant sensory consequences are proprioceptive. If a subject is aware of the representation of the predicted proprioceptive consequences they should be able to tell that the circles are in fact the same size. So if the results from the Ebbinghaus circle

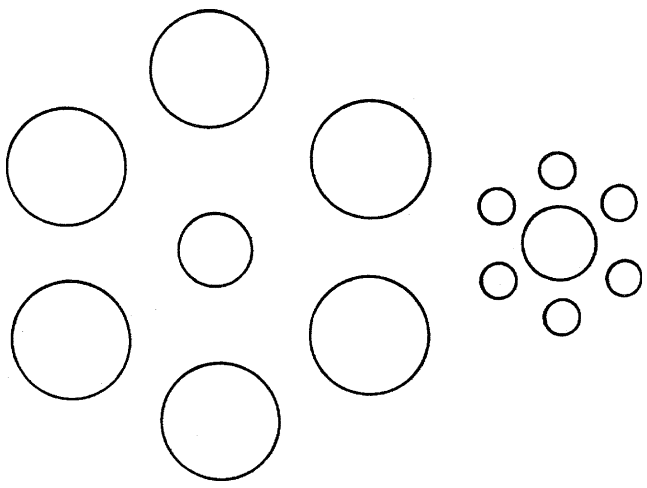

Fig. 4. The Ebbinghaus circles illusion. Despite appearances to the contrary, the circles in the middle of each ring are the same size.

illusion generalise it seems that *neither* the motor commands nor the predicted proprioceptive consequences contribute significantly to our awareness. One possible response to this problem is to deny that these results generalise. This may be the case, but I know of no evidence that suggests this is true. Even if this is the case, these data still must be explained. The other possibility is to suggest that the formation of either a motor command or predicted proprioceptive consequences allows the subject to represent that a movement has been performed, but they are not sufficient for the subject to know exactly which movement has been performed. If we accept this proposal then it seems we have a problem in explaining how healthy subjects are able to correct their movements in the absence of visual feedback in tests such as Frith and Done's above. What other than the motor command and the predicted proprioceptive consequences could allow those subjects to recognise that they performed the *wrong* movement?

The answer is obvious; we can suppose that subjects come to know they have performed the wrong action based on actual proprioceptive feedback. The comparison of the actual proprioceptive feedback from the predicted proprioceptive feedback indicates an error. The subject must represent this information, or else they would not be able to correct their errors. So we can suppose that those with delusions of control need visual feedback to correct their errors as there is a problem with the representation of the comparison between the actual the predicted *proprioceptive* feedback (Frith et al., 1998, p. 172; Frith et al., 2000, p. 1784). Hence, they need to use visual feedback to compensate for the information they lack.

Could a failure to properly represent the comparison between the actual and predicted proprioceptive feedback explain the delusion of alien control? In line with Frith[13] I believe it can. As we saw above the comparison between the predicted sensory consequences and the actual sensory consequences allows a subject to represent that they initiated a movement. Thus, if a subject cannot properly represent this comparison there will be no way to generate the experience of having initiated the movement (Frith et al., 2000, p. 1784). Thus in delusions of control, the damage to the agentive$_B$ self is ultimately a deficit in the motor control system.

I propose that the cognitive capacity to compare the predicted proprioceptive consequences of a movement to the actual proprioceptive consequences is important to the agentive$_B$ self. Specifically, one experiences a sense of control when there is a negligible difference between the representation of the predicted sensory consequences and the actual sensory consequences of a movement. Hence, the agentive$_B$ self is the capacity to represent this similarity.

4.2. The boundary$_B$ self

The boundary$_B$ self involves the sense of being distinct from other objects. It is the capacity to represent the 'in my body/not in my body' distinction. This includes the sense of having a perspective from inside the body. As we have just seen, somatosensory and proprioceptive representations play a role in the agentive$_B$ self, by providing information that the motor control system can use to predict the sensory consequences of a movement. We will see that this information plays an even more vital role in the construction of the boundary$_B$ self.

Consider body integrity identity disorder (First, 2005). Patients with this disorder form a strong desire to amputate one or more limbs. Such people explain this by stating that the limb is not 'really' a part of them (First, 2005, p. 919). They seem to have a problem forming the 'in my body/not in my body' distinction, erroneously experiencing some parts of their body as not truly their own or as a needless addition to their body. One subject reported that she had to focus on remaining "connected" when her legs (which she desired to amputate) were touched (First, 2005, p. 926). She called this "doing body work" (First, 2005, p. 926). Occasionally sufferers of this disorder are able to obtain an amputation. After surgery patients describe themselves as feeling "whole" or "paradoxically...more complete" (First, 2005, pp. 919–920). These appear to be cases where the sense of a limb being part of the body is somehow diminished. This apparently leads to a weak form of denial of ownership. This is a weak form of denial as those suffering body integrity identity disorder know in some sense that the limb is their limb, but they deny that it is 'truly' their limb (First, 2005, p. 922). There is no consensus on an explanation of this disorder. Body integrity identity disorder is only recently being studied

[13] This is not exactly Frith's proposal. He talks of the experience of control arising from the mere attenuation of the 'signal' that 'arises from' the actual sensory consequences of an action (Frith et al., 2000, p. 1784). However, for this information to be used it must be represented. Hence my focus on representations.

and it is not recognised in the DSM-IV-TR. So, it is not settled as to exactly what form the disorder takes. Adopting work from Damasio I believe it is possible to provide an account that gives insight into this disorder and the boundary$_B$ self.

The body and the brain interact via a variety of chemical and neural pathways (Damasio, 1994, p. 87 ; Sherwood, 2001, pp. 175, 222–224 & 635). The body provides information to the brain, which in turn regulates the activity of the body (Damasio, 1994, p. 88). For example, mean arterial blood pressure is monitored by baroreceptors (neurons sensitive to changes in pressure) in the aortic arch. These neurons signal directly to the hypothalamus. The brain can then act to alter blood pressure in a variety of ways (Sherwood, 2001, pp. 355-356). Chemical levels are also monitored. For example, there are neurons in the hypothalamus that are sensitive to sodium concentration (osmoreceptors) which monitor the osmolarity of the blood (Sherwood, 2001, p. 359). If we add to this kind of information the senses of touch, heat, cold, pain and proprioception, then there is enough information to build a very detailed map of the activity of the body (Damasio, 1994, p. 65). The brain integrates this information to produce a detailed map of the activity of the body (Damasio, 1994, p. 66).

I propose that these representations of the body include any information that enters the brain via those channels discussed above. Any representations not gained via these channels are not included in the integrated representation of the body. As such a representation of, say, a car in the visual field is not included in the integrated representation of the body as no information about the car is obtained via the relevant channels. Within the brain the body is defined as whatever is represented in this way. Using these representations we are able to represent the 'in my body/not in my body' distinction. These representations are involved in generating the phenomenology associated with the boundary$_B$ self.

This model offers us a way into understanding both boundary$_B$ self and body integrity identity disorder. First boundary$_B$ self. We have already seen how this model gives rise to the 'in my body/not in my body' distinction. Once the integrated representation of the body is in place and can be used by other parts of the brain, any sensory stimulus can be represented as occurring to the body at a particular place. That all sensory experiences are represented as occurring to the body at a particular place may be all there is to the experience of having a perspective from inside the body. Thus, this model can explain two aspects of the phenomenology of the self, namely how the body is felt to be distinct from the world and the feeling of having perspective from inside the body.[14]

This model also provides the first step in attempting to explain body integrity identity disorder. I propose that sufferers of this disorder fail to properly integrate information about the limb they desire to amputate into their representation of the body. The information must still be present, in some form, as this is only a weak form of denial of ownership. In order to explain this properly a full computational model of the process leading to the integrated representation of the body is required. I suspect that it may be important that some subjects can wilfully 'connect' to representations of the limb they wish amputated (First, 2005, p. 926). Even in the infancy of our understanding of the boundary$_B$ self it seems reasonable to suppose that body integrity identity disorder is a disorder of the boundary$_B$ self.

Body integrity identity disorder is by no means the only disorder of the boundary$_B$ self. Consider Cotard's delusion. Paradigmatically Cotard's delusion is described as a patient's belief that they have died (Gerrans, 2002, p. 50). Other patients suggest that their body no longer exists, that they have become mere objects or that they have been transported into hell (Butler, 2000, Davies, Coltheart, Langdon, & Breen, 684; Davies et al., 2002, p. 140). The following is a typical description from a patient:

> Things just happen to me now and I have no control over them. . .At times I can't even control what I want to think about. I am starting to feel pretty numb about everything because I am becoming an object and objects don't have feelings (Davies et al., 2002, p. 140).[15]

[14] Damasio appears sympathetic to this approach. There are similarities between this and his 'body as a frame of reference' approach (Damasio, 1994, pp. 235–241).

[15] It may appear that this subject has lost their sense of agency. However, this is probably an artefact of the imprecision of language. It is orthodox to interpret statements like this from Cotard's patients as a feeling of alienation from the world rather than a loss in the sense of agency (Gerrans, 2002).

Philip Gerrans has proposed that Cotard's delusion is to be explained as a lack of felt emotional responses to all sensory stimuli, at least to those sensory stimuli that are about the external world (Gerrans, 2002, p. 51). This is why:

> The Cotard patient experiences her perceptions and cognitions, not as changes in her *self*, but as changes in the states of the universe, one component of which is her body, which now feels like an inanimate physical substance, first decomposing and finally disappearing (Gerrans, 2002, p. 50).

Described as such it appears the Cotard patient is failing to distinguish between their self (in particular their body) and the rest of the world. In the framework described here, this is a failure of the boundary$_B$ self.

Why should a lack of emotional responses to external objects cause a failure of boundary$_B$ self? The answer has to do with what emotions are about. Damasio proposes that emotions are representations of body states (Damasio, 1994, pp. 131–139). Gerrans takes a similar approach, suggesting that emotions have their basis in the regulation of the body (Gerrans, 2002, p. 51). On this account when one has a felt emotional experience about some environmental object one is evaluating how good that object is for the body (Damasio, 1994, pp. 133–134). So if you feel happy upon seeing your partner, you are judging that your partner is good for your body. If you feel disgust at the smell of spoilt milk, you are judging that the milk is bad for your body. These kinds of evaluative emotional judgements involve complex representations of body states (potential or actual) that are used to continuously update the integrated body map. The Cotard patient lacks such responses to objects in their environment. Thus this kind of information is not available to update their integrated body map as they interact with the environment. As this representation becomes more and more out of date, the Cotard patient has less and less reason to suppose that they are distinct from the world. It seems to them that they are no longer interacting with the world; as such it is reasonable for them to suppose that they have died or that their body is no longer really a body.

4.3. The agentive$_M$ self

The agentive$_M$ self gives rise to the sense of generating one's own thoughts. This sense is lost in cases of the delusion of thought insertion. Patients with this delusion believe that another agent is inserting thoughts into their mind. They may describe their mind as like a screen on which the thoughts of others appear (Mellor, 1970, p. 17). We saw above that delusions of control can be explained by the patient lacking awareness of the comparison between the predicted sensory consequences of a movement and the actual sensory consequences. This explanation cannot be extended to delusions of thought insertion as thoughts do not necessarily have sensory consequences (Frith et al., 1998, p. 175) as they are not movements. By examining Frith's proposed explanation of this delusion I will shed light on the agentive$_M$ self.

We can attempt to explain the delusion of thought insertion by considering mental practice. On the model of the motor control system described in the discussion of the agentive$_B$ self, mental practice involves sending a motor command to the predictor *without sending it to the body*. This produces a representation of the predicted sensory consequences of the movement, but no actual sensory changes occur. The motor control system can then compare the predicted sensory consequences to the target sensory changes (the target state from which the motor commands are derived). From this, error signals can be derived which are used to improve the functioning of the controller (Frith et al., 1998, p. 176; Frith et al., 2000, p. 1773). If this comparison is explicitly represented it may contribute to two kinds of thoughts coming to attention, namely the mental practice of movements and mental speech. Mental speech seems to be a special case of the mental practice of movement. In particular, it is the practice of the movements of the larynx that generate vocalisations. The internal dialogue that one experiences is a representation of the predicted auditory consequences of the motor commands that generate a particular vocalisation.[16] These representations contribute greatly to our awareness (Dennett, 1991, p. 195; O'Brien & Opie, 2002, p. 11).

[16] I hold that this mechanism generates the experience of thoughts in words. It does not follow from this that this is all there is to the mind. Thinking with words is only one of the activities of the mind.

The sense of control over one's thoughts may arise from the representation of the comparison between the target state of a movement and the predicted sensory consequences of that movement. Patients suffering delusions of thought insertion experience a thought, but do not recognise that they generated it. The fact that they experience the thought suggests that they are capable of forming representations of the predicted auditory consequences of a movement. After all, it is the predicted auditory consequences that we experience during internal speech. The comparison between the predicted sensory consequences of a vocal motor command and the target state of that movement (i.e. the movement of the larynx) is what gives rise to the sense of control over one's own thoughts. If so, this suggests that patients suffering delusions of thought insertion have a deficit in the representation of this comparison. Thus I propose that the agentive$_M$ self is the capacity to be represent the comparison between the target state of a movement and the predicted sensory consequences of that movement.

It may be objected to this proposal that it cannot account for the experience of thought insertion for thoughts that are not linguistic or other forms of mental practice. Indeed this is the case; however, it is not obvious from descriptions of the delusion, such as that in Mellor (1970), that experiences of thought insertion for non-linguistic thoughts are the same disorder. It is plausible that conditions such as 'made' feelings (which some may consider a kind of non-linguistic inserted thought) are distinct disorders from delusions of thought insertion (Mellor, 1970, p. 17). However, if we hold that thought insertion occurs only for linguistic thoughts it becomes a problem for this proposal that no one experiences insertion of non-linguistic mental practice. Before we can solve this problem more empirical work is required to determine exactly what kinds of thoughts can be experienced as inserted.

In the mean time, this proposal is testable. If those with delusions of thought insertion have a deficit in forming the representation of the comparison between the target state of a movement and the predicted sensory consequences of that movement, they should be worse at mental practice than controls. To test this prediction, those with delusions of thought insertion should be compared to healthy subjects on mental practice tasks. I predict that those with delusions of thought insertion will do worse than controls. If they do not, and the theoretical problems are not overcome, we must look for an alternative explanation. It is worth noting that there is some evidence that patients with schizophrenia do have some difficulty in learning motor control tasks (Scherer, Stipp, Paquet, & Bedard, 2003; Weickert et al., 2002). However, these results are not specific to delusions of thought insertion or to mental practice.

5. The boundary$_M$ self

The boundary$_M$ self is the cognitive capacity that gives rise to the sense of having a perspective from within one's own mind. It allows one to understand what is a part of one's mind and what is not. In this sense it involves the 'in my mind/not in my mind' distinction. It is tempting to suppose that what underlies this basis for representing the mind is some form of integrated map, as used in the case of the boundary$_B$ self. However, there are none of the obvious sources of information. The brain does not scan itself to determine what is a part of it and what is not. Nor is there a place to integrate the information into a single representation, as there was for the body. It seems unlikely that there is an 'integrated mind representation' of a kind similar to that for the body.

At this point, we understand so little about the boundary$_M$ self that all I can do is suggest a way of approaching the phenomenon. With no obvious starting point I am lead to ask: what is involved in the breakdown of the boundary$_M$ self? I propose that the boundary$_M$ self is damaged in schizophrenic patients suffering the delusion of thought broadcast. The delusion of thought broadcast is defined by the belief that one's thoughts are not contained within one's own mind (Mellor, 1970, p. 17). The following is a typical description:

> As I think, my thoughts leave my head on a type of mental ticker tape. Everyone around has only to pass the tape through their mind and they know my thoughts. (Mellor, 1970, p. 17).

Here, there is a clear failure to identify the thought as occurring solely in one's mind. The subject knows that they generated the thought, hence there is no problem with their agentive$_M$ self. Furthermore, this is not an extension of the mind to include inappropriate objects. Subjects do not say that all people are in their

mind, for example. Rather the subject feels that their thought becomes external to their mind. What might explain this? Certainly it is not a failure to form an integrated body map, as it involves the representation of thoughts. Nor does it seem likely that this is a motor disorder. Since we are considering thoughts in language perhaps Frith's model of mental practice will offer some insight. Unfortunately this is not so. Clearly patients with delusions of thought broadcast form predicted sensory consequences as they do experience their thoughts. Furthermore, they have no problem with the comparison between the target and predicted sensory consequences, as their sense of control is intact. There is nowhere else for Frith's model of motor control to fail for linguistic thoughts, as there is no more to mental practice than the formation of predicted sensory consequences and the comparison of those predicted consequences to the target state. Thus, this system does not offer us a way to understand the broadcast of thoughts. However, I remain hopeful that once we understand what failures underlie this delusion we should gain some insight into the cognitive capacities that compose the $\text{boundary}_{\text{M}}$ self.

6. Conclusion

In this paper, I have argued that there is a rich phenomenology of the self. In fact, there are several senses of self. Underlying each of these various senses of self are dissociable cognitive capacities. The self, then, is the set of cognitive capacities that underlie the various senses of self. I have found that we, us healthy adult human beings, possess a rich phenomenology that composes our sense of self. We all know that we are distinct entities bounded in both body and mind. We have a sense of perspective from inside the body and from inside the mind. We feel that we initiate and control our actions and our thoughts. Furthermore we have a sense of ownership over our thoughts and actions. Beyond these moment-by-moment experiences we recognise that we are extended in time. We experience that we existed in the past and we hope to exist in the future.

Underlying this phenomenology is a set of cognitive capacities. Each of these capacities can be thought of as a distinct part of the self. For example, the diachronic self underlies experiences of temporal extendedness. In this paper, I have focused on the aspects of the self that underlie one's experience of the self as it exists at a particular moment. This is not to deny the importance of feelings of temporal extendedness. Work is needed to identify the cognitive capacities that explain one's feelings of being extended in time. Before a study into the diachronic self can proceed, we need an answer to these questions for the synchronic self. I have begun to provide such an answer here.

Within the synchronic self I have identified four empirically dissociable capacities. The $\text{boundary}_{\text{B}}$ self underlies one's experience of one's own body being distinct from the world and the sense of perspective from inside the body. It is in virtue of the $\text{boundary}_{\text{M}}$ self that one experiences the sense of being bounded within one's mind and having perspective from inside the mind. In addition to the sense of boundedness there is the sense of control. The $\text{agentive}_{\text{B}}$ self gives rise to the feeling that one controls one's bodily actions. It is in virtue of the capacities that are constitutive of the $\text{agentive}_{\text{M}}$ self that one feels that one controls one's thoughts. The interaction between these cognitive capacities explains the sense of ownership. I feel that a movement is mine if it is my body that moves, as recognised due to the $\text{boundary}_{\text{B}}$ self, *and* I initiated the movement, as recognised using the $\text{agentive}_{\text{B}}$ self. Similarly, a thought is mine if the thought occurs in my mind, as recognised due to the $\text{boundary}_{\text{M}}$ self, *and* I initiated the thought, recognised using the $\text{agentive}_{\text{M}}$ self. Once this framework is in place it is possible to begin to explain the various cognitive capacities that make up the self. I have attempted to provide a model of each of these parts of the synchronic self.

Each of these selves is to be identified with a specific part of the computational machinery that makes up the mind. The $\text{agentive}_{\text{B}}$ self arises from the capacity to represent the comparison the predicted sensory consequences of a movement to the actual sensory consequences of that movement. When there is a negligible difference between these representations one will feel that one initiated and controlled the movement. The $\text{agentive}_{\text{B}}$ self is the capacity to recognise this similarity. The $\text{agentive}_{\text{M}}$ self is also a part of the motor control system. It is the capacity to compare the predicted sensory consequences of a movement to the target state of that movement. It is a consequence of this proposal that those with delusions of thought insertion could only experience mental practice or thoughts in language as inserted. Disorders such as 'made' emotions—which may seem on first inspection to resemble thought insertion—must be distinct disorders. If the empirical predictions of this model are not borne out we would have reason to look for an alternative explanation. Testing

the predictions of this model should be a target for future research. The $\text{boundary}_{\text{B}}$ self is a somewhat different capacity. It is the capacity to form an integrated representation of the body as a whole. By forming such a representation and implicitly or explicitly defining everything not in that representation as not the body, this capacity gives rise to the 'in my body/not in my body' distinction. Furthermore, by allowing the brain to represent every sensory experience as occurring *to the body*, this integrated representation gives rise to the sense of perspective from inside the body. The $\text{boundary}_{\text{M}}$ self has proven to be far more difficult to identify. However, I am confident that through the consideration of disorders such as the delusion of thought broadcast that we will come to model this system.

Having explained the phenomenology of self in terms of a set of computational capacities we are left with another question. Have I explained the self or explained it away? This is not an issue that can be resolved conclusively here. However, the prima facie implication of this model is that after the phenomenology of the self has been explained there is nothing of the self left to explain. Thus it would seem that rather than explain the self, this model explains it away.

Acknowledgments

I thank Gerard O'Brien for supervising the project from which this paper is taken, Elizabeth Schier for countless hours of reading various manuscripts and Philip Gerrans and three anonymous reviewers for extensive comments on pervious drafts of this paper.

References

Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast circles deceive the eye and not the hand. *Current Biology, 5*, 679–685.

Barth, Jochen, Povinelli, Daniel J., Cant, & John G. H. (2005). Bodily Origins of SELF. In Beike, D., Lampinen, J. & Behrend, D. (Eds.), *the Self and memory*.

Butler, Peter V. (2000). Diurnal variation in Cotard's syndrome (Copresent with Capgras delusion) following traumatic brain injury. *Australian and New Zealand Journal of Psychiatry, 34*, 684–687.

Campbell, John (1995). Body image and self consciousness. In N. Eilan, A. Marcel & J. L. Bermudez (Eds.), *The body and the Self*.

Campbell, John (1999). Schizophrenia, the space of reasons and thinking as a motor process. *Monist, 82*(4).

Campbell, John (2002). The ownership of thoughts. *Philosophy, Psychiatry, Psychology, 9*(1).

Clark, Andy (2001). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.

Damasio, Antonio, R. (1994). Descartes' Error: Emotion, Reason and the Human Brain. Quill.

Davies, Martin, Coltheart, Max, Langdon, Robyn, & Breen, Nora (2002). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, Psychology, 8*(2–3).

de Vignemont, F., & Fourneret, P. (2004). The sense of agency: A philosophical and empirical review of the "Who" system. *Consciousness and Cognition, 13*, 1–19.

Dennett, D. (1989). The origins of selves. In D. Kolak & R. Martin (Eds.), *Self and identity: Contemporary philosophical issues*. Macmillan.

Dennett, D. (1991). *Consciousness explained*. Penguin.

Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.), *Self and consciousness multiple perspectives*. Hilsdale: Erlbaum.

Eilan, Naomi, Marcel, Anthony, & Bermudez, Jose Luis (1995). Self consciousness and the body: an interdisciplinary introduction. In N. Eilan, A. Marcel, & J. L. Bermudez (Eds.), *The body and the self*. MIT press.

First, Michael B. (2005). Desire for amputation of a limp: paraphillia, psychosis or a new kind of identity disorder. *Psychological Medicine, 35*(6), 919.

Frith, Chris, Rees, Geraint, & Friston, Karl (1998). Psychosis and the experience of the self. *Proceedings of the New York Acadamy of Science: Neuroscience of the Mind on the Centennial of Freud's Project for a Scientific Psychology, 843*, 170–178.

Frith, Christopher D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*.

Frith, Christopher D., Blakemore, Sarah-Jayne, & Wolpert, Daniel M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London, 355*, 1771–1788.

Frith, Christopher D., & Done, D. John (1989). Experiences of alien control in scizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine, 19*, 359–363.

Gallagher, Shaun (2000a). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science, 4*(1), 14–21.

Gallagher, Shaun (2000b). Self-reference and schizophrenia: cognitive model of immunity to error through misidentification. In D. Zahavi (Ed.), *Exploring the self: Philosophical and psychopathological perspectives on self-experience* (pp. 203–239). Jon Benjamins.

Gallup, Gordon G. (1970). Chimpanzees: self recognition. *Science, 167*(3914), 86–87.

Gerrans, Philip (2002). A one-stage explanation of cotard's delusion. *Philosophy, Psychiatry, Psychology, 9*(1).

Hume, David (1739). Of personal identity. *A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects*.

James, William (1890). Chapter X: the consciousness of self. *The Principles of Psychology*.

Jeannerod, Marc, Farrer, Chloe, Frank, Nicolas, Fourneret, Pierre, Posada, Andres, Daprati, Elena, & Georgieff, Nicolas, (2003). Action recognition in normal and scizophrenic subjects. In Kircher, T. & David, A. (Eds.), *The Self in Neuroscience and Psychiatry*.

Kopp, B., & Rist, F. (1994). Error-correcting behaviour in schizophrenics. *Schizophrenia Research, 13*, 11–22.

Legerstee, Maria (1998). Mental and bodily awareness in infancy: consciousness of self existence. *The Journal of Consciousness Studies, 5*(5–6), 627–644.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): the unconscious initiation of a freely voluntary act. *Brain, 106*, 623–642.

Martin, M. G. F. (1995). Bodily awareness: a sense of ownership. In N. Eilan, A. Marcel, & J. L. Bermudez (Eds.), *The body and the self*. MIT Press.

Mellor, C. S. (1970). First rank symptoms of schizophrenia. *British Journal of Psychiatry, 117*, 15–23.

Metzinger, Thomas (2003). *Being no one: The self model theory of subjectivity*. MIT Press.

O'Brien, Gerard, & Opie, Jon (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences, 22*, 127–196.

O'Brien, Gerard, & Opie, Jon (2000). Notes toward a structuralist theory of representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation*. Greenwood Publishers.

O'Brien, Gerard, & Opie, Jon (2002). Radical connectionism: thinking with (not in) language. *Language and Communication*.

O'Brien, Gerard & Opie, Jon (forthcoming). How do connectionist networks compute? *Cognitive Processing* **7**.

Povinelli, Daniel J. (2001). The self: elevated in consciousness and extended in time. In K. Skene & C. Moore (Eds.), *The development of the extended self in preschool children: Theory and research*. Cambridge University Press.

Povinelli, Daniel J., & Cant, John G. H. (1995). Aboreal clambering and the evolution of self-conception. *The Quarterly review of Biology, 70*(4), 393–421.

Proust, Joelle (2003). Thinking of oneself as the same. *Consciousness and Cognition, 12*, 495–509.

Scherer, Helene, Stipp, Emmanuel, Paquet, Franscios, & Bedard, Marc-Andre (2003). Mild Procedural Learning Deficits in Neuroleptic-Naive Patients with Schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences, 15*(1).

Sherwood, Lauralee (2001). *Human physiology: Cells to systems*. Brooks/Cole.

Stephens, G. Lynn, & Graham, George (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. MIT.

Strawson, Galen (1997). The self. *The Journal of Consciousness Studies, 4*(5–6), 405–428.

Weickert, Thomas W., Terazas, Alejandro, Bigelow, Llewellyn B., Malley, James D., Hyde, Thomas, Egan, Michael F., Weinberger, Daniel R., & Goldberg, Terry G. (2002). Habit and skill learning in schizophrenia: evidence for normal striatal processing with abnormal cortical input. *Learning and memory, 9*, 430–442.